Untangling the wires: A strategy to trace functional interactions in signaling and gene networks

Boris N. Kholodenko*†, Anatoly Kiyatkin*, Frank J. Bruggeman‡, Eduardo Sontag§, Hans V. Westerhoff‡, and Jan B. Hoek*

*Department of Pathology, Anatomy and Cell Biology, Thomas Jefferson University, 1020 Locust Street, Philadelphia, PA 19107; ‡Department of Microbial Physiology, Free University, Biocentrum, 1081 HV, Amsterdam, The Netherlands; and §Department of Mathematics, Rutgers University, Piscataway, NJ 08854

Communicated by Rudolf E. Kalman, Swiss Federal Institute of Technology, Gainesville, FL, July 25, 2002 (received for review March 30, 2002)

Emerging technologies have enabled the acquisition of large genomics and proteomics data sets. However, current methodologies for analysis do not permit interpretation of the data in ways that unravel cellular networking. We propose a quantitative method for determining functional interactions in cellular signaling and gene networks. It can be used to explore cell systems at a mechanistic level or applied within a ''modular'' framework, which dramatically decreases the number of variables to be assayed. This method is based on a mathematical derivation that demonstrates how the topology and strength of network connections can be retrieved from experimentally measured network responses to successive perturbations of all modules. Importantly, our analysis can reveal functional interactions even when the components of the system are not all known. Under these circumstances, some connections retrieved by the analysis will not be direct but correspond to the interaction routes through unidentified elements. The method is tested and illustrated by using computer-generated responses of a modeled mitogen-activated protein kinase cascade and gene network.

Advances in high-throughput genomics and proteomics analysis facilitate the monitoring of the expression levels of large gene sets and the activity states of signaling proteins in living cells. The explosive growth in the amount of data calls for the development of novel quantitative approaches for analysis. Thus far, our understanding of cellular signaling and gene networks has been almost exclusively qualitative. Recently, both qualitative and mechanistic mathematical modeling have been applied to better understand network molecular organization and kinetics in quantitative terms (1–8). The mechanistic "bottom-up" approach has the advantage of being readily testable against experiments as a computer replica of cellular networks. However, a major disadvantage of a mechanistic modeling is the large number of molecular processes to be considered, complicated by the fact that values for multiple kinetic parameters are unknown. Moreover, a bottom-up approach inevitably misses the interactions and regulatory feedbacks still awaiting discovery.

As an approach to studying cellular networks, we developed a form of "top-down" sensitivity analysis to quantify the input–output relations and molecular interactions in regulatory networks (9). The control of the input signal over the output target was quantified as the ratio of the input-to-output changes at steady state (called the response coefficient in the limit of infinitesimal changes). The signal may be a hormone, growth factor, neurotransmitter, or experimental intervention (e.g., an inhibitor), and the target process may be the phosphorylation state or activity of a protein, mRNA level, or transcription rate. If a full set of molecular interactions is known, the (*global*) network response to a signal or experimental perturbation can be predicted and expressed in terms of the individual (*local*) responses by using a "map" of network connections (9). However, when detailed information about molecular mechanisms is lacking, a top-down analysis has advantages, because it can be applied experimentally to any cellular network regardless of its degree of complexity (10, 11).

The daunting challenge of understanding the coordinated behavior of numerous molecular interactions can be facilitated by analyzing them within a "modular" framework (12, 13). A complex cellular network can be divided conceptually into reaction groups referred to as functional units or modules. Each module consists of several signaling or gene interactions and performs one or more identifiable tasks. For instance, each of the three tiers of the mitogen-activated protein kinase (MAPK) cascade can be considered as a functional module that involves unphosphorylated, monophosphorylated, and bisphosphorylated forms of a protein kinase and the reactions converting these forms. Modules need not be rigid, and entire MAPK cascades can serve as functional modules in a signaling network that involves growth factor and stress-activated pathways. For gene networks, modules can involve mRNAs of a particular gene or gene cluster with regulatory interaction loops running through metabolic and signaling pathways (14). Modules can be interconnected in multiple ways, many of which may be unknown, even when the network components are identified in genetic and biochemical studies. Fig. 1 illustrates such potential interactions for a three-module cascade and dynamic connections as well as possible unknown components for a gene-expression network.

One of the fundamental problems in cell biology is to infer and quantify interconnections in complex regulatory networks. The present paper proposes a powerful method for attacking such questions, assuming that knowledge of at least some network components is on hand. Specifically, we develop a methodology capable of unraveling and quantifying unknown molecular or modular connections in signaling and gene networks. We demonstrate how, by making systematic perturbations (using inhibitors, activators, changes in external signals, etc.) and measuring global responses only, one can discover a network "interaction map" that can be expressed in terms of module-to-module connection strengths. Importantly, we select experimental interventions that directly perturb single modules, and we apply as many perturbations as there are modules. We illustrate this approach by applying perturbations to model networks and comparing quantitative reconstructions to known interaction maps.

Methods

Fundamentals of Top-Down Regulatory Analysis of Modular Cellular Networks. *Quantitation of a network interaction map.* We conceptually divide a signaling or gene network into modules (m). The degree of complexity of each module is not restricted, and generally a module involves many cellular components (intermediates) connected by chemical reactions (intramodular interactions). We assume that only a single intermediate, referred to as "communicating," serves as the module output (this simplifying assumption is relaxed in *Appendix 1*, which is published as supporting information on the PNAS web site, www.pnas.org). A communicating molecule

Abbreviations: MAPK, mitogen-activated protein kinase; MKK, MAPK kinase; MKKK, MKK kinase; P, monophosphorylated; PP, bisphosphorylated.

†To whom reprint requests should be addressed. E-mail: Boris.Kholodenko@mail.tju.edu.

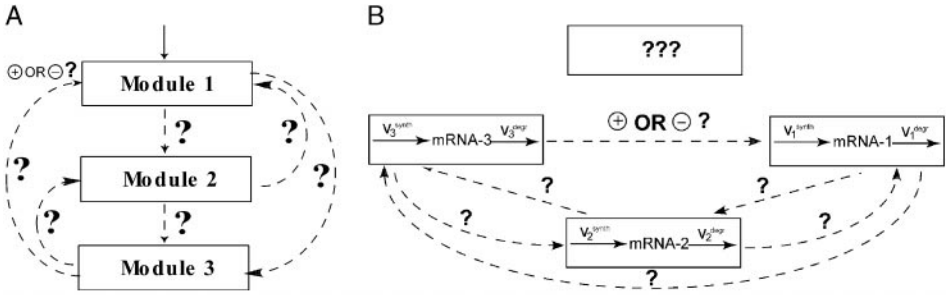

Fig. 1. A three-module cascade (*A*) and a gene network (*B*). The question marks stand for unknown connections and additional network components (e.g., uncharacterized genes), which can influence and in turn be affected by the known components.

may be the active form of a kinase, a second messenger, mRNA, or transcription factor influencing other modules. Thus, communicating intermediates form molecular connections between modules, referred to as intermodular interactions.

