

Systems biology

Sensitivity, principal component and flux analysis applied to signal transduction: the case of epidermal growth factor mediated signaling

Gang Liu, Mark T. Swihart and Sriram Neelamegham*

Department of Chemical and Biological Engineering, State University of New York at Buffalo, Buffalo, NY 14260, USA

Received on April 16, 2004; revised on September 29, 2004; accepted on October 20, 2004

Advance Access publication November 5, 2004

ABSTRACT

Motivation: Novel high-throughput genomic and proteomic tools are allowing the integration of information from a range of biological assays into a single conceptual framework. This framework is often described as a network of biochemical reactions. We present strategies for the analysis of such networks.

Results: The direct differential method is described for the systematic evaluation of scaled sensitivity coefficients in reaction networks. Principal component analysis, based on an eigenvalue–eigenvector analysis of the scaled sensitivity coefficient matrix, is applied to rank individual reactions in the network based on their effect on system output. When combined with flux analysis, sensitivity analysis allows model reduction or simplification. Using epidermal growth factor (EGF) mediated signaling and trafficking as an example of signal transduction, we demonstrate that sensitivity analysis quantitatively reveals the dependence of dual-phosphorylated extracellular signal-regulated kinase (ERK) concentration on individual reaction rate constants. It predicts that EGF mediated reactions proceed primarily via an Shc-dependent pathway. Further, it suggests that receptor internalization and endosomal signaling are important features regulating signal output only at low EGF dosages and at later times.

Contact: neel@eng.buffalo.edu

Supplemental data: http://www.eng.buffalo.edu/~neel/bio_reaction_network.html

INTRODUCTION

Mathematical modeling and simulation of complex biological processes has gained momentum in the area of quantitative cell biology and bioinformatics. *In silico* models have been developed for various cells including *Escherichia coli* (Edwards *et al.*, 2001), *Saccharomyces cerevisiae* (Vaseghi *et al.*, 1999), *Mycoplasma genitalium*-like cells (Tomita *et al.*, 1999) and erythrocytes (Rapoport *et al.*, 1977; Kauffman *et al.*, 2002). Complex reaction network models for signal transduction (Bhalla and Iyengar, 1999; Lee *et al.*, 2003), apoptosis (Fussenegger *et al.*, 2000) and blood coagulation (Kuharsky and Fogelson, 2001) have also emerged. Gene networks have also been described (Kholodenko *et al.*, 2002; Gardner *et al.*, 2003). While many of the above models require knowledge of reaction stoichiometry or detailed kinetics, models based on Boolean and

self-similarity networks have also been developed. The objective of these studies is to quantitatively understand the features regulating cellular metabolism, biochemical reaction networks and cell response to stimulus.

Strategies developed and applied to date for analyzing biological reaction networks include metabolic flux analysis (Stephanopoulos *et al.*, 1998), metabolic control analysis (Kacser and Burns, 1973; Heinrich and Rapoport, 1974; Fell, 1997), biochemical systems theory (Savageau, 1976), genetic algorithms (Kikuchi *et al.*, 2003), flux balance analysis (Edwards *et al.*, 2001), phase plane analysis and time-lagged correlation analysis (Kauffman *et al.*, 2002), and Bayesian network approaches (Sachs *et al.*, 2002). Among these methods, while metabolic flux analysis, flux balance analysis and metabolic control analysis are appropriate for understanding steady state and quasi-steady-state processes, they are not well suited for transient processes. Biochemical systems theory can be applied to steady-state problems, but estimation of kinetic orders and rate constants under unsteady state conditions is challenging. Phase plane and statistical time-lagged correlation analysis allow recognition of patterns emerging in complex reaction systems by identifying metabolic ‘pools’. While phase plane analysis is easy to visualize, it misses features when the concentrations of two or more species move in tandem with a time delay. Correlation analysis overcomes this limitation, but it is more difficult to interpret. Bayesian methods require vast amounts of experimental data for model development.

To complement the above approaches, we suggest that sensitivity analysis along with principal component analysis (PCA) may provide a simple, systematic and powerful methodology for analysis of biological networks. Such analysis can be applied to study the characteristics of non-steady-state phenomena, such as are typical during signal transduction. Drawing on knowledge in the field of reaction engineering (Varma *et al.*, 1999), in this paper we demonstrate the utility of the direct differential method for evaluation of sensitivity coefficients. Eigenvalue–eigenvector analysis of these results using PCA reveals the set of strongly interacting reactions that together regulate system output. PCA is analogous to the singular value decomposition method used for analysis of gene array data (Alter *et al.*, 2000; Holter *et al.*, 2000). When combined with flux analysis, sensitivity analysis allows model reduction. Model reduction extends sensitivity and PCA by identifying the necessary and sufficient reactions required to simulate any biochemical network.

