BIOINFORMATICS

Vol. 20 no. 12 2004, pages 1877–1886
doi:10.1093/bioinformatics/bth173

Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data

Eduardo Sontag[1], Anatoly Kiyatkin[2] and Boris N. Kholodenko[2,*]

[1]Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA and [2]Department of Pathology, Anatomy and Cell Biology, Thomas Jefferson University, 1020 Locust Street, Philadelphia, PA 19107, USA

Received on September 26, 2003; revised on January 29, 2004; accepted on February 3, 2004
Advance Access publication March 22, 2004

ABSTRACT

Motivation: High-throughput technologies have facilitated the acquisition of large genomics and proteomics datasets. However, these data provide snapshots of cellular behavior, rather than help us reveal causal relations. Here, we propose how these technologies can be utilized to infer the topology and strengths of connections among genes, proteins and metabolites by monitoring time-dependent responses of cellular networks to experimental interventions.

Results: We demonstrate that all connections leading to a given network node, e.g. to a particular gene, can be deduced from responses to perturbations none of which directly influences that node, e.g. using strains with knock-outs to other genes. To infer all interactions from stationary data, each node should be perturbed separately or in combination with other nodes. Monitoring time series provides richer information and does not require perturbations to all nodes. Overall, the methods we propose are capable of deducing and quantifying functional interactions within and across cellular gene, signaling and metabolic networks.

Contact: boris.kholodenko@jefferson.edu

Supplementary Information: Supplementary material is available at http://www.dbi.tju.edu/bioinformatics2004.pdf

INTRODUCTION

Remarkable progress in genetics and molecular biology has permitted the sequencing of the genomes of a number of species and the determination of a plethora of protein and lipid components of intracellular signaling networks (Li *et al*., 2002). High-throughput technologies are capable of monitoring the expression levels of large gene sets and the activity states of signaling proteins giving us snapshots of transcriptional and signaling behavior of living cells. However, the web of regulatory interactions among components of cellular networks remains largely unknown at the present time. For instance, to relate genes to one another, current techniques use clustering algorithms, which group genes that appear to be coherently activated or inactivated (Claverie, 1999; Niehrs and Pollet, 1999). Genes with similar expression patterns can be placed together into a cluster, yet the functional interactions between these and other genes are unknown. Another example comes from the biology of signaling networks. Although the basic architecture of mitogen-activated protein kinase (MAPK) cascades has been worked out, the complete pattern of feedback regulatory loops and cross-talk between various MAPK pathways and other signaling systems remain elusive (Bagowski and Ferrell, 2001; Ferrell and Machleder, 1998; Langlois *et al*., 1995; Yu *et al*., 2002).

There is an intrinsic difficulty in capturing network interactions using traditional genetic experiments or pharmacological interventions. Any perturbation to a particular gene or a signaling component rapidly propagates through a network. In fact, it is practically impossible to carry out an experiment designed to observe how a change in one node directly affects another node, because interconnections will cause widespread (global) changes in a network. Following experimental perturbations to intact cells, only global responses of an entire network can be assessed. The question is then how to use the observed global changes to derive interactions between individual nodes.

This problem has generated an effort by many research groups whose goal is to infer mechanistic relationships underlying the observed behavior of complex molecular networks. A series of studies was concerned with the determination of reaction mechanisms and the deduction of biochemical pathways from measurements of the time-course of species concentrations (Arkin and Ross, 1995; Arkin *et al*., 1997; Chevalier *et al*., 1993; Samoilov *et al*., 2001; Tyson, 1975; Vance *et al*., 2002). Boolean networks, genetic algorithms, dynamic simulations and Bayesian models have been applied to infer the gene circuitry (Bhan *et al*., 2002; D'Haeseleer

*To whom correspondence should be addressed.

et al., 2000; Gat-Viks and Shamir, 2003; Husmeier, 2003; Ideker *et al*., 2000; Jayaraman *et al*., 2000; Repsilber *et al*., 2002; Tegner *et al*., 2003; Wahde and Hertz, 2000; Wiggins and Nemenman, 2003; Yeung *et al*., 2002). Metabolic Control Analysis (MCA) demonstrated that measurements of sensitivities of the stationary fluxes and concentrations to changes in each enzyme activity (the so-called control coefficients) allow for determining interactions between metabolites and enzymes (in terms of the so-called elasticity coefficients) (Cascante *et al*., 1989; Kholodenko and Westerhoff, 1995; Westerhoff *et al*., 1994). However, an implicit assumption of the MCA approach is that the pathway stoichiometry is known (Schuster *et al*., 2000), which is not the case for signaling and gene networks. de la Fuente *et al*. (2002) advanced steady-state MCA techniques to reconstruct gene networks. At the same time, this technique imposed restrictions on the network architecture, disallowing each network node to be generated and consumed by multiple processes, and cannot be applied to signaling and metabolic networks.

A more general strategy, capable of inferring the architecture of signaling and gene networks and allowing the employment of a modular framework, was recently reported (Kholodenko *et al*., 2002) [the methods proposed by de la Fuente *et al*. (2002) and Kholodenko *et al*. (2002) were reviewed by Stark *et al*. (2003)]. Yet, both approaches dealt only with steady-state behavior and implied that network nodes are connected through regulatory interactions that exclude mass flow. Thus, the tools developed so far have limited applications to *in vivo* systems, particularly when a biological process is intrinsically time dependent, such as the cell cycle, or includes both information feedbacks and mass flow connections through biochemical conversions (Kholodenko *et al*., 1999; Tyson *et al*., 2001). In addition, many previous methods required experimental interventions to be absolutely specific, influencing one node at a time, whereas more often experimental interference simultaneously affects several nodes.

The present paper develops powerful quantitative techniques to unravel functional interactions between genes, transcription factors and metabolites from time series data on cellular responses to perturbations. Compared to stationary data, time series offer particularly rich opportunities for understanding the dynamics of biological processes and may help to deal with the problem of noise. We demonstrate that monitoring time-dependent responses enables the determination of causal relationships even when not all of the network nodes can be perturbed. Avoiding the nodes that do not permit direct experimental interventions should be accompanied by the application of two or more independent perturbations to other nodes. Our strategy is illustrated and tested *in silico* using computer-generated responses corresponding to realistic experimental protocols.