A top-down regulatory analysis "black-boxes" the molecular organization of network modules, considering only communicating intermediates (the module outputs). We designate by x_i, $i = 1, \ldots, m$, the activities (concentrations) of communicating intermediates. Following our previous work (9), we quantify intermodular interactions in terms of the fractional changes ($\Delta x_i/x_i$) in the activity of communicating intermediate (x_i) of a particular module (i) brought about by a change in the (output) activity (x_j) of another module (j). Output activities of all other modules (x_k, $k \neq i, j$) are assumed to remain fixed, whereas the affected module (i) is allowed to relax to its steady state. A mathematical definition requires the changes ($\Delta x/x$) to be infinitesimally small, resulting in log-to-log derivatives,

$$r_{ij} = \lim_{\Delta x_j \to 0} \left(\frac{\Delta x_i/x_i}{\Delta x_j/x_j} \right) = \left(\frac{\partial \ln x_i}{\partial \ln x_j} \right)_{\text{module } i \text{ steady state}}, \; i \neq j. \qquad [1]$$

The coefficient r_{ij} is referred to as the local response (coefficient), which quantifies the sensitivity of module i to module j. The term "local" indicates that the response results from immediate interactions between two modules when all other network interactions are held constant. A response coefficient r_{ij} less than 1 means that (small) fractional changes in module j output are attenuated in module i, whereas a response greater than 1 means that these fractional changes are amplified by the factor r_{ij}. A response coefficient of 0 means that module j has no direct effect on module i, whereas a negative response coefficient means inhibition.

Because each module is assumed to have a single communicating intermediate, all interactions between network modules are quantified by $m \cdot (m - 1)$ intermodular response coefficients, r_{ij}. These "connection" coefficients indicate how the network is "wired" and compose the $m \times m$ matrix, $\mathbf{r}$, hereafter referred to as the network interaction map. The ith row of the matrix $\mathbf{r}$ quantifies how module i is affected by each network module through immediate interaction, whereas the jth column of $\mathbf{r}$ measures how module j directly influences each network module. We assign values of -1 to the diagonal elements (r_{ii}) of the matrix $\mathbf{r}$, $r_{ii} = -1, i = 1, \ldots, m$.

Local and global network responses to perturbations. Conceptually considering module i "in isolation" from the network, we determine the local response coefficient (r_{ip_i}) of x_i to a perturbation of parameter p_i, intrinsic to module i as follows: $r_{ip_i} = (\partial \ln x_i / \partial p_i)_{\text{module } i \text{ steady state}}$. When module i is isolated from the network, changes in parameters p_j, influencing other modules (j), have no effect on module i, and therefore the local response of x_i to a perturbation in p_j equals zero. Local responses to perturbations, affecting single modules only, form the diagonal $m \times m$ matrix, $dg\mathbf{r_p}$, with diagonal elements r_{ip_i} and all off-diagonal elements equal to zero.

If following a parameter (p_i) perturbation intrinsic to module i an entire network is allowed to relax, this perturbation not only causes changes in those modules directly affected by module i but also propagates further into the network through interactions between other modules. The resulting time-dependent or stationary responses are called "global" responses of the network. We designate by R_{jp_i} the global response coefficient of module j to a perturbation in p_i and by $\mathbf{R_p}$ the $m \times m$ matrix composed of these coefficients:

$$R_{jp_i} = \left(\frac{d \ln x_j}{dp_i} \right)_{\text{system steady state}}, \; j, i = 1, \ldots, m. \qquad [2]$$

The difference between the local $dg\mathbf{r_p}$ and global $\mathbf{R_p}$ responses is that only module i is allowed to reach the steady state to determine r_{ip_j}, whereas an entire network is permitted to relax to its steady state to measure R_{ip_j}.

Models of Signaling and Gene Networks Used to Test and Illustrate the Proposed Approach. *Computer simulation of MAPK cascade responses to specific perturbations.* MAPK cascades consist of several levels, where the activated kinase at each level phosphorylates the kinase at the next level down the cascade. The three-tiered MAPK cascade comprises MAPK (the terminal level), MAPK kinase (MKK) and MKK kinase (MKKK) (Fig. 2). MAPKs are activated by MKKs, which phosphorylate them at conserved threonine and tyrosine residues. At one level upstream, MKKs themselves are phosphorylated at serine and threonine residues by MKKKs. The kinases of the first level, MKKKs, are activated by incompletely understood mechanisms, involving interactions with the membrane-bound GTPase Ras (in the case of the MKKK Raf-1) and phosphorylation of Raf-1 at a tyrosine residue by an unknown protein kinase (15). Thus, Ras-GTP and unknown membrane kinase(s) function as the input signal that activates MKKK (Raf-1). At each cascade level, protein phosphatases inactivate the corresponding kinases (Fig. 2). Our computational model of the MAPK cascade resembles models

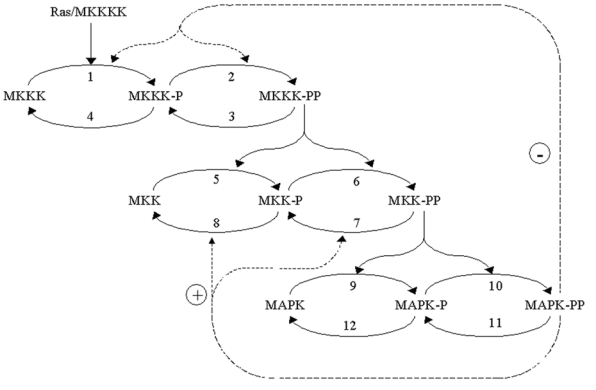

Fig. 2. Kinetic scheme of the MAPK cascade. Feedback effects of MAPK on the rate of MKKK and MKK phosphorylation are shown schematically by dashed lines. MKKKK, MKKK kinase; P and PP, monophosphorylated and bisphosphorylated forms.

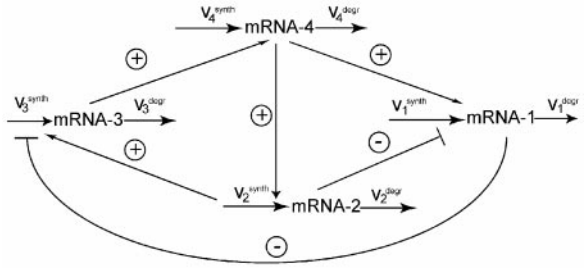

Fig. 3. Kinetic scheme of a four-gene network. Arrows correspond to activation interactions, whereas lines with blunt ends represent inhibitions.

developed previously (16, 17) but includes two negative feedbacks. The first is formed by bisphosphorylated MAPK (MAPK-PP)-mediated inhibition of the MKKK-activating reaction, and the second results from MAPK-PP-induced activation of the MKK phosphatases (18, 19). The kinetic equations, moiety conservation relations, and rate expressions are presented in Table 1, which is published as supporting information on the PNAS web site. We used the model to generate global responses of the cascade communicating intermediates to perturbations, which imitated experimental interventions.

Computer simulation of responses of a gene network to specific perturbations. A kinetic scheme of a four-gene network is depicted in Fig. 3. The level of each mRNA species is determined by the rate of transcription and degradation, $d[\mathrm{mRNA}_i]/dt = v_i^{\mathrm{synth}} - v_i^{\mathrm{degr}}$. Gene interactions result in nonlinear dependences of transcription rates (v_i^{synth}) on other mRNA_j concentrations, which act as communicating intermediates (x_i). The rates are described by the Hill-type equations (1, 3) and presented in Table 2, which is published as supporting information on the PNAS web site. Network responses to perturbations in transcription rates were used to infer functional interactions between genes.