*To whom correspondence should be addressed.

For illustrative purposes, we choose the model of epidermal growth factor (EGF) mediated signaling, since signaling via the receptor tyrosine kinase family is important for cell growth and differentiation (Starbuck and Lauffenburger, 1992; Kholodenko *et al.*, 1999; Schoeberl *et al.*, 2002; Resat *et al.*, 2003). Of the published models, we choose the model by Schoeberl *et al.* (2002) as a starting point for our calculations since it simultaneously considers receptor trafficking, signaling and degradation. Our analysis reveals the rate-limiting reactions in this network as a function of time and EGF stimulus dose. It compares the relative importance of Shc-dependent and -independent pathways in regulating cell signaling. Further, it illustrates that simulation of gene knock-out models based on sensitivity analysis can allow development of experimentally testable hypothesis. While the example of signal transduction is discussed here, it is apparent that this method can be extended to analysis of gene networks as well.

SYSTEMS AND METHODS

EGF signaling reaction network

The EGF reaction network described by Schoeberl *et al.* (2002) includes 94 reactants and 125 reversible reactions (Supplemental Fig. S1 and Table S1). Of the reactants, 48 are cytoplasmic species, 33 are endosomal species, 3 are degraded species and 10 species are either coated pit proteins or are molecules complexed with these proteins. Among the reactions, 48 are signaling reactions in the cytoplasm, 43 in the endosome, 30 are internalization reactions, 3 are degradation reactions, and 1 reaction mediates EGF receptor (EGFR) synthesis. The entire reaction network begins with the binding of EGF to EGFR, and culminates with the double phosphorylation of ERK to form ERK-PP. This network can be conceptually divided into four modules (Fig. 1, Supplemental Fig. S1). In module A, the EGF-EGFR complex is formed and undergoes dimerization and auto-phosphorylation. In module B, the dimerized EGFR sequentially recruits adaptor proteins Shc (Src homolog and collagen domain protein), Grb2 (growth factor receptor-binding protein 2), and exchange factor Sos (Son of Sevenless homolog protein). This module, which is termed 'Shc-dependent', mediates the conversion of Ras-GDP to form Ras-GTP. In module C, Ras-GTP is formed in a similar fashion as in module B, except that the adaptor protein Shc is not involved. The module is thus 'Shc-independent'. In module D, Raf activated by Ras-GTP triggers the MAP-kinase pathway, which finally produces ERK-PP (dual-phosphorylated extracellular signal-regulated kinase). The phosphorylation of Raf at the start of module D thus forms a critical bridge linking upstream binding and trafficking with cytoplasmic signaling. The above reactions take place at both the cell-surface and endosomal compartments following EGFR internalization. A fraction of internalized reactants also undergo lysosomal degradation. We note that ERK-PP is formed both in the cytoplasm and in endosomes. In our example, the cytoplasmic ERK-PP is considered to be the final network output. The concentration of this species is denoted as $C_{\text{ERK-PP}}$. $C_{\text{MEK-PP}}$ denotes cytoplasmic concentration of double phosphorylated MEK.

The prototypic reaction in this model is second-order with respect to the reacting species. Thus, the velocity (v) of the reaction where A and B combine to form product X with forward and reverse rate constants k^f and k^r respectively is described as $v = k^f C_A C_B - k^r C_X$, where C_A , C_B and C_X denote the concentrations of the reactants and products. In general, the velocity of the i th reaction (v_i) is given by the difference between the i th forward (r_i^f) and reverse reaction rates (r_i^r):

$$v_i = r_i^f - r_i^r = k_i^f \prod_{l=1}^m C_l^{\mu_{il}^f} - k_i^r \prod_{l=1}^m C_l^{\mu_{il}^r} \quad (1)$$

Here, k_i^f and k_i^r are the forward and reverse rate constants for the i th reaction, C_l is the concentration of the l th species, and μ_{il}^f (μ_{il}^r) refers to the forward (reverse) reaction order of the i th reaction with respect to the l th species.