SYSTEM AND METHODS

Quantifying direct connections between network nodes

A conceptual framework has been developed by us to quantify molecular interactions in cellular networks (Brown *et al*., 1997; Bruggeman *et al*., 2002; Kholodenko *et al*., 1997b). A basic concept is to analyze the direct effect of a small change in one network node on the activity of another node, while keeping all remaining nodes (variables) 'frozen'. Here, these ideas are applied to reconstructing cellular networks from time-series of gene expression, signaling and metabolite data.

The dynamic behavior of a network is often described by a set of differential equations,

$$d\boldsymbol{x}/dt = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{p}), \quad \boldsymbol{x} = (x_1, \ldots, x_n), \quad \boldsymbol{p} = (p_1, \ldots, p_m), \tag{1}$$

where a single state variable x_i is assigned to each network node, representing its concentration or activity level, and the corresponding function f_i describes how the rate of change of x_i depends on all other elements of the network. The parameters ($\boldsymbol{p}$) represent any external or internal condition maintained constant, e.g. external concentrations, rate constants, pH, temperature. If all molecular interactions were identified by a mechanistic ('bottom-up') approach, we would be able to reconstruct the functions f_i in terms of the so-called chemical kinetic equations, provided all forward and backward rate constants were known (Moehren *et al*., 2002; Schoeberl *et al*., 2002; Tyson *et al*., 1996). In practice, our knowledge of both network interactions and rate constants is far from complete, and here we develop a 'top-down' approach to estimate the derivatives $\partial f_i/\partial x_j$, which precisely describe the influence of each variable x_j upon the rate (f_i) of change of every other variable x_i. These derivatives form the Jacobian matrix, **F**, which is well known in mathematics and engineering,

$$\mathbf{F} = (\partial \boldsymbol{f}/\partial \boldsymbol{x}).$$

If an entry F_{ij} of the matrix **F** is zero at any time, component x_j has no direct effect on component x_i. In this case, there is no edge from node j to node i at the connection graph associated with the network. For a non-zero element F_{ij}, node j connects to node i at the connection graph (Fig. 1). This interaction has both direction and sign. Node j affects node i, but not vice versa if $F_{ji} = 0$. In fact, a gene (x_j) may influence another gene (x_i), whereas (x_i) may have no influence on (x_j). If $F_{ji} > 0$, node j activates node i by enhancing the net rate of x_i production, and if $F_{ji} < 0$, node j inhibits node i. The F's values depend on a particular state ($\boldsymbol{x}$) of the system at a given time, and specify the dynamics of the positive and negative interaction strengths between network nodes.

Our goal is to find these interaction strengths (given by the Jacobian matrix **F**) from experimental time series obtained by monitoring network dynamic responses. Biochemistry has a long history of exploring the stationary behavior of cellular pathways. Steady-state kinetic studies have provided

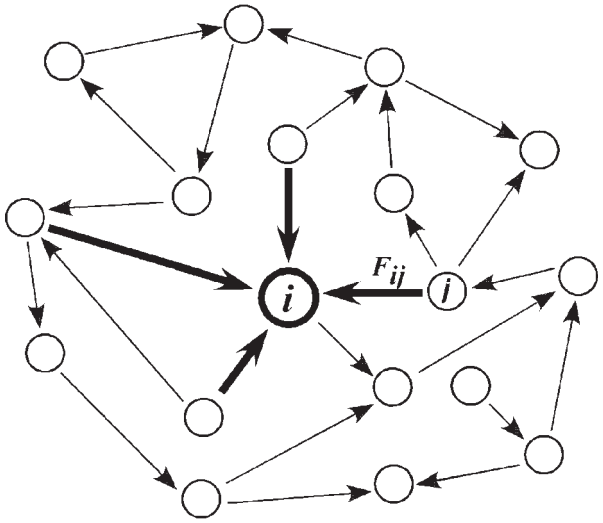

Fig. 1. Network interaction map. Connections leading to node i of the network correspond to non-zero elements F_{ij} of the ith row of the Jacobian matrix. These connections are shown in bold. The F_{ij} value quantifies the interaction strength, indicating how the rate of change in the activity of node i depends on node j.

pertinent information on pathway regulation and protein and metabolite interactions (Edwards *et al*., 2001; Kholodenko and Westerhoff, 1995). At the same time, measurements of system variables at steady states cannot aid in determining the elements F_{ij}, which describe local interactions. Indeed, the multiplication of functions f_i in Equation (1) by any constant factor does not affect the steady-state behavior, whereas it changes the Jacobian matrix **F**. Experimental data on stationary responses to perturbations can only determine the Fs up to arbitrary scaling factors. For instance, it is convenient to divide F_{ij} by the negative diagonal elements $-F_{ii}$. The resulting dimensionless coefficients, $r_{ij} = -F_{ij}/F_{ii}$, are referred to as the connection coefficients (Kholodenko *et al*., 2002). When stable elements are interconnected into a network ($F_{ii} < 0$), r_{ij} and $F_{ij} (i \neq j)$ have the same sign, which is plus or minus depending on whether network node j activates or inhibits node i. In biological terms, the connection coefficient r_{ij} tells us how much x_i will change in response to a causative change in x_j, when all other network nodes are kept constant, while node i is relaxing to its new steady state (meaning a conceptual 'isolation' of node i from other network interactions while characterizing the direct influence of x_j). The ratio of these changes $(\partial x_i/\partial x_j)$ is evaluated from the steady-state condition for node i, $f_i(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n, \boldsymbol{p}) = 0$, considered at constant values of x_k for $k \neq i, j$. The differentiation with respect to x_j gives $\partial x_i/\partial x_j = -(\partial f_i/\partial x_j)/(\partial f_i/\partial x_i) = -F_{ij}/F_{ii} = r_{ij}, i \neq j$. Therefore, in the limit of infinitesimal changes, the ratio $\Delta x_i/\Delta x_j$ equals r_{ij} and quantifies the direct impact of node j on node i at steady states. Previously, we demonstrated how the connection coefficients r_{ij} can be inferred from steady-state measurements (Kholodenko *et al*., 2002). Here, we develop a technique allowing for determination of the dynamics of gene interactions described by the Fs from time-dependent network responses to experimental interventions (Kholodenko and Sontag, 2002). The inferred interaction dynamics offers a more valuable and complete description of a cellular network than the less informative steady-state data.