Results

Relation Between the Local and Global Network Responses. Global responses to perturbations can be measured in experiments with intact cellular systems. However, local responses governed by the network interaction map cannot be captured by using intact cells. To measure the kinetics of local interactions between two modules (proteins) directly, they should be isolated from the network. Sometimes the interaction of interest can be reconstituted "*in vitro*," but often only an entire system is accessible experimentally. We are left with the question of how to determine quantitatively the network interaction map if only global responses can be assessed. We demonstrate that by making parameter perturbations to all modules and measuring the global network responses, we can retrieve the unknown interaction map (see *Appendix 2*, which is published as supporting information on the PNAS web site, for the abstract mathematical derivation).

An experimental intervention to perturb a parameter (p_i) intrinsic to module i can employ a specific inhibitor or activator of a reaction within module i, an antisense mRNA affecting the expression level of a protein, or a plasmid changing the rate of transcription. A parameter change, Δp_i, first causes a local perturbation in x_i, which subsequently propagates through intermodular interactions described by the local response coefficients (r_{ij}). After the network has relaxed to a new steady state, the resulting global changes in communicating intermediates (x_k) brought about by a perturbation (Δp_i) are related through

$$\frac{\Delta x_k}{x_k} = \sum_{j \neq k} r_{kj} \cdot \frac{\Delta x_j}{x_j} + r_{kp_i} \cdot \Delta p_i,\ r_{kp_i} = 0,\ \text{if } i \neq k,\ k, i = 1, \ldots, m. \quad [3]$$

Dividing both sides of Eq. **3** by Δp_i and using matrix notations, we arrive at

$$\mathbf{r} \cdot \mathbf{R_p} = -dg\mathbf{r_p}. \quad [4]$$

In intact cells, only the global response matrix, $\mathbf{R_p}$, can be monitored experimentally, whereas neither the network interaction map, $\mathbf{r}$, nor local responses to parameter perturbations, $dg\mathbf{r_p}$, can be measured. We demonstrate how the elements of both matrices, $\mathbf{r}$ and $dg\mathbf{r_p}$, can be calculated by using the matrix $\mathbf{R_p}$. After multiplying both sides of Eq. **4** by the inverse matrix $\mathbf{R_p}^{-1}$, we obtain, $\mathbf{r} = -dg\mathbf{r_p} \cdot \mathbf{R_p}^{-1}$. Because all the diagonal elements of the matrix $\mathbf{r}$ are equal to -1, the elements of the diagonal matrix, $dg\mathbf{r_p}$, are expressed readily in terms of the diagonal elements of the matrix, $\mathbf{R_p}^{-1}$. Designating by $dg(\mathbf{R_p^{-1}})$ the diagonal matrix with diagonal elements $(\mathbf{R_p}^{-1})_{ii}$ and all off-diagonal elements equal to zero, we have $\mathbf{I} = dg\mathbf{r_p} \cdot dg(\mathbf{R_p}^{-1})$, where $\mathbf{I}$ is the identity matrix. By expressing $dg\mathbf{r_p}$ from this equation, we obtain

$$\mathbf{r} = -[dg(\mathbf{R_p}^{-1})]^{-1} \cdot \mathbf{R_p}^{-1}. \quad [5]$$

This final expression gives us the answer: if the (global) responses of a cellular network to perturbations to all modules have been measured, the network interaction map ($\mathbf{r}$) can be retrieved by the inversion of the response matrix ($\mathbf{R_p}$).

Importantly, our method does not require the parameter changes (Δp_i) to be measured or estimated. Instead of response coefficients, one can simply consider the global ($\Delta_i \ln x_j$) fractional changes in communicating intermediates (x_j) caused by a parameter change Δp_i. Accordingly, we redefine the global response matrix, $\mathbf{R_p}$, with coefficients R_{jp_i} to be determined by the global fractional changes brought about by a perturbation Δp_i,

$$R_{jp_i} = (\Delta_i \ln x_j)_{\text{system steady state}},\ i, j = 1, \ldots, m. \quad [6]$$

Here the derivatives, which were considered in Eq. **2**, are substituted by the finite changes (divided by the initial or the mean value). However, the crucial distinction is that according to Eq. **2**, the parameter changes (Δp_i) should be known, whereas Eq. **6** merely considers the differences in intermediates x_j after and before perturbation to determine the global response matrix, $\mathbf{R_p}$. Using Eq. **6**, one obtains exactly the same relationship (Eq. **5**) that expresses the network interaction map in terms of the measured changes in the levels of communicating intermediates without requiring any knowledge about the values of parameter changes. This technique enhances the applicability of the proposed analysis in cases where it is difficult or impossible to quantify the values of parameter perturbations.

Practical Application of the Proposed Methodology. We now outline three steps of experimental applications for the proposed method.

i. Conceptually divide the network under consideration into interacting modules and identify communicating intermediates.

ii. Use an inhibitor or other perturbation that affects a single network module only, e.g., module 1, and measure the difference in the steady-state levels of communicating intermediates before $[x_j^{(0)}]$ and after $[x_j^{(1)}]$ the perturbation. Then, calculate the first column of the matrix $\mathbf{R_p}$ by using, e.g., the central fractional differences defined as the finite difference in the activities divided by the mean value,

$$\Delta_1 \ln x_j \approx 2(x_j^{(1)} - x_j^{(0)})/(x_j^{(1)} + x_j^{(0)}) = 2\frac{(x_j^{(1)}/x_j^{(0)} - 1)}{(x_j^{(1)}/x_j^{(0)} + 1)}. \quad [7]$$

Repeat for remaining network modules ($i = 2, \ldots, m$) by using a perturbation directly affecting that module only, and calculate the remaining columns of the matrix $\mathbf{R_p}$ $(\Delta_i \ln x_1, \ldots, \Delta_i \ln x_m)^T$.

CELL BIOLOGY

-7.4^{a}	-55.5^{b}	6.9^{a}	42.9^{b}	3.7^{a}	23.8^{b}
-7.4^{c}	-55.5^{d}	-7.0^{c}	-46.3^{d}	-3.7^{c}	-25.0^{d}
-6.2^{a}	-44.8^{b}	-3.1^{a}	-21.1^{b}	8.9^{a}	56.6^{b}
-6.2^{c}	-44.8^{d}	3.1^{c}	20.3^{d}	-8.9^{c}	-56.8^{d}
-12.7^{a}	-85.7^{b}	-6.3^{a}	-42.4^{b}	-3.4^{a}	-22.5^{b}
-12.7^{c}	-85.7^{d}	6.3^{c}	39.4^{d}	3.4^{c}	21.8^{d}

Fig. 4. Global fractional responses obtained by simulating experimental perturbations to the MAPK cascade. Four global response matrices ($\mathbf{R_p}$·100, designated by superscripts a–d) were generated by applying the following 12 parameter perturbations. a or b, 10 or 50% decrease, respectively, in: [Ras-GTP], perturbation to module 1; the catalytic activities of steps 5 and 6 (k_5^{cat} and k_6^{cat}; see Table 1, which is published as supporting information on the PNAS web site), module 2; k_9^{cat} and k_{10}^{cat}, module 3. c or d, 10 or 50% decrease, respectively, in: [Ras-GTP], module 1; the maximal activities of steps 7 and 8 (V_7^{max} and V_8^{max}; see Table 1), module 2; V_{11}^{max} and V_{12}^{max}, module 3.

The presentation in terms of the *relative* values given in Eq. **7** may help where quantitation of the absolute activities is difficult, e.g., when Western blotting is used to quantify the relative amount of a protein or determining the ratio of the fluorescence intensities from gene arrays (14).

iii. Apply Eq. **5** to reveal and quantify the network interaction map in terms of the matrix **r** of intermodular (local) response coefficients.