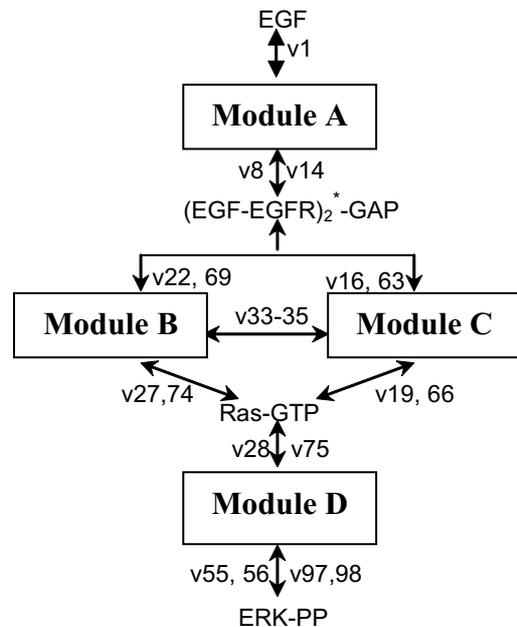


Fig. 1. EGF mediated signaling: Signaling through the EGF reaction network can be divided into four modules, A–D. Signal transduction is initiated with the binding of EGF (system input) to its receptor, and it culminates with the formation of ERK-PP (system output). Double headed arrows depict the reversible nature of the reactions. Key reactions (labelled v8, etc.) that form the connections between the modules are shown in the figure using the nomenclature of Schoeberl *et al.* (2002). Detailed reaction equations, rate constants and initial concentrations are shown in supplemental Figure S1 and Table S1. It is noted that the interconnections between the modules described here is based only on reactions connecting them. Some reactants like Grb2 and Sos are common to modules B and C. Thus, the consumption of these molecules in module B affects the reaction rates in C and vice versa. This provides additional coupling between the modules.

Taken together, the above network is described as an initial value problem, which in matrix form includes m reactants and n reactions:

$$\frac{dC}{dt} = f(k^f, k^r, C) = \alpha^T \cdot v \quad (2)$$

$$C(\text{at } t = 0) = C_0$$

Here, vector v consists of the n individual reaction velocities, C the concentrations of the m reactants, k^f and k^r are $n \times n$ diagonal matrices with the forward and reverse rate constants arranged along the diagonal, and the $m \times n$ matrix α^T contains the stoichiometric coefficients for the reaction network. For our calculations, we modified the Matlab code provided by Schoeberl *et al.* along with their manuscript. Changes made to reaction rate constants and initial concentrations are discussed in Section I of Supplemental Material. Additional codes were also written for sensitivity, principal component and flux distribution analysis as discussed below.

We note some limitations of the model discussed above. First, the EGF signaling pathway analyzed here is relatively simple with regard to receptor degradation. The effects of EGFR ligation on other signaling molecules, like phospholipase-C γ , are also not considered. The complexity of this network can be expanded in the future to account for more biomolecular interactions. Second, the rate constants and initial reactant/receptor concentrations for the simulation may vary with the cell-type being studied and this may affect data interpretation. Third, the reaction network above treats the cells as a 'perfectly mixed' system without accounting for concentration gradients and mass transfer limitations which likely occur in real cases. In

spite of these limitations, the model does represent a reasonable starting point for meaningful biological analysis.

Sensitivity analysis

Sensitivity analysis is a linear perturbation method that is applied to study the effect of infinitesimal changes in system parameters on system variables (Varma *et al.*, 1999). Examples of system parameters include the individual reaction rate constants (k_j^f and k_j^r), the pathway structure and the initial conditions of the system. System variables include individual species concentration (C_i) and reaction velocity (v_i). In our case, the system variable considered is C_i and system parameter is either k_j^f or k_j^r . Sensitivity coefficients are derived for forward and reverse rate constants independently. For forward rates, $Z_{ij}^f \equiv \partial C_i / \partial k_j^f \cdot \mathbf{Z}^f$, the $m \times n$ matrix containing the sensitivity coefficients for the forward reaction rates, can be derived by solving the initial value problem in Equation (3a) (detailed derivation provided in Supplemental Material, Section II):

$$\frac{d\mathbf{Z}^f}{dt} = \boldsymbol{\alpha}^T \cdot \left[\mathbf{r}^f \cdot \mathbf{k}^{f-1} + (\mathbf{r}^f \cdot \boldsymbol{\mu}^f \cdot \mathbf{C}^{-1} - \mathbf{r}^r \cdot \boldsymbol{\mu}^r \cdot \mathbf{C}^{-1}) \cdot \mathbf{Z}^f \right] \quad (3a)$$

with initial condition \mathbf{Z}^f (at $t = 0$) = 0. Similarly for the reverse rate constant, it can be shown that

$$\frac{d\mathbf{Z}^r}{dt} = \boldsymbol{\alpha}^T \cdot \left[-\mathbf{r}^r \cdot \mathbf{k}^{r-1} + (\mathbf{r}^f \cdot \boldsymbol{\mu}^f \cdot \mathbf{C}^{-1} - \mathbf{r}^r \cdot \boldsymbol{\mu}^r \cdot \mathbf{C}^{-1}) \cdot \mathbf{Z}^r \right] \quad (3b)$$

with initial condition \mathbf{Z}^r (at $t = 0$) = 0. Here \mathbf{r}^f and \mathbf{r}^r are $n \times n$ diagonal matrices with the forward and reverse rates as the diagonal elements. \mathbf{C}^{-1} is an $m \times m$ diagonal matrix with $1/C_i$ arranged along the diagonal. $\boldsymbol{\mu}^f$ and $\boldsymbol{\mu}^r$ are $n \times m$ reaction order matrices for the forward and reverse reactions respectively.

