Genomics 102 (2013) 195–201

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Discovering dominant pathways and signal–response relationships in signaling networks through nonparametric approaches

Isar Nassiri [a], Ali Masoudi-Nejad [a,*], Mahdi Jalili [b], Ali Moeini [c]

[a] Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran
[b] Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
[c] Department of Algorithms and Computation, College of Engineering, University of Tehran, Tehran, Iran

ARTICLE INFO

Article history:
Received 10 June 2013
Accepted 26 July 2013
Available online 2 August 2013

Keywords:
Signaling pathway
Dose–response relationship
Neuronal plasticity
Molecular targeted therapy

ABSTRACT

A signaling pathway is a sequence of proteins and passenger molecules that transmits information from the cell surface to target molecules. Understanding signal transduction process requires detailed description of the involved pathways. Several methods and tools resolved this problem by incorporating genomic and proteomic data. However, the difficulty of obtaining prior knowledge of complex signaling networks limited the applicability of these tools. In this study, based on the simulation of signal flow in signaling network, we introduce a method for determining dominant pathways and signal response to stimulations. The model uses topology-weighted transit compartment approach and comprises four main steps which include weighting the edges, simulating signal transduction in the network (weighting the nodes), finding paths between initial and target nodes, and assigning a significance score to each path. We applied the proposed model to eighty-three signaling networks by using biologically derived source and sink molecules. The recovered dominant paths matched many known signaling pathways and suggesting a promising index to analyze the phenotype essentiality of molecule encoding paths. We also modeled the stimulus–response relations in long and short-term synaptic plasticity based on the dominant signaling pathway concept. We showed that the proposed method not only accurately determines dominant signaling pathways, but also identifies effective points of intervention in signal transduction.

1. Introduction

One of major challenges in systems biology is to understand complex processes of signal transduction and cell response to environmental signals such as hormones, growth factors, nutrients, and other stimuli [1]. Recent technologies enable a systematic mapping of signaling interactions and constructing large signal transduction networks [2–4]. Networks generated as such are very complex and often require extensive computational tools for dissecting to the pathways and yield biological insights [4]. Parametric and non-parametric modeling methods are two general approaches in these computational tools [5]. The difficulty of obtaining parameters for broad operating conditions limits the applicability of parametric methods in signal-response relationship studies [6]. Non-parametric methods use topological data to estimate the unknown quantities , and thus, are valid for most physiological conditions and capable of predicting the system response for possible signal inputs.

Signal transduction in cells is an adaptive process [7]. When a stimulus arrives, a signaling network reassembles and large numbers of molecular contacts are established. In this step, for routing, local signal transmission tends to be stochastic. If a stimulus arrives regularly, cellular networks exhibit global routing property by developing a structural memory and learning the most efficient links to transfer the signal [8,9]. These preferred paths of signal transduction are dominant signaling pathways [10,11]. Several methods were developed for mining dominant paths within cellular signaling networks. The major challenge facing these methods is to determine which of the signaling paths within a given pair of starting and ending molecules are biologically significant [11]. Several recently developed methods addressed this issue by integrating the network structure with confidence values, sequence information, or gene-expression profiles [12–14]. In this study, we proposed a non-parametric method to determine the dominant paths, i.e., paths between internal nodes that are most effected by input stimulus. To determine the dominant signaling path, we need to consider paths from a source to a sink molecule. The difficulty of this problem is that many paths can link two molecules in a signaling network [15]. We use two established assumptions to simplify the problem. First, in biological systems, the signal flow prefers the path of least physicochemical resistance, and signals are dominantly transferred across the rationally short cascades [16]. Second, propagation of signal through edges on overlapping parallel paths increases the efficiency of signal transduction [17]. To our knowledge, there is no non-parametric and

* Corresponding author. Fax: +98 21 6640 4680.
E-mail address: amasoudin@ibb.ut.ac.ir (A. Masoudi-Nejad).
URL: http://Lbb.ut.ac.ir (A. Masoudi-Nejad).

0888-7543/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2013.07.012

probabilistic model has been reported to determine the signal transduction pathways within networks based on simulation of signal flow and routing. We used literature-mined mammalian signaling networks to systematically analyze the efficiency of the proposed modeling method for inferring pathways and signal transduction-related responses.

2. Methods

2.1. Weighting of the edges and nodes

To model a signaling network, we used directed graph model with nodes (n) representing molecules and edges (e) showing type and direction of interactions (activation/blockage) between the molecules [18]. The input (source) nodes of signaling networks represent the ligands or their receptors, the intermediate nodes consist of various kinases and second messengers, and the output (sink) nodes represent cellular responses (e.g. transcription factors) [4]. During simulation, we used normalized similarity index (NSI) to weight the edges and calibrate their efficiency for signal transduction in network [18,19]. Let us consider molecules i and j in a directed network with adjacency matrix A, *where,* A_{ij} *is the number of edges from node i to node j*. Nodes i and j can be linked by a direct interaction A_{ij} and by indirect interaction with their shared neighbor. The weight of direct interaction between node i and node j is computed as the proportion of the common shared neighbors:

$$NSI(ij) = (|A_{in} \cap A_{jn}| + A_{ij}) / \left(k_i + k_j - (|A_{in} \cap A_{jn}| + A_{ij})\right) \quad (1)$$