RESULTS

Network reconstruction from time series data does not require perturbations to all nodes

A key issue is to develop an experimental protocol capable of collecting all necessary data required to determine the network architecture. A priori, it is not obvious what kind of perturbations should be applied and how many experiments should be carried out to infer all network connections. The key to our method and a distinguishing feature from other network identification approaches is the following experimental design. To determine connections leading to each component x_i, a set of experimental interventions that do not directly influence x_i is selected. Each of these perturbations may directly affect one or several nodes different from x_i. Formally, for each $x_i (i = 1, \ldots, n)$, we choose a subset $\boldsymbol{P}_i$ of parameters p_j known to have the property that the function f_i in Equation (1) does not depend upon p_j,

$$\partial f_i/\partial p_j(\boldsymbol{x}, \boldsymbol{p}) = 0, \quad p_j \in \boldsymbol{P}_i. \tag{2}$$

This prior information about the system is far less restrictive than it may first appear. Indeed, it is usually the case that biological information is available, for instance, telling us that a certain protein has no direct influence on an unrelated gene or biochemical interaction, or a certain inhibitor of a membrane kinase has no direct influence on a cytoplasmic phosphatase, and so on.

For each perturbation, we will measure the original and perturbed time series (called trajectories) describing the time dependence of network variables (e.g. the gene activities and the functional states of proteins). These trajectories are the solutions to Equation (1) corresponding to parameter values p_j and $p_j + \Delta p_j$, respectively, and to the same initial condition $\boldsymbol{x}^0$ (unless a perturbation is a change in the initial condition). The time-dependent response $R_{ij}(t)$ of each network variable x_i to a perturbation of p_j is defined as the parameter sensitivity of the solution $x_i(t, \boldsymbol{x}^0, \boldsymbol{p})$ to Equation (1):

$$R_{ij}(t) = \partial x_i(t, \boldsymbol{x}^0, \boldsymbol{p})/\partial p_j = \lim_{\Delta p_j \to 0} \left(\frac{x_i(t, p_j + \Delta p_j) - x_i(t, p_j)}{\Delta p_j} \right), \quad i = 1, \ldots, n. \tag{3}$$

These sensitivities are computed using the variational system along the corresponding trajectory and are routinely employed in differential equation theory and systems analysis for parameter optimization and identification of dynamical systems (e.g. Bryson and Ho, 1969; Lee and Markus, 1968; McBride and Narendra, 1965; Saltelli *et al*., 2000; Sontag, 1993). They have been used before in the context of metabolic networks (see Ingalls and Sauro, 2003; Kholodenko *et al*., 1997a).

Our objective is to determine dynamic connections, given by the Jacobian elements $F_{ij}(t)$ (Fig. 1), from the experimental time series that evaluate global response coefficients $R_{ij}(t)$. We will also need the second-order sensitivities, $\rho_{ij}(t)$, which can be estimated from the measurements of the responses $R_{ij}(t)$ at two successive time points t and $t + \Delta t$,

$$\begin{aligned}\rho_{ij}(t) &= \partial^2 x_i(t, \boldsymbol{x}^0, \boldsymbol{p})/\partial p_j \partial t = \partial R_{ij}(t)/\partial t \\ &\approx [R_{ij}(t + \Delta t) - R_{ij}(t)]/\Delta t. \end{aligned} \quad (4)$$

With the subset $\boldsymbol{P}_i$ of parameters p_j that do not directly affect the rate of change $f_i(\boldsymbol{x}, \boldsymbol{p})$ of node i [Equation (2)], we associate the matrix $\mathbf{R}(t, \boldsymbol{P}_i)$ composed of the measured time-dependent responses, $R_{kj}(t)$. Each jth column of the response matrix $\mathbf{R}(t, \boldsymbol{P}_i)$ corresponds to a single experiment, in which the parameter p_j is perturbed and the time-course of the response of each network node x_k to a change in p_j is evaluated (see Experimental design). Therefore, the matrix $\mathbf{R}$ has n rows and as many columns as selected parameters p_j that immediately influence either a single node different from x_i, or any combination of such nodes. For each perturbation experiment (p_j), the second-order sensitivity $\rho_{ij}(t)$ is evaluated from time series for node x_i. By using the matrix $\mathbf{R}$ and the $\rho_{ij}(t)$ values, the problem that we posed can be solved as follows.

The dynamic behavior of node i is described by the solution $x_i(t, \boldsymbol{x}^0, \boldsymbol{p})$ to Equation (1),

$$dx_i(t, \boldsymbol{x}^0, \boldsymbol{p})/dt = f_i(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n, \boldsymbol{p}).$$

Taking the derivatives on both sides with respect to p_j and using Equations (2)–(4), we find that the unknown elements of the ith row $(F_{i1}, \ldots, F_{in})$ of the Jacobian matrix $\mathbf{F}$ satisfy the following system of linear equations:

$$\rho_{ij}(t) = \sum_{k=1}^{n} R_{kj}(t) \cdot F_{ik}(t), \quad p_j \in P_i. \quad (5)$$

This equation gives us the answer we were looking for: the n Jacobian entries, F_{ik}, quantifying the influence of every node k on node i (Fig. 1) can be determined from the measured responses, $R_{kj}(t)$ and $\rho_{ij}(t)$, to perturbations that do not directly affect node i, provided that the rank of the matrix $\mathbf{R}(t, P_i)$ equals n. Importantly, to deduce dynamic connections leading to node i, the second-order sensitivities ρ_{ij} should be measured only for component x_i, whereas for all other components x_k the first-order sensitivities $R_{kj}(t)$ should be determined. Supplementary Proof 1 demonstrates that for any n independent perturbations, the rank of the response matrix $\mathbf{R}(t, \boldsymbol{P}_i)$ generically equals n at any given time. Moreover, this rank generically equals n even when only a single network node is directly affected by n experimental interventions, each of which changes an independent parameter influencing that particular node. For instance, a purely uncompetitive inhibitor can change the $V_{\max}$ of a Michaelis–Menten reaction, whereas the K_{m} can be affected by a competitive inhibitor, and time-dependent responses to both perturbations can be used to unravel the network architecture. Interestingly, one of the perturbations could be a change in the initial condition (e.g. the protein abundance). We conclude that estimations of the time-varying sensitivity coefficients allow us to completely infer and quantify the network connections, even if experimental interventions can directly perturb only selected network components, the number of which is less than n.