Unraveling the MAPK Cascade Interaction Map: An Illustration. MAPK cascades are widely involved in eukaryotic signal transduction, and these pathways are conserved from yeast to mammals (20). Mammalian cells express at least four different MAPK families including the ERK cascade (which is our primary example) and the c-Jun N-terminal kinase (JNK) and p38 MAPK cascades. In many cell types, MAPK cascades are regulated by multiple feedbacks. For instance, in mammalian cells inhibitory phosphorylation of the GDP/GTP exchange factor, SOS, by ERK provides a mechanism for switching off Ras and, thereby, Raf-1 signaling, creating a negative feedback as shown schematically in Fig. 2 (18). In *Xenopus* oocytes, two MAPK pathways, the p42 MAPK and JNK cascades, appear to be embedded in positive-feedback loops (21, 22). Some regulatory feedbacks are well documented, but the complete interaction map of the MAPK pathways is unknown. For example, it is not understood yet which interactions form positive feedbacks in the JNK cascade (22). Also, both negative- and positive-feedback interactions may differ in various cell types.

Our method may provide a universal tool to analyze the interaction map of MAPK pathways in various cells. To test and illustrate the method, we retrieve the interaction map from computer-generated responses of a kinetic model of the MAPK cascade to perturbations, which simulate experimental interventions. The first step is to identify modules and communicating intermediates based on biological information. We define three MAPK cascade modules that involve different phosphorylation forms of MKKK, MKK, and MAPK, respectively, and the reactions converting these forms (e.g., module 2 includes MKK, MKK-P, and MKK-PP and reactions 5–8, Fig. 2). The bisphosphorylated forms (such as MKK-PP) play the role of communicating intermediates (x_i) influencing other modules. Importantly, the concentration x_i does not determine the concentration of the remaining forms within a module, because two of the three forms are independent variables within a mechanistic description (17). Our method has the advantage of monitoring only communicating intermediates to untangle and quantify the web of intermodular interactions.

A

-1.0	0.0^{a} 0.0^{b} / -0.0^{c} -0.0^{d}	-1.1^{a} -1.0^{b} / -1.1^{c} -1.2^{d}
1.9^{a} 1.8^{b} / 1.8^{c} 1.8^{d}	-1.0	-0.6^{a} -0.6^{b} / -0.6^{c} -0.6^{d}
-0.0^{a} -0.0^{b} / 0.0^{c} -0.0^{d}	2.0^{a} 2.0^{b} / 2.0^{c} 1.9^{d}	-1.0

B

-1	0	-1.1
1.9	-1	-0.6
0	2.0	-1

Fig. 5. Retrieved ''experimental'' interaction maps (*A*) and known ''theoretical'' interaction map (*B*). Four experimental local interaction matrices (**r**) were retrieved from the data shown in Fig. 4. The response coefficients marked by superscripts a–d correspond to perturbations indicated in the legend to Fig. 4.

In the second step, we apply three different perturbations, each affecting a single module. As a perturbation to the first module, we inhibited the input signal by decreasing the Ras-GTP concentration. As relevant interventions to module 2, either the maximal activities of the phosphatase, which dephosphorylates MKK-PP and MKK-P, or the kinase that acts on MKK were inhibited. The different perturbations were applied to illustrate that network interactions to be detected with the method would not depend on what particular molecular processes within a module are affected. Module 3 was perturbed by inhibiting either the maximal activity of the MAPK phosphatase or the kinase. After each perturbation, the MAPK cascade was allowed to reach a new steady state, and the global responses of communicating intermediates were calculated according to Eq. **7**. Fig. 4 presents the matrices $\mathbf{R_p}$ obtained for inhibition values of 10 and 50%. It is convenient to multiply the elements of $\mathbf{R_p}$ by 100, which would correspond to changes in x_i expressed as a percentage of the mean. As follows from Eq. **5**, this multiplication does not change the resulting interaction map, **r**. The 10% perturbation brought about (global) fractional changes of communicating intermediates of less than 13%, whereas a 2-fold inhibition (50%) resulted in up to 86% changes. Perturbations of this magnitude are not justified mathematically, but the simulation results show that our method can handle them well.

Four different matrices ($\mathbf{R_p}$), displayed in Fig. 4, were substituted into Eq. **5** to retrieve the network interaction map (**r**). Notably, both different simulated inhibitors and perturbation values, which brought about widely diverse global changes in communicating intermediates (Fig. 4), resulted in four nearly identical ''experimental'' interaction maps (rounded to the nearest tenth, Fig. 5*A*). Module 1 was found to affect directly module 2, which in turn affects module 3. Both local interactions appear ''ultrasensitive'' with response coefficients r_{21} and r_{32} ranging from 1.8 to 1.9 and 1.9

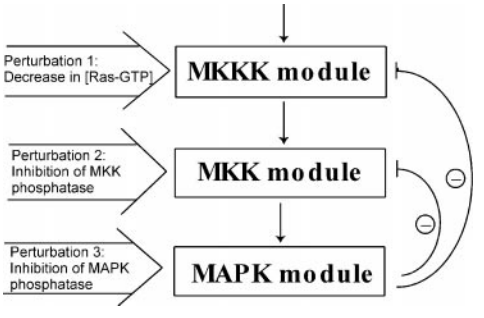

Fig. 6. Revealed interactions between MAPK cascade modules. Responses of every module to perturbations to each single module were measured, and intermodular connections were discovered and quantified by using Eqs. **8** and **10**.

-34.8[a]	39.3[b]	15.9[a]	-22.8[b]	-2.9[a]	2.6[b]	-7.6[a]	11.4[b]
2.2[a]	-3.3[b]	-39.9[a]	46.0[b]	-6.3[a]	5.9[b]	-16.1[a]	27.3[b]
14.7[a]	-20.7[b]	-27.7[a]	42.4[b]	-38.4[a]	43.5[b]	-6.3[a]	14.0[b]
4.2[a]	-6.6[b]	-9.1[a]	10.6[b]	-12.8[a]	10.9[b]	-37.5[a]	44.5[b]

Fig. 7. Global fractional responses obtained by simulating experimental perturbations to the gene network shown in Fig. 3. Two global response matrices ($\mathbf{R_p}$·100, designated by superscripts a and b) were generated by applying the following eight parameter perturbations. a or b, 30% decrease or 50% increase, respectively, in: the maximal activity of the transcription rate v_1^{synth} (V_1^s; see Table 2, which is published as supporting information on the PNAS web site), perturbation to module 1; V_2^s, module 2; V_3^s, module 3; and V_4^s, module 4.

to 2.0, respectively, for different perturbations used. The local interactions, r_{12} and r_{31}, which describe potential effects of modules 2 and 1 on modules 1 and 3, respectively, appear to be zero. Our method unraveled and quantified negative feedbacks from module 3 to modules 1 and 2 (Fig. 6). The response coefficient, r_{13}, ranged from -1.0 to -1.2, and r_{23} was equal to -0.6 for all perturbations used. It is instructive to compare the network interaction map retrieved from "measured" global responses with the correct map, which was calculated according to Eq. **1** for the model example, where molecular interactions were known. As shown in Fig. 5 *A* and *B*, both experimental and theoretical interaction maps appear nearly identical.