where, n is the neighboring node, and k_i is degree of node i [18]. Simulation of signal transduction starts from ligand molecule(s) and iteratively traverses the whole network by a modified breadth-first-search (BFS) method [18]. The state of a node j, (X_j), is the activity-level of given molecule. At each time interval [$X(t)$], state of a node j is updated based on its value in the previous time step [$X(i-1)$] states of activator(s) ($X_{A_{i\to j}}$) and repressor(s) ($X_{R_{i\to j}}$) nodes feeding into it as:

$$X_j(t) = \left[X_j(t-1) \times (0.7 \le R \le 1)\right] + \left[1 - \prod_{i \epsilon A}\left(1 - X_{A_{i\to j}}.NSI_{ij}\right)\right] \times \prod_{i \epsilon R}\left(1 - X_{R_{i\to j}}.NSI_{ij}\right) \times \left[1 - X_j(t-1)\right] \quad (2)$$

where, each directed edge (i, j) models the transduction of signals from node i to node j of activator or repressor type [18].

2.2. Finding signaling paths between two molecules

A walk on a directed graph is an intermittent sequence of nodes and edges, beginning and ending with node ($\mathcal{N} = \{n_0, n_1, \ldots, n_{l-1}, n_l\}$). Walks without repeated edges are called trails, and those without repeated edges and nodes are called paths [20]. We refer to n_0 as the source node and n_l as the sink (or target) node of the signal transduction path. To find signaling paths between source and target molecules, and compare the properties of dominant and secondary paths, we used approximation traversal algorithms available in graph theory [20]. This approximation algorithm finds optimal paths from all possible paths.

The algorithm uses breadth-first search, and iterates over potential paths between source and target nodes in the given directed network. During the search, among routes with intermediate node(s) in their center-back parts, the algorithm tries to choose those reaching the target node faster (Fig. 1). This method is well suited for our problem, as in biological systems the signal flow prefers the path of least physicochemical resistance [21,22]. Without setting a limit on path length, the algorithm finds the partially or completely non-overlapping paths with polynomial runtime. Table 1 shows the pseudo code for the algorithm.

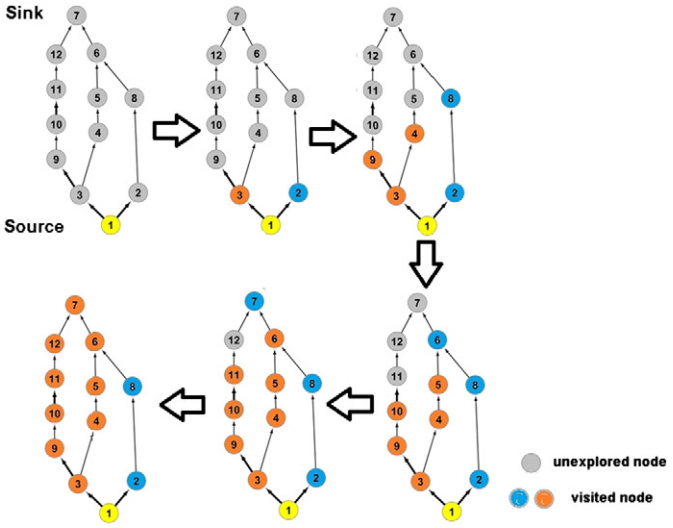

Fig. 1. To illustrate our approach to select optimal signaling paths between two molecules, suppose that we are concerned with finding paths, which transfer signal from molecule 1 into 7. In the given network, first, algorithm generates trails starting from source to target node. In the network G, the resulting trail sets are {a:1,2,8,6,7}, {b:1,3,4,5,6,7}, and {c:1,3,9,10,11,12,7}. Between trails a, and b with intermediate node 6, we select path a, since it reaches the sink node faster.

2.3. Topology-weighted transit compartments function to score signal transduction paths

Because cellular signal transduction is a path of pair-wise interactions, we applied the transit compartment model and queuing theory to score

Table 1
Non-parametric computing dominant signaling pathway (NSIP) algorithm. First, algorithm simulates the signal flow in the connected sub-network that begins with the first node, and records the efficiency of signal transduction between all node pairs and proportion of active form for each node. According to the weights of nodes and edges, paths between source and sink nodes are scoring by a specific function. Path with the highest score is a plausible candidate for a dominant signaling path.

Algorithm: Nonparametric computing dominant signaling pathways (NSIP)
Problem: simulation of signal flow and inferring dominant signal transduction pathway
Inputs:
$N(n)$: set of nodes in Graph G to which n is connected directly as parent
$X(N)$: Nodes and edges weighting functions
$P(i)$: is a linking list includes paths from source to node i
so: source node
si: sink node
Outputs:
C: paths topology
D: dominant path

Body of algorithm:
1: calculate the weight of nodes and edges according to the $X(N)$
2: q
3: enqueue (q, so)
4: visited(so) true
5: visited(si) true
6: **while**(q)
7: n dequeue(q)
8: **for each** $w \in N(n)$ **do**
9: **if** !visited(w) **then**
10: enqueue(w)
11: visited(w) true
12: **end if**
13: Add $P(n)$ to $P(w)$
14: Add 0 to $P(w)$
15: Insert n into $P(w)$ before each 0
16: **end for**
17:**end while**
18:C P( )
19:Calculate the score for all paths in C
20:D the path in C with the highest score