Avoiding derivatives Importantly, by employing genetic or pharmacological tools to perturb cells, one does not need to measure the sizes of the resulting parameter changes (Kholodenko *et al*., 2002). In fact, such measurements would be difficult if not impossible to make *in vivo*. Instead of the parameter sensitivities defined in terms of derivatives, we can simply consider the global changes (Δx_i) in network variables caused by a perturbation (Δp_j) and introduce the quantities $\Delta R_{ij}(t)$ and $\Delta\rho_{ij}(t)$ as follows:

$$\begin{aligned}\Delta R_{ij}(t) &= x_i(t, p_j + \Delta p_j) - x_i(t, p_j), \\ \Delta\rho_{ij}(t) &= [\Delta R_{ij}(t + \Delta t) - \Delta R_{ij}(t)]/\Delta t. \end{aligned} \quad (6)$$

Using these finite differences, as an approximation of mathematically correct infinitesimal changes, one obtains exactly the same relationship as Equation (5), but for $\Delta R_{ij}(t)$ and $\Delta\rho_{ij}(t)$,

$$\Delta\rho_{ij}(t) \approx \sum_{k=1}^{n} \Delta R_{kj}(t) \cdot F_{ik}(t), \quad p_j \in \boldsymbol{P}_i. \quad (7)$$

Therefore, network connections can be expressed in terms of the measured changes in the levels of intermediates, without requiring any knowledge about the values of parameter changes. Indeed, Equation (5), which used the derivatives, differs from the (approximate) estimates Equation (7) involving finite changes by only multiplication of both sides by Δp_j.

Steady-state conditions necessitate perturbations to each network node If a network under consideration approaches a stable steady state, the rank of the matrix $\mathbf{R}(t, \boldsymbol{P}_i)$ decreases as time approaches infinity. In fact, the maximal possible rank of the steady-state response matrix, $\mathbf{R}(\boldsymbol{P}_i) = \lim_{t\to\infty} \mathbf{R}(t, \boldsymbol{P}_i)$, equals $n - 1$, and to achieve this rank at least $n - 1$ perturbation experiments are required, in which all nodes except node i are directly perturbed, separately or in combination.

Therefore, when (quasi)steady-state behavior is displayed, the Jacobian elements cannot be determined, and only the scaled Jacobian elements, $r_{ij} = -F_{ij}/F_{ii}$, will be determined using the following equation (cf. Kholodenko *et al*., 2002):

$$\sum_{k=1,\ k\neq i}^{n} \Delta R_{kj} \cdot r_{ik} \approx \Delta R_{ij}. \qquad (8)$$

Equation (8) determines the connection coefficients r_{ij} using the finite differences ΔR_{ij}, which are equal to changes in state variables following a transition from an initial to a new steady state. The specificity of a particular perturbation is not important, as long as this perturbation does not directly affect node i. Notably, replacing the response coefficients by finite changes helps use results of systematic gene knock-outs to infer the architecture of gene networks (de la Fuente *et al*., 2002; Kholodenko *et al*., 2002).

Biological noise Since our strategy relies on the measurement of the differences between perturbed and unperturbed dynamics, and involves determining the vectors orthogonal to experimentally obtained changes, robustness to experimental noise is an important issue awaiting theoretical and experimental analyses. A comparison of the connection strengths determined at multiple time points using time series can help one to rule out false positive or false negative connections.

Experimental design: practical steps to infer the strengths of dynamic connections leading to any given network node

(1) Apply an experimental setup, where the states (activities) of nodes can be measured during a transient process, for instance, throughout a transition from a resting to an active state of a network. This transient process can be initiated by cell stimulation with a ligand, or the system behavior may be inherently transient as for cell cycle or circadian rhythm oscillators.

(2) Select an experimental intervention known to have no direct influence on a given node (x_i). It can be a change in external ligand concentration, a change in the initial concentration (activity) of a component different from x_i, a pharmacological manipulation or the use of nucleic acid-based technologies, such as tetracycline-inducible expression and small RNA interference. At selected time points, monitor the perturbed and unperturbed values of all n network nodes and determine the differences according to Equation (6). These data provide one column of the matrix $\mathbf{R}(t, \boldsymbol{P}_i)$ and one coefficient $\rho_{ij}(t)$. Note that it is often convenient to normalize the differences by the mean values; hence, determine the fractional changes (Kholodenko *et al*., 2002).

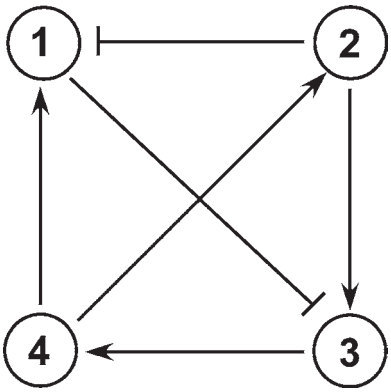

Fig. 2. Schematic four-gene network. Arrows indicate activation, and lines with blunt ends represent repressions.

(3) Repeat step (2), in full carrying out as many perturbation experiments as there are nodes (n), and obtain the matrix $\mathbf{R}(t, \boldsymbol{P}_i)$ and n coefficients $\rho_{ij}(t)$.

(4) At each selected time point, $t > 0$, solve Equation (5) to find the Jacobian elements $F_{ij}(t)$.

(5) To obtain a complete architecture of network connections, steps (2) and (3) should be repeated for each node $x_i, i = 1, 2, \ldots n$. Note that the same measurements (columns of the response matrix) can be used for different nodes provided these nodes were not directly affected by the selected perturbations.

Network models used to test the proposed approach *in silico*

Reverse engineering gene networks First, the validity of our approach is demonstrated for a four-gene network (Fig. 2). We constructed this *in silico* network using kinetic mechanisms and parameters for gene interactions drawn from the literature (Hargrove *et al*., 1991; Reinitz and Sharp, 1995; von Dassow *et al*., 2000). Network nodes represent the mRNA concentrations, which are determined by the transcription (v_i^{synth}) and degradation (v_i^{degr}) rates, $\mathrm{d}[\text{mRNA}_i]/\mathrm{d}t = v_i^{\text{synth}} - v_i^{\text{degr}}$. Gene interactions result in non-linear dependences of these rates on other mRNA_j concentrations. As previously reported (Kholodenko *et al*., 2002; von Dassow *et al*., 2000), we describe the rates by the Hill-type equations (Supplementary Table 1).