Unraveling the Wiring of a Gene Network. Our approach can be applied to untangle gene network interactions (wiring) by carrying out specially designed gene microarray experiments. Gene networks are high-level conceptual representations of interactions between genes (14). These interactions proceed through multiple protein products (e.g., transcription factors) and metabolic intermediates, which are not considered explicitly in the analysis, such that the mRNAs themselves act as communicating intermediates. Fig. 3 illustrates this for a four-gene network. Assuming that no preliminary knowledge is available about gene interactions, we performed two series of four different perturbations to the network as required by the method. The transcription rate of each gene was perturbed independently by decreasing the corresponding maximal activity by 30% or by increasing it by 50%. After each perturbation, the gene network relaxed to its new steady state, and mRNA responses were calculated according to Eq. **7**. The global response matrices ($\mathbf{R_p}$) obtained for perturbation values equal to either 30 or 50% are shown in Fig. 7.

The network interaction map was determined by taking the inverse of $\mathbf{R_p}$ and substituting it into Eq. **5**. Both simulated perturbations, i.e., inhibition or activation, resulted in nearly identical experimental interaction matrices (rounded to the nearest tenth; Fig. 8*A*). All the gene interactions shown in Fig. 2 were retrieved successfully (Fig. 8). Importantly, for both examples considered here, inevitable mistakes related to the substitution of the infinitesimal changes by finite ones did not lead to erroneously predicted interactions, e.g., absent in the network but found by the proposed method. Fig. 8 demonstrates that experimentally obtained network wiring and its quantitation nearly coincides with the known (correct) interaction map for this model system. We conclude that the proposed method can be a powerful tool for unraveling interactions in gene networks.

Discussion

Recently, high-throughput technologies have enabled the acquisition of data on the expression of thousands of genes and the functional state of hundreds of proteins. However, there are no methods capable of providing quantitative interpretations of genomics and proteomics data sets in a manner that unravels the wiring of cellular machinery. This paper proposed a powerful quantitative method to unravel interactions in signaling and gene networks. A dynamic connection between two network components is quantified by the extent to which a small change in one component affects the level or activity of the other, provided all remaining interactions are kept unchanged. The resulting quantifier is known as a response coefficient, which is a convenient and unambiguous measure of the sensitivity of a particular component to a local, direct effect by another component. A network component may be a protein, a gene, or a module involving a number of interacting proteins and genes when considered within a modular framework (12, 13). The present paper demonstrates that monitoring of signaling and gene-expression responses to systematic perturbations is sufficient to infer and quantify signal transduction maps and gene connections.

A series of studies was concerned with the determination of complex reaction mechanisms by experimental evaluation of the Jacobian matrix elements from time-series analysis (23, 24). Methods for the deduction of chemical reaction pathways from measurements of species concentrations were pioneered by Ross and coworkers (25, 26). These studies used ranked time-lagged correlation functions among pairs of chemical species coupled with a multidimensional scaling analysis and heuristic algorithms to deduce a diagram describing the interactions between chemical species. The method that we propose here exploits the modular organization of signaling and gene networks and the absence of

A

-1.0		-0.5[a] -0.6[b]	0.0[a] -0.0[b]	0.4[a] 0.6[b]
-0.0[a] 0.0[b]	-1.0	0.0[a] -0.0[b]	0.4[a] 0.6[b]
-0.4[a] -0.5[b]	0.5[a] 0.7[b]	-1.0	0.0[a] 0.0[b]
0.0[a] -0.0[b]	0.0[a] -0.0[b]	0.3[a] 0.3[b]	-1.0

B

-1	-0.6	0	0.5
0	-1	0	0.5
-0.4	0.6	-1	0
0	0	0.3	-1

Fig. 8. Retrieved experimental interaction maps (*A*) and known theoretical interaction map (*B*). Two experimental local interaction matrices (**r**) were retrieved from the data shown in Fig. 7. The response coefficients marked by superscript a and b correspond to perturbations indicated in the legend to Fig. 7.

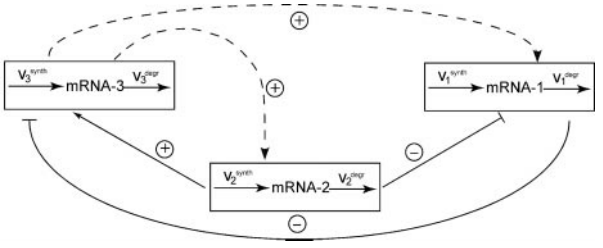

Fig. 9. Retrieved wiring for a three-gene network. Because gene 4 was an unknown component, (quasi)direct effects of gene 3 on genes 1 and 2 were revealed.

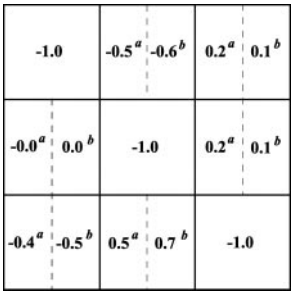

Fig. 10. Interaction maps retrieved from incomplete data. Two experimental local interaction matrices (**r**) were retrieved from the global response matrix corresponding to the first three rows and columns of the matrices shown in Fig. 7. The response coefficients marked by superscripts a or b correspond to perturbations indicated in the legend to Fig. 7.

mass flow between modules (assuming proteins are not significantly sequestered in intermodular interactions). We demonstrated that steady-state measurements of only communicating intermediates (the number of which is much smaller than the number of all independent protein forms) are sufficient to quantify interactions within a modular framework.

Our technique involves a matrix inversion (Eq. **5**). This inversion may give rise to numerical errors if the experimentally measured matrix $\mathbf{R_p}$ is ill-conditioned. Various preconditioning methods might be used to rescale data, but a singular-value decomposition of $\mathbf{R_p}$ can avoid these potential errors by dropping the least meaningful modes. In general, a matrix of lower rank will result, which will constrain the estimates of the local response matrix **r** (which is the normalized Jacobian matrix, see *Appendix 2*, which is published as supporting information on the PNAS web site) to a lower dimensional subspace. A similar approach is based on the observation that a vector that quantifies dynamic connections leading to a particular module (i.e., a row of the matrix **r**) is orthogonal to the linear subspace (H) spanned by vectors composed of measured network responses to perturbations influencing other modules (columns of $\mathbf{R_p}$) (27). If the rank of H decreases because of ill conditioning, additional experiments should be performed by applying different perturbations. Any perturbation directly affecting a single module is appropriate. For instance, in signaling networks one can inhibit or mutate enzymatic activities or change the abundance of a protein operating within a single module. For gene networks, suitable experimental interventions involve the inhibition of a transcription rate or transfection with a plasmid expressing a gene that results in an increase in the mRNA synthesis rate. Applying such perturbations to a model gene network, all gene interactions were unraveled and quantified (Fig. 3). By applying perturbations to the MAPK pathway conceptually partitioned into modules, we detected all existing interactions between modules including the inhibition of module 2 by module 3 (Fig. 6). Mechanistically, this negative feedback occurred as the activation of an enzyme within module 2 (MKK phosphatase) by a communicating intermediate of module 3 (MAPK-PP, Fig. 2). Clearly, the molecular mechanisms cannot be predicted by the method. However, any manifestation of interaction detected by the method can be investigated further mechanistically to advance understanding in molecular terms.