Following the experimental setup described above, we analyzed $[\text{mRNA}_i]$ transients from a resting state, where all four genes were inactive, to a stable activity state (Fig. 3). Four perturbations to the transcription and degradation rates were applied (indicated in Supplementary Table 1), and the kinetic model generated non-linear network responses in lieu of experimental measurements. Using these data, the finite differences between the control and perturbed transitions were obtained according to Equation (6) (as schematically

Table 1. A snapshot of the retrieved 'experimental' and known 'theoretical' interaction strengths for the gene network model of Figure 2.

F_{11}= -2.0[a] / -1.3[b]	F_{12}= -1.1[a] / -1.0[b]	F_{13}= 0.0[a] / 0.0[b]	F_{14}= 0.6[a] / 0.5[b]
F_{21}= 0.0[a] / 0.0[b]	F_{22}= -0.3[a] / -1.7[b]	F_{23}= 0.0[a] / 0.0[b]	F_{24}= 0.4[a] / 0.5[b]
F_{31}= -0.5[a] / -0.5[b]	F_{32}= 0.4[a] / 0.4[b]	F_{33}= -3.7[a] / -2.9[b]	F_{34}= 0.0[a] / 0.0[b]
F_{41}= 0.0[a] / 0.0[b]	F_{42}= 0.0[a] / 0.0[b]	F_{43}= 2.6[a] / 3.0[b]	F_{44}= -1.5[a] / -2.0[b]

The Jacobian elements F_{ij} are determined at 0.5 h after a transition of the gene network from a resting state to an active state began.
[a]Retrieved 'experimental' interaction strength.
[b]'Theoretical' interaction strength.

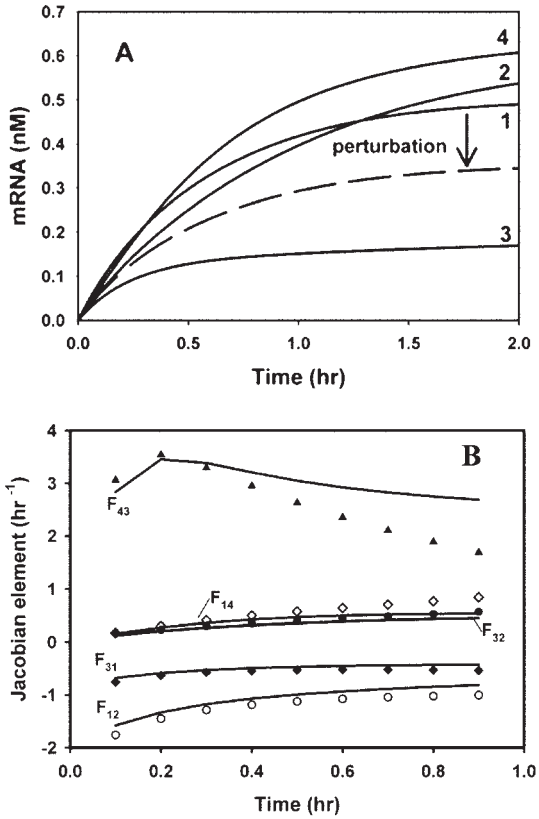

Fig. 3. Reverse engineering of dynamic gene interactions. (**A**) Computer-generated transients of mRNA concentrations. Numbers correspond to the genes in Figure 2. The dashed line indicates the response of gene 1 to a 30% decrease in its transcription rate. (**B**) Dynamics of activation or repression of gene i by gene j is quantified by the Jacobian element F_{ij}. Correct 'theoretical' values (solid lines) and 'experimental' estimates of non-zero elements F_{ij}, deduced using 10% perturbations, are plotted as functions of time during the transition from a resting state to a stable activity state of the network, open circles—F_{12}; open diamonds–F_{14}; closed diamonds—F_{31}; closed circles—F_{32}, closed triangles—F_{43}.

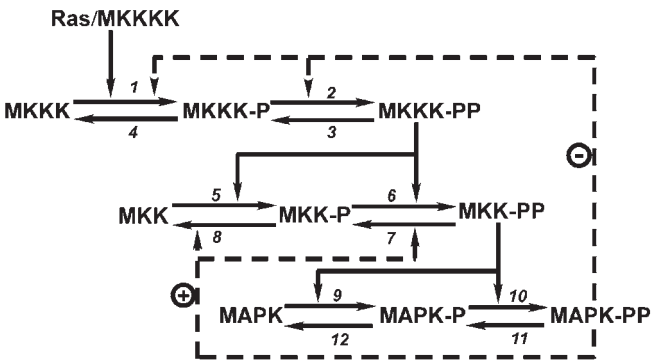

Fig. 4. Kinetic scheme of an MAPK cascade. P and PP designate monophosphorylated and bisphosphorylated protein forms. Feedback effects of MAPK on the rate of MKKK phosphorylation and MKK dephosphorylation are shown schematically by dashed lines.

illustrated in Fig. 3A). Finally, using Equation (7) we inferred both the architecture and the time dependence of the strength of functional interactions between genes during rest-to-activity transition. Figure 3B depicts the dynamics of non-zero elements of the Jacobian matrix and demonstrates the accuracy with which the topology and the strength of dynamic connections were retrieved. Table 1 provides a snapshot of the entire Jacobian matrix at 0.5 h after the transient began. As steady state is approached, the deviation between the retrieved and correct interaction strengths begins to rise (Fig. 3B). Indeed, only the normalized Jacobian elements (connection coefficients r_{ij}) can be determined when quasi-stationary responses are monitored (Kholodenko *et al.*, 2002).

Reverse engineering signaling networks Signaling through MAPK pathways plays a crucial role in many cellular processes, as diverse as growth, proliferation, differentiation and apoptosis (Chang and Karin, 2001; Levchenko *et al.*, 2000). MAPK cascades are evolutionarily conserved from yeast to mammals and usually consist of three levels, where the activated kinase at each level phosphorylates the kinase at the next level down the cascade (Fig. 4). At the third (terminal) level, the bisphosphorylated, active form of a MAPK kinase (MKK–PP) phosphorylates MAPK. At the second level, MKK is phosphorylated by active MAPK kinase kinase (MKKK–PP), which, in turn, is phosphorylated at the first level by plasma membrane kinases in a sequence of steps involving also the small GTPase Ras (shown as Ras/MKKKK). A kinetic

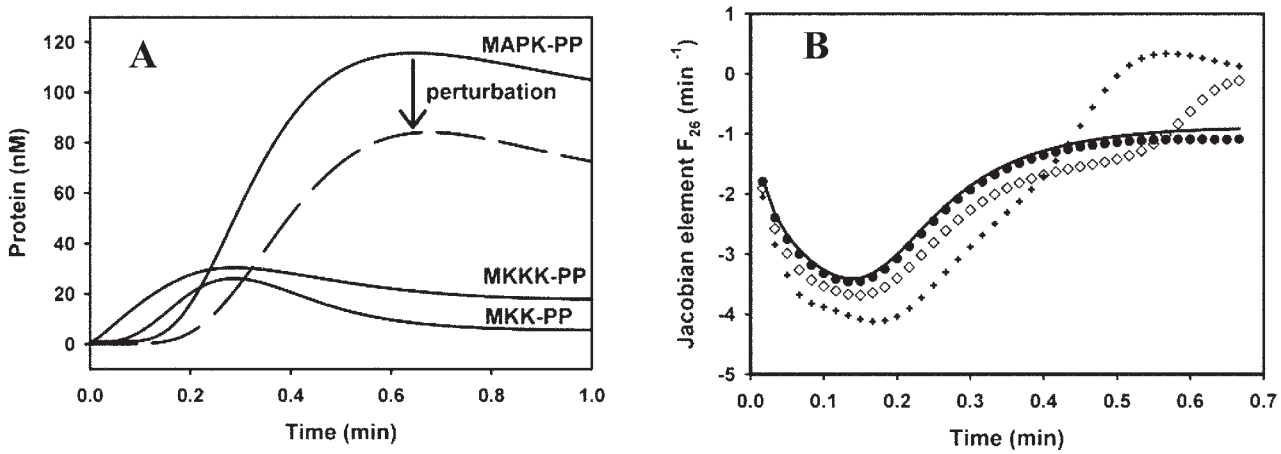

Fig. 5. Inferring dynamic connections successfully: the retrieved time-dependent strength of a negative feedback from MAPK–PP to MKKK–PP. (**A**) Computer-generated transients of MAPK cascade components, MAPK–PP, MKK–PP and MKKK–PP, are shown by solid lines. The dashed line indicates the perturbed MAPK–PP transient (50% decrease in k_{cat} for MKK kinase and 50% increase in V_{max} for MAPK phosphatase). (**B**) Time-dependent connection strength during the transition of the cascade from a resting state to an active state. The Jacobian element F_{26} quantifies a negative feedback from MAPK–PP to MKKK–PP. Solid line, correct 'theoretical' value of F_{26} is plotted as a function of time. Symbols show how accurately the Jacobian element F_{26} is determined by using 'experimental' interventions with different magnitudes of parameter perturbations, closed circles—5% , open diamonds—25% and plus—50% , (see Supplementary Tables 1 and 2 for further details).

model for this pathway is given in Supplementary Table 2. This model possesses two feedback loops (Fig. 4) and was published elsewhere (Kholodenko *et al.*, 2002). We used the model to generate time-dependent responses of the network variables. Because of moiety conservation, there are only two independent variables at each cascade level, meaning that this network has six nodes in total. It is important to realize that there is mass flow between unphosphorylated, monophosphorylated and bisphosphorylated protein forms at each MAPK cascade level, and information flow between different levels, which communicate through regulatory interactions. In contrast with prior approaches, our method can handle well both mass flow and regulatory interactions.

We monitored a transition of the MAPK pathway from a resting state to a stable activity state (Fig. 5A). To infer connections leading to each node we applied six different perturbations affecting one or several reactions (and nodes, respectively) in the MAPK pathway. Using computer-generated responses, we calculated the finite differences between the control and perturbed transitions (Fig. 5A). Finally, using these data, the Jacobian elements were retrieved by numerical calculation of the solution to Equation (7). Figure 5B illustrates that 5, 25 and even 50% perturbations unravel a negative feedback from MAPK–PP to MKKK–PP and quantify its dynamic strength with a good fidelity at practically any time before a stable steady state is reached. Supplementary Table 3 compares the correct (theoretical) Jacobian elements with 'experimentally' deduced interaction strengths and demonstrates that the architecture of the entire MAPK pathway is deduced by our method from the observed time series.

Importantly, the method described here enables us to reconstruct the dynamics of the connection strengths not only for gene and signaling networks that exhibit stimulus-induced transitions from one stable state to another, but also when no stable steady state exists and a system displays sustained oscillatory behavior. Here, we highlight the power of our method using a model of a protein kinase/phosphatase cascade, such as MAPK cascade (Supplementary Table 4), where a strong negative feedback brings about sustained oscillations of the activity of protein kinases (Kholodenko, 2000). Figure 6 demonstrates that the oscillatory dynamics of the connection strengths in the cascade can be efficiently retrieved by our perturbation approach from time series data on oscillations of protein kinase activities.

CONCLUSIONS

The results presented here demonstrate how unknown gene and signaling networks can be reconstructed from time series measured as responses of intact cells to perturbations. The experimental interventions need not to be specific, but can directly influence many components. Our approach can be applied to a cellular network regardless of its degree of complexity, the presence or the lack of mass flow between nodes and whether or not a mechanistic description is available. At the same time, prior information about a network can be utilized in our method. For instance, if it is known that two particular nodes do not directly influence a given node i,