Other biological applications of the method include systems in which some components are unknown or uncharacterized. To illustrate these applications, we revisit the example of a four-gene network, assuming that only three genes (1, 2, and 3) are known (cf. Figs. 1*B* and 3). If we were unaware of the existence of an additional gene (number 4) or simply assumed that this gene did not interact with the system under study, we would bring about perturbations to only three genes and measure the global responses of only those genes. As a result, we would obtain the global response matrix ($\mathbf{R_p}$) corresponding to the first three rows and columns of the matrices presented in Fig. 7. Taking the inverse of this reduced matrix $\mathbf{R_p}$ and substituting into Eq. **5**, we obtain the network interaction map shown in Figs. 9 and 10. We can see that the connections between three known genes, which were identified previously by using perturbations to all four genes (Figs. 3 and 8), also were retrieved by using incomplete information. Importantly, new connections were found for the system with only three identified genes. In the four-gene network, gene 3 directly affected neither gene 1 nor gene 2. However, using incomplete information, we found that gene 3 affects both gene 1 and gene 2. This finding reflects interaction paths from gene 3 through gene 4. The results of our previous work (9) imply that the corresponding responses r_{13} and r_{23} determined for this three-gene network are equal to the (mathematical) products $r_{14}{\cdot}r_{43}$ and $r_{24}{\cdot}r_{43}$, respectively, determined for a four-gene network (cf. Figs. 8 and 10). If it were known that neither the protein product of gene 3 nor proteins interacting with this product could affect gene 1 or 2 directly, this would imply the existence of unidentified gene(s) that perform those interactions. Therefore, the proposed method is able to provide an unbiased analysis to indicate the existence of unknown or uncharacterized components in the system.

We thank J. Pastorino, B. Ingalls, and H. Sauro for stimulating discussions. This work was supported by National Institutes of Health Grants GM59570, AA08714, and P20-GM6437. E.S. also acknowledges support from the BioMaPS Institute at Rutgers University.

1. Smolen, P., Baxter, D. A. & Byrne, J. H. (1998) *Am. J. Physiol.* **274,** C531–C542.
2. Kholodenko, B. N., Demin, O. V., Moehren, G. & Hoek, J. B. (1999) *J. Biol. Chem.* **274,** 30169–30181.
3. von Dassow, G., Meir, E., Munro, E. M. & Odell, G. M. (2000) *Nature (London)* **406,** 188–192.
4. Haugh, J. M., Wells, A. & Lauffenburger, D. A. (2000) *Biotechnol. Bioeng.* **70,** 225–238.
5. Hasty, J., McMillen, D., Isaacs, F. & Collins, J. J. (2001) *Nat. Rev. Genet.* **2,** 268–279.
6. Tyson, J. J., Chen, K. & Novak, B. (2001) *Nat. Rev. Mol. Cell. Biol.* **2,** 908–916.
7. Moehren, G., Markevich, N., Demin, O., Kiyatkin, A., Goryanin, I., Hoek, J. B. & Kholodenko, B. N. (2002) *Biochemistry* **41,** 306–320.
8. Shvartsman, S. Y., Hagan, M. P., Yacoub, A., Dent, P., Wiley, H. S. & Lauffenburger, D. A. (2002) *Am. J. Physiol.* **282,** C545–C559.
9. Kholodenko, B. N., Hoek, J. B., Westerhoff, H. V. & Brown, G. C. (1997) *FEBS Lett.* **414,** 430–434.
10. Brand, M. D. (1996) *J. Theor. Biol.* **182,** 351–360.
11. Krauss, S. & Brand, M. D. (2000) *FASEB J.* **14,** 2581–2588.
12. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. (1999) *Nature (London)* **402,** C47–C52.
13. Lauffenburger, D. A. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 5031–5033.
14. De la Fuente, A., Brazhnik, P. & Mendes, P. (2002) *Trends Genet.* **18,** 395–398.
15. Marais, R. & Marshall, C. J. (1996) *Cancer Surv.* **27,** 101–125.
16. Huang, C. Y. & Ferrell, J. E., Jr. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 10078–10083.
17. Kholodenko, B. N. (2000) *Eur. J. Biochem.* **267,** 1583–1588.
18. Hu, Y. & Bowtell, D. D. (1996) *Oncogene* **12,** 1865–1872.
19. Kolch, W. (2000) *Biochem. J.* **351,** 289–305.
20. Chang, L. & Karin, M. (2001) *Nature (London)* **410,** 37–40.
21. Ferrell, J. E., Jr., & Machleder, E. M. (1998) *Science* **280,** 895–898.
22. Bagowski, C. P. & Ferrell, J. E., Jr. (2001) *Curr. Biol.* **11,** 1176–1182.
23. Tyson, J. J. (1975) *J. Chem. Phys.* **62,** 1010–1015.
24. Chevalier, T., Schreiber, T. & Ross, J. (1993) *J. Phys. Chem.* **97,** 6776–6787.
25. Arkin, A. & Ross, J. (1995) *J. Phys. Chem.* **99,** 970–979.
26. Samoilov, M., Arkin, A. & Ross, J. (2001) *Chaos* **11,** 108–114.
27. Kholodenko, B. N. & Sontag, E. D. (2002) arXiv: physics/0205003.

Supporting information for Kholodenko *et al.* (2002) *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.192442699

Appendix 1: Generalization of the Method for the Case of Several Communicating Intermediates in a Module

Our method is readily generalized for the case when modules have more than one communicating intermediate as the output. In this case, the number of independent perturbations applied to a module should be more than one and equal to the number of communicating intermediates in that module. The resulting expression for the network interaction map becomes slightly more complicated. Instead of the diagonal matrix $[dg(\mathbf{R_p}^{-1})]^{-1}$, the block-diagonal matrix, $\mathbf{B} = [bldg(\mathbf{R_p}^{-1})]^{-1}$, is determined by using the elements of the inverse matrix, $\mathbf{R_p}^{-1}$. For a network with m modules, the matrix $\mathbf{B}$ has m central minors (square blocks) with nonzero elements, whereas all other elements, which do not enter these blocks, are equal to zero. The dimension of each block corresponds to the number of communicating intermediates in the corresponding module. Assuming k such intermediates in module i, the ith principal minor of the block-diagonal matrix $\mathbf{B}$ is determined as follows:

$$\begin{bmatrix} B_{ii} & B_{i,i+1} & \dots & B_{i,i+k} \\ B_{i+1,i} & B_{i+1,i+1} & \dots & B_{i+1,i+k} \\ \dots & \dots & \dots & \dots \\ B_{i+k,i} & B_{i+k,i+1} & \dots & B_{i+k,i+k} \end{bmatrix} = \begin{bmatrix} R_{ii}^{-1} & R_{i,i+1}^{-1} & \dots & R_{i,i+k}^{-1} \\ R_{i+1,i}^{-1} & R_{i+1,i+1}^{-1} & \dots & R_{i+1,i+k}^{-1} \\ \dots & \dots & \dots & \dots \\ R_{i+k,i}^{-1} & R_{i+k,i+1}^{-1} & \dots & R_{i+k,i+k}^{-1} \end{bmatrix}^{-1} . \qquad [1]$$

For the case a single communicating intermediate in each module, the matrix $\mathbf{B}$ is identical to the diagonal matrix $[dg(\mathbf{R_p}^{-1})]^{-1}$ in Eq. **8** of the main text. In the general case, the network interaction map, $\mathbf{r}$, can be obtained by using measured values of system responses (matrix $\mathbf{R_p}$) as follows:

$$\mathbf{r} = -\mathbf{B}\cdot\mathbf{R_p}^{-1} . \qquad [2]$$

Note that the analysis also can be conducted in terms of nonnormalized responses ($\partial x_i/\partial x_j$) defined as the quotients of absolute rather than fractional changes.