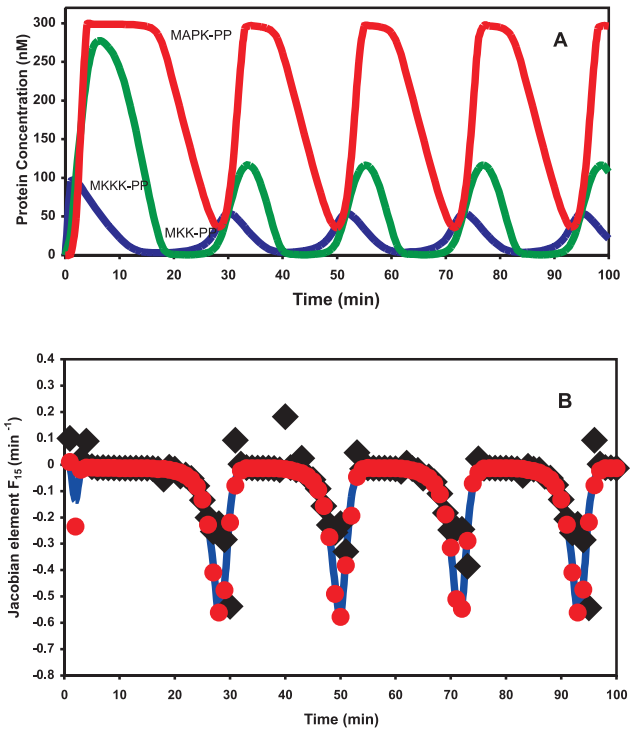

Fig. 6. Oscillatory dynamics of connection strengths is effectively deduced. (**A**) Sustained oscillations in the activities of the MAPK pathway proteins, MAPK–PP, red line; MKK–PP, green; MKKK–PP, blue. (**B**) Sustained oscillations of the strength of a negative feedback. A negative effect of MAPK–PP on MKKK phosphorylation is quantified in terms of the Jacobian element F_{15}. Solid line, the correct 'theoretical' value of F_{15} is plotted as a function of time. Symbols show F_{15} retrieved from the responses to 'experimental' perturbations of 1% (red dot) and 10% (black diamond) magnitudes (see Supplementary Table 4 for further details). Perturbations were repeated every 3 min, and the perturbed trajectory was restarted from the unperturbed solution (time scale corresponds to the unperturbed trajectory).

the corresponding connection strengths (Jacobian elements) are equal to zero. Then, the number of unknown interactions decreases by two, and so does the number of perturbation experiments required to quantify all connections leading to module i. Importantly, our method is fully scalable. In fact, an increase in the node connectivities does not change the number of required perturbation experiments and calculations, which is proportional to the amount of nodes, and it does not involve a combinatorial increase in computations with an increase in the network complexity.

Hierarchical control of cell machinery spreads over metabolome, proteome and genome. In cellular networks, metabolic and lipid transformations and post-translational modification events result in mass flow connections between nodes. Mass flow interactions cause any perturbation of the rate going to or from a particular node to immediately affect other nodes connected by that rate. Prior approaches, which required perturbations affecting each node individually, could not apply to these networks. A distinction of our approach is its ability to reconstruct not only regulatory (information) pathways, but also networks involving mass flow connections. By measuring metabolic, proteomic and gene time series data, the method presented enables us to unveil the signaling circuitry of protein–protein interactions and to infer which metabolites and proteins affect which genes. This strategy may bring to realization the ambitious goal of a quantitative understanding of how the complete metabolic/proteomic/genetic 'supernetwork' of living cells is wired.

ACKNOWLEDGEMENTS

We thank Drs A. Davies and J. Pastorino for critical reading of the manuscript. This work was supported by Grants GM59570, AA08714 and P20-GM64375 from the National Institute of Health. E.S. also acknowledges support from NSF grant CCR-0206789, AFOSR grant F49620-01-1-0063 and a grant from Aventis Pharmaceuticals.

REFERENCES

Arkin,A. and Ross,J. (1995) Statistical construction of chemical reaction mechanisms from measured time-series. *J. Phys. Chem.*, **99**, 970–979.

Arkin,A., Shen,P. and Ross,J. (1997) A test case of correlation metric construction of a reaction pathway from measurements. *Science*, **277**, 1275–1279.

Bagowski,C.P. and Ferrell,J.E.,Jr. (2001) Bistability in the JNK cascade. *Curr. Biol.*, **11**, 1176–1182.

Bhan,A., Galas,D.J. and Dewey,T.G. (2002) A duplication growth model of gene expression networks. *Bioinformatics*, **18**, 1486–1493.

Brown,G.C., Hoek,J.B. and Kholodenko,B.N. (1997) Why do protein kinase cascades have more than one level? *Trends Biochem. Sci.*, **22**, 288.

Bruggeman,F.J., Westerhoff,H.V., Hoek,J.B. and Kholodenko,B.N. (2002) Modular response analysis of cellular regulatory networks. *J. Theor. Biol.*, **218**, 507–520.

Bryson,A.E. and Ho,Y.-C. (1969) *Applied Optimal Control*. Wiley, New York.

Cascante,M., Franco,R. and Canela,E.I. (1989) Use of implicit methods from general sensitivity theory to develop a systematic approach to metabolic control. II. Complex systems. *Math. Biosci.*, **94**, 289–309.

Chang,L. and Karin,M. (2001) Mammalian MAP kinase signalling cascades. *Nature*, **410**, 37–40.

Chevalier,T., Schreiber,T. and Ross,J. (1993) Toward a systematic determination of complex reaction mechanisms. *J. Phys. Chem.*, **97**, 6776–6787.

Claverie,J.M. (1999) Computational methods for the identification of differential and coordinated gene expression. *Hum. Mol. Genet.*, **8**, 1821–1832.

de la Fuente,A., Brazhnik,P. and Mendes,P. (2002) Linking the genes: inferring quantitative gene networks from microarray data. *Trends Genet.*, **18**, 395–398.

D'Haeseleer,P., Liang,S. and Somogyi,R. (2000) Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics*, **16**, 707–726.

Edwards,J.S., Ibarra,R.U. and Palsson,B.O. (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.*, **19**, 125–130.

Ferrell,J.E.,Jr. and Machleder,E.M. (1998) The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. *Science*, **280**, 895–898.

Gat-Viks,I. and Shamir,R. (2003) Chain functions and scoring functions in genetic networks. *Bioinformatics*, **19**, I108–I117.

Hargrove,J.L., Hulsey,M.G. and Beale,E.G. (1991) The kinetics of mammalian gene expression. *Bioessays*, **13**, 667–674.

Husmeier,D. (2003) Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics*, **19**, 2271–2282.

Ideker,T.E., Thorsson,V. and Karp,R.M. (2000) Discovery of regulatory interactions through perturbation: inference and experimental design. *Pac. Symp. Biocomput.*, 305–316.

Ingalls,B.P. and Sauro,H.M. (2003) Sensitivity analysis of stoichiometric networks: an extension of metabolic control analysis to non-steady state trajectories. *J. Theor. Biol.*, **222**, 23–36.

Jayaraman,A., Yarmush,M.L. and Roth,C.M. (2000) Dynamics of gene expression in rat hepatocytes under stress. *Metab. Eng.*, **2**, 239–251.