In summary, we present an experimental strategy that provides a systematic approach to analyzing functional interactions in complex signaling or gene network systems in a quantitative manner. It offers the potential of generating a complete description of the relevant network interactions and can even identify the contribution of those components that may have escaped identification thus far. It also offers the flexibility of being applicable at different levels of organization by redefining the modules of which the interactions are being explored. For instance, whereas the examples analyzed for illustrative purposes were focused on intracellular signaling and gene expression pathways, the method can be adapted to the analysis of cell–cell interactions (or beyond) by taking whole cells or tissues as single modules and considering hormone and cytokine

signals or similar mediators as communicating intermediates. Hence, it becomes possible to generate a quantitative analysis of a complex biological system at progressively deeper levels by applying the analysis at sequential levels of modular organization.

Supporting information for Kholodenko *et al.* (2002) *Proc. Natl. Acad. Sci. USA,* 10.1073/pnas.192442699

Table 1. Rate equations and parameter values of the MAPK cascade model

Rate equation	Parameter values
$v_1 = \frac{k_1^{cat} \cdot [RasGTP] \cdot [MKKK]}{(K_{11} + [MKKK] + [MKKKP] \cdot K_{11}/K_{12}) \cdot (1 + [MAPKPP]/K_i)}$	k_1^{cat} = 1; [RasGTP] =20; K_{11} = 300; K_{12} = 20; K_i = 100
$v_2 = \frac{k_2^{cat} \cdot [RasGTP] \cdot [MKKKP]}{(K_{11} + [MKKK] + [MKKKP] \cdot K_{11}/K_{12}) \cdot (1 + [MAPKPP]/K_i)}$	k_2^{cat} = 15; $[MKKK]_{total}$ = 200
$v_3 = \frac{V_3^{max} \cdot [MKKKPP]}{(K_{31} + [MKKKPP] + [MKKKP] \cdot K_{31}/K_{32} + [MKKK] \cdot K_{31}/K_{33})}$	V_3^{max} = 18.8; K_{31} = 22; K_{32} = 18; K_{33} = 80
$v_4 = \frac{V_4^{max} \cdot [MKKKP]}{(K_{31} + [MKKKPP] + [MKKKP] \cdot K_{31}/K_{32} + [MKKK] \cdot K_{31}/K_{33})}$	V_4^{max} = 16.4
$v_5 = \frac{k_5^{cat} \cdot [MKK] \cdot [MKKKPP]}{(K_{51} + [MKK] + [MKKP] \cdot K_{51}/K_{52})}$	k_5^{cat} = 1; K_{51} = 300; K_{52} = 20
$v_6 = \frac{k_6^{cat} \cdot [MKKP] \cdot [MKKKPP]}{(K_{51} + [MKK] + [MKKP] \cdot K_{51}/K_{52})}$	k_6^{cat} = 15; $[MKK]_{total}$ = 180
$v_7 = \frac{V_7^{max} \cdot [MKKPP] \cdot (1 + A \cdot [MAPKPP]/K_{mp})}{(K_{71} + [MKKPP] + [MKKP] \cdot K_{71}/K_{72} + [MKK] \cdot K_{71}/K_{73}) \cdot (1 + [MAPKPP]/K_{mp})}$	V_7^{max} = 18.8; K_{71} = 22; K_{72} = 18; K_{73} = 80; A = 5; K_{mp} = 100
$v_8 = \frac{V_8^{max} \cdot [MKKP] \cdot (1 + A \cdot [MAPKPP]/K_{mp})}{(K_{71} + [MKKPP] + [MKKP] \cdot K_{71}/K_{72} + [MKK] \cdot K_{71}/K_{73}) \cdot (1 + [MAPKPP]/K_{mp})}$	V_8^{max} = 16.4
$v_9 = \frac{k_9^{cat} \cdot [MKKPP] \cdot [MAPK]}{(K_{91} + [MAPK] + [MAPKP] \cdot K_{91}/K_{92})}$	k_9^{cat} = 1; K_{91} = 300; K_{92} = 20
$v_{10} = \frac{k_{10}^{cat} \cdot [MKKPP] \cdot [MAPKP]}{(K_{91} + [MAPK] + [MAPKP] \cdot K_{91}/K_{92})}$	k_{10}^{cat} = 15; $[MAPK]_{total}$ = 360
$v_{11} = \frac{V_{11}^{max} \cdot [MAPKPP]}{(K_{111} + [MAPKPP] + [MAPKP] \cdot K_{111}/K_{112} + [MAPK] \cdot K_{111}/K_{113})}$	V_{11}^{max} = 8.4; K_{111} = 22; K_{112} = 18; K_{113} = 80
$v_{12} = \frac{V_{12}^{max} \cdot [MAPKP]}{(K_{111} + [MAPKPP] + [MAPKP] \cdot K_{111}/K_{112} + [MAPK] \cdot K_{111}/K_{113})}$	V_{12}^{max} = 7.3

Concentrations and the Michaelis constants (K_{ij}, i = 1, 3, 5, 7, 9, 11; j = 1, 2, 3; K_{mp}; K_i) are given in nM. The catalytic rate constants (k_i^{cat} , i = 1, 2, 5, 6, 9, 10) and maximal enzyme rates (V_i^{max}, i = 3, 4, 7, 8, 11, 12) are expressed in s^{-1} and $nM \cdot s^{-1}$, respectively. The kinetic equations and moiety conservations derived from the stoichiometry are the following: d[MKKK-P]/dt = $v_1 - v_2 + v_3 - v_4$; d[MKKK-PP]/dt = $v_2 - v_3$; d[MKK-P]/dt = $v_5 - v_6 + v_7 - v_8$; d[MKK-PP]/dt = $v_6 - v_7$; d[MAPK-P]/dt = $v_9 - v_{10} + v_{11} - v_{12}$; d[MAPK-PP]/dt = $v_{10} - v_{11}$; $[MKKK]_{total}$ = [MKKK] + [MKKK-P] + [MKKK-PP]; $[MKK]_{total}$ = [MKK] + [MKK-P] + [MKK-PP]; $[MAPK]_{total}$ = [MAPK] + [MAPK-P] + [MAPK-PP]. MAPK, mitogen-activated protein kinase; MKK, MAPK kinase; MKKK, MKK kinase; P, monophosphorylated form; PP, bisphosphorylated form.

Supporting information for Kholodenko *et al.* (2002) *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.192442699