Kholodenko,B.N. (2000) Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur. J. Biochem.*, **267**, 1583–1588.

Kholodenko,B.N., Demin,O.V., Moehren,G. and Hoek,J.B. (1999) Quantification of short term signaling by the epidermal growth factor receptor. *J. Biol. Chem.*, **274**, 30169–30181.

Kholodenko,B.N., Demin,O.V. and Westerhoff,H.V. (1997a) Control analysis of periodic phenomena in biological systems. *J. Phys. Chem.*, **101**, 2070–2081.

Kholodenko,B.N., Hoek,J.B., Westerhoff,H.V. and Brown,G.C. (1997b) Quantification of information transfer via cellular signal transduction pathways [Erratum (1997) *FEBS Lett.* **419** 150.] *FEBS Lett.*, **414**, 430–434.

Kholodenko,B.N., Kiyatkin,A., Bruggeman,F.J., Sontag,E., Westerhoff,H.V. and Hoek,J.B. (2002) Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proc. Natl Acad. Sci., USA*, **99**, 12841–12846.

Kholodenko,B.N. and Sontag,E.D. (2002) Determination of Functional Network Structure from Local Parameter Dependence Data. *arXiv: physics/0205003*.

Kholodenko,B.N. and Westerhoff,H.V. (1995) The macroworld versus the microworld of biochemical regulation and control. *Trends Biochem. Sci.*, **20**, 52–54.

Kholodenko,B.N., Westerhoff,H.V., Puigjaner,J. and Cascante,M. (1995) Control in channelled pathways. A matrix method calculating the enzyme control coefficients. *Biophys. Chem.*, **53**, 247–258.

Langlois,W.J., Sasaoka,T., Saltiel,A.R. and Olefsky,J.M. (1995) Negative feedback regulation and desensitization of insulin- and epidermal growth factor-stimulated p21ras activation. *J. Biol. Chem.*, **270**, 25320–25323.

Lee,E.B. and Markus,L. (1968) *Foundations of Optimal Control Theory*. Wiley, New York.

Levchenko,A., Bruck,J. and Sternberg,P.W. (2000) Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc. Natl Acad. Sci., USA*, **97**, 5818–5823.

Li,J., Ning,Y., Hedley,W., Saunders,B., Chen,Y., Tindill,N., Hannay,T. and Subramaniam,S. (2002) The Molecule Pages database. *Nature*, **420**, 716–717.

McBride,L.E. and Narendra,K.S. (1965) Optimization of time-varying systems. *IEEE Trans. Autom. Control*, **10**, 289–294.

Moehren,G., Markevich,N., Demin,O., Kiyatkin,A., Goryanin,I., Hoek,J.B. and Kholodenko,B.N. (2002) Temperature dependence of the epidermal growth factor receptor signaling network can be accounted for by a kinetic model. *Biochemistry*, **41**, 306–320.

Niehrs,C. and Pollet,N. (1999) Synexpression groups in eukaryotes. *Nature*, **402**, 483–487.

Reinitz,J. and Sharp,D.H. (1995) Mechanism of eve stripe formation. *Mech. Dev.*, **49**, 133–158.

Repsilber,D., Liljenstrom,H. and Andersson,S.G.E. (2002) Reverse engineering of regulatory networks: simulation studies on a genetic algorithm approach for ranking hypotheses. *Biosystems*, **66**, 31–41.

Saltelli,A., Chan,K. and Scott,E.M. (2000) *Sensitivity Analysis*. Wiley, New York.

Samoilov,M., Arkin,A. and Ross,J. (2001) On the deduction of chemical reaction pathways from measurements of time series of concentrations. *Chaos*, **11**, 108–114.

Schoeberl,B., Eichler-Jonsson,C., Gilles,E.D. and Muller,G. (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.*, **20**, 370–375.

Schuster,S., Fell,D.A. and Dandekar,T. (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.*, **18**, 326–332.

Sontag,E.D. (1993) Neural networks for control. In Trentelman,H.L. and Willems,J.C. (eds), *Essays on Control: Perspectives in the Theory and its Applications*. Birkhauser, Boston, pp. 339–380.

Stark,J., Callard,R. and Hubank,M. (2003) From the top down: towards a predictive biology of signalling networks. *Trends Biotechnol.*, **21**, 290–293.

Tegner,J., Yeung,M.K., Hasty,J. and Collins,J.J. (2003) Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling. *Proc. Natl Acad. Sci., USA*, **100**, 5944–5949.

Tyson,J.J. (1975) Classification of instabilities in chemical reaction systems. *J. Chem. Phys.*, **62**, 1010–1015.

Tyson,J.J., Chen,K. and Novak,B. (2001) Network dynamics and cell physiology. *Nat. Rev. Mol. Cell. Biol.*, **2**, 908–916.

Tyson,J.J., Novak,B., Odell,G.M., Chen,K. and Thron,C.D. (1996) Chemical kinetic theory: understanding cell-cycle regulation. *Trends Biochem. Sci.*, **21**, 89–96.

Vance,W., Arkin,A. and Ross,J. (2002) Determination of causal connectivities of species in reaction networks. *Proc. Natl Acad. Sci., USA*, **99**, 5816–5821.

von Dassow,G., Meir,E., Munro,E.M. and Odell,G.M. (2000) The segment polarity network is a robust developmental module. *Nature*, **406**, 188–192.

Wahde,M. and Hertz,J. (2000) Coarse-grained reverse engineering of genetic regulatory networks. *Biosystems*, **55**, 129–136.

Westerhoff,H.V., Hofmeyr,J.H. and Kholodenko,B.N. (1994) Getting to the inside of cells using metabolic control analysis. *Biophys. Chem.*, **50**, 273–283.

Wiggins,C. and Nemenman,I. (2003) Process pathway inference via time series analysis. *Exp. Mechan.*, **43**, 361–370.

Yeung,M.K., Tegner,J. and Collins,J.J. (2002) Reverse engineering gene networks using singular value decomposition and robust regression. *Proc. Natl Acad. Sci., USA*, **99**, 6163–6168.

Yu,C.F., Liu,Z.X. and Cantley,L.G. (2002) ERK negatively regulates the epidermal growth factor-mediated interaction of Gab1 and the phosphatidylinositol 3-kinase. *J. Biol. Chem.*, **277**, 19382–19388.