Table 2. Rate equations and parameter values of the gene network model

Rate equation	Parameter values
$V_1^{synth} = \frac{V_1^s \cdot \left(1 + A_{14} \cdot \left([mRNA_4]/K_{14}^a\right)^{n_{14}}\right)}{\left(1 + \left([mRNA_4]/K_{14}^a\right)^{n_{14}}\right) \cdot \left(1 + \left([mRNA_2]/K_{12}^I\right)^{n_{12}}\right)}$	$V^s_1 = 5$; $A_{14} = 4$; $K^a_{14} = 1.6$; $n_{14} = 2$; $K^I_{12} = 0.5$; $n_{12} = 1$
$V_2^{synth} = \frac{V_2^s \cdot \left(1 + A_{24} \cdot \left([mRNA_4]/K_{24}^a\right)^{n_{24}}\right)}{1 + \left([mRNA_4]/K_{24}^a\right)^{n_{24}}}$	$V^s_2 = 3.5$; $A_{24} = 4$; $K^a_{24} = 1.6$; $n_{24} = 2$
$V_3^{synth} = \frac{V_3^s \cdot \left(1 + A_{32} \cdot \left([mRNA_2]/K_{32}^a\right)^{n_{32}}\right)}{\left(1 + \left([mRNA_2]/K_{32}^a\right)^{n_{32}}\right) \cdot \left(1 + \left([mRNA_1]/K_{31}^I\right)^{n_{31}}\right)}$	$V^s_3 = 3$; $A_{32} = 5$; $K^a_{32} = 1.5$; $n_{32} = 2$; $K^I_{31} = 0.7$; $n_{31} = 1$
$V_4^{synth} = \frac{V_4^s \cdot \left(1 + A_{43} \cdot \left([mRNA_3]/K_{43}^a\right)^{n_{43}}\right)}{1 + \left([mRNA_3]/K_{43}^a\right)^{n_{43}}}$	$V^s_4 = 4$; $A_{43} = 2$; $K^a_{43} = 0.15$; $n_{43} = 2$
$V_1^{degr} = V_1^d \cdot [mRNA_1]/\left(K_1^d + [mRNA_1]\right)$	$V^d_1 = 200$; $K^d_1 = 30$
$V_2^{degr} = V_2^d \cdot [mRNA_2]/\left(K_2^d + [mRNA_2]\right)$	$V^d_2 = 500$; $K^d_2 = 60$
$V_3^{degr} = V_3^d \cdot [mRNA_3]/\left(K_3^d + [mRNA_3]\right)$	$V^d_3 = 150$; $K^d_3 = 10$
$V_4^{degr} = V_4^d \cdot [mRNA_4]/\left(K_4^d + [mRNA_4]\right)$	$V^d_4 = 500$; $K^d_4 = 50$

Concentrations ([mRNA$_i$], $i = 1 - 4$) and Michaelis constants (K^a_i, K^I_i; K^d_i) are given in nM. Maximal enzyme rates (V^s_i, V^d_i) are expressed in nM·s^{-1}. Kinetic equations comprising the model are: $d[\text{mRNA}_i]/dt = V_i^{synth} - V_i^{degr}$.

Supporting information for Kholodenko *et al.* (2002) *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.192442699

Appendix 2: Mathematical Description

In this Appendix we provide an abstract mathematical derivation of the method presented in the paper. This presentation uses the implicit function theorem and matrix notation. A more general version that relaxes the assumption that perturbations considered affect only single modules is published elsewhere (1).

The problem studied in this paper can be reformulated, in an abstract mathematical way, as follows. We consider a dynamic system,

$$dx/dt = f(x,p), \quad x = x_1,...,x_n, \quad p = p_1,...,p_n, \qquad \textbf{[1]}$$

where the vector of variables $x = (x_1,\ldots, x_n)$ and the vector of parameters $p = (p_1,\ldots, p_m)$ belong to open subsets of Euclidean spaces. It is assumed that the system has a stable steady state (x^0, p^0),

$$f(x,p) = 0, \qquad \textbf{[2]}$$

where the Jacobian matrix, **J**, is nonsingular,

$$\mathbf{J} = (\partial f/\partial x). \qquad \textbf{[3]}$$

According to the implicit function theorem, there is a unique vector $x(p)$ solving the set of Eq. **2** in some neighborhood of a particular value p^0. One objective would be to determine the Jacobian (**J**), assuming that one can determine the global response matrix ($\mathbf{R_p}$) experimentally given by Eq. **4** of the main text (hereafter nonnormalized derivatives are used),

$$\mathbf{R_p} = (\partial x/\partial p). \qquad \textbf{[4]}$$

Unfortunately, such an objective is impossible to achieve: the equation $f(x,p) = 0$ is equivalent to the equation $2f(x,p) = 0$, and thus there will be no way to distinguish between $(\partial f/\partial x)$ and $2(\partial f/\partial x)$. Thus, we will restate our objective as that of finding the matrix **r** of the local response coefficients (r_{ij}). Eq. **1** of the main text defines r_{ij} in terms of the fractional changes in x_i brought about by a change in x_j, provided that all other variables (x_k, $k \neq i,j$) remain unperturbed. The coefficients r_{ij} correspond to the elements $(\partial f_i/\partial x_j)$ of the Jacobian matrix (**J**) "normalized" by the diagonal elements, $(\partial f_i/\partial x_i)$, i.e., $r_{ij} = \partial x_i/\partial x_j = -(\partial f_i/\partial x_j)/(\partial f_i/\partial x_i)$, as calculated by the differentiation of the equation, $f_i(x, p) = 0$. In matrix notations,

$$\mathbf{r} = -(dg\mathbf{J})^{-1}\cdot\mathbf{J}. \qquad \textbf{[5]}$$

Eq. **2** allows us to relate the global response matrix ($\mathbf{R_p}$), the Jacobian matrix (**J**), and the matrix of the partial derivatives of functions *f* with respect to the vector of parameter (*p*),

$$\mathbf{R_p} = (\partial x/\partial p) = -(\partial f/\partial x)^{-1}\cdot(\partial f/\partial p) = -(\mathbf{J})^{-1}\cdot(\partial f/\partial p). \quad \textbf{[6]}$$

It is assumed that the matrix $(\partial f/\partial p)$ is nonsingular in the vicinity of the state (x^0, p^0). As explained in the main text, we perturb specific parameters (p_i) that affect only single modules (*i*), which makes the matrix $(\partial f/\partial p)$ diagonal, $(\partial f/\partial p) = dg\mathbf{f_p}$. It is related to the local response matrix, $dg\mathbf{r_p}$, which is defined by Eq. **3** of the main text, as follows,

$$dg\mathbf{r_p}\cdot = -(dg\mathbf{J})^{-1}\cdot dg\mathbf{f_p}. \quad \textbf{[7]}$$

[Eq. **7** is obtained by the differentiation of the equation, $f_i(x,p) = 0$, with respect to p_i assuming that all other variables except x_i (x_k, $k \neq i$) remain fixed]. Using Eqs. **5**–**7**, we find

$$\mathbf{R_p} = -(\mathbf{J})^{-1}\cdot dg\mathbf{f_p} = -(\mathbf{J})^{-1}\cdot(dg\mathbf{J})\cdot(dg\mathbf{J})^{-1}\cdot dg\mathbf{f_p} = -\mathbf{r}^{-1}\cdot dg\mathbf{r_p}. \quad \textbf{[8]}$$

From Eq. **8**, the matrix **r** is expressed as follows,

$$\mathbf{r} = -dg\mathbf{r_p}\cdot\mathbf{R_p}^{-1}. \quad \textbf{[9]}$$

Because all the diagonal elements of the matrix, **r,** are equal to –1 (see Eq. **5**), we can write

$$\mathbf{I} = dg\mathbf{r_p}\cdot dg(\mathbf{R_p}^{-1}), \quad \textbf{[10]}$$

where **I** is the identity matrix, and $dg(\mathbf{R_p}^{-1})$ is the diagonal matrix with diagonal elements $(\mathbf{R_p}^{-1})_{ii}$ and all off-diagonal elements equal to zero. By expressing $dg\mathbf{r_p}$ from this equation, we obtain final Eq. **8** of the main text,

$$\mathbf{r} = -[dg(\mathbf{R_p}^{-1})]^{-1}\cdot\mathbf{R_p}^{-1}. \quad \textbf{[11]}$$

1. Kholodenko, B. N. & Sontag, E. D. (2002) arXiv: physics/0205003.