The Matlab ODE15s function was used to solve Equations (2) and (3) simultaneously. Once Z_{ij}^f and Z_{ij}^r were obtained, we normalized the result to obtain the scaled sensitivity coefficients, W_{ij} [Equation (4)]. These coefficients are dimensionless and they vary with time and stimulus dose. In order to keep the presentation of our results simple, we typically plot the maximum or minimum W_{ij} in a given time interval.

$$W_{ij} \equiv \frac{k_j^f}{C_i} \frac{\partial C_i}{\partial k_j^f} \quad \text{or} \quad \frac{k_j^r}{C_i} \frac{\partial C_i}{\partial k_j^r} \quad (4)$$

All our calculations were performed in two steps. In the first step, the reactants in the system were equilibrated for long times by setting the EGF concentration to zero. This resulted in a decrease in the number of EGFR per cell from 50000 to 36888, and the formation of complexes among the adaptor proteins. In support of this equilibration step, others have also shown that Sos forms a stable complex with Grb2 in resting cells even in the absence of EGF stimulation (Sastry *et al.*, 1995). Following equilibration, EGF was added at a fixed dosage and the effect of this stimulation was studied. Here, we only provide results for the second step.

By default, in this paper, C_i corresponds to the concentration of ERK-PP in the cellular cytoplasm ($C_{\text{ERK-PP}}$, species 59 in Supplemental Fig. S1) since this is the output of the reaction network. In some cases, to validate the direct differential method, finite perturbations of 1, 5 or 50%, were applied to the rate constant. The effect of this change on network output signal was quantified in terms of % change in $C_{\text{ERK-PP}}$ ($= 100 * [C_{\text{ERK-PP, perturbed}} - C_{\text{ERK-PP, original}}] / C_{\text{ERK-PP, original}}$).

Principal component analysis

Sensitivity analysis results in vast amounts of W_{ij} data. The number of sensitivity coefficients studied depends on the choice of system variables and system parameters. These coefficients also vary with both time and stimulus dosage. Manual analysis of W_{ij} data can thus be complex and tedious. With the objective of addressing this issue, PCA is applied to quantitatively weight the effects of the above features (i.e. time, stimulus dosage and choice of system variable) on the reaction network. Briefly, this method involves the construction of a scaled sensitivity coefficient matrix \mathbf{S} (and its transpose

\mathbf{S}^T) whose elements are W_{ij} derived from the sensitivity analysis. Detailed explanation is provided in Section III (Supplemental Material).

The product matrix $\mathbf{S}^T \mathbf{S}$ can be diagonalized using its eigenvalues and eigenvectors as shown in Equation (5), where the individual eigenvalues (λ_l) form the diagonal elements of the diagonal matrix $\boldsymbol{\Lambda}$, and \mathbf{U} denotes the matrix whose columns are the normalized eigenvectors. u_{jl} represents an element of \mathbf{U} .

$$\mathbf{S}^T \mathbf{S} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T \quad (5)$$

In such an analysis, the eigenvalue provides an absolute measure of the significance of some part of the biological system that is composed of strongly coupled reactions. Each eigenvector is a linear combination of reactions, and the relative magnitude of the elements of each eigenvector measures the relative importance of each reaction for the corresponding eigenvalue. In the current paper, we use the PCA parameter e_j ($= \sum_{l=1}^{2n} \lambda_l u_{jl} / \sum_{l=1}^{2n} \lambda_l$) as a measure of the importance of the j th reaction. Taken together, eigenvalues and eigenvectors of $\mathbf{S}^T \mathbf{S}$ evaluated using PCA provide a measure of the significance of individual reactions.

Model reduction

With the goal of devising a systematic method for eliminating reactions that do not contribute significantly to network output, we combined flux analysis with sensitivity analysis. The rationale for this is that reactions deleted during model reduction must not only have low W_{ij} but also low flux. As an example, if a sequence of reactions occurs in series and one reaction is rate limiting, all of these reactions will have comparable flux but only one will have a large value of W_{ij} . Clearly all the reactions must be retained in the reduced model since we would otherwise break the series of reactions.

For any given EGF concentration, model reduction was performed independently in two distinct time-scales or phases: One in the first several minutes during which $C_{\text{ERK-PP}}$ increased (Phase I), and the next when this species concentration decreased (Phase II). These phases are described in detail in the next section. For model reduction, in any phase, the mean reaction flux \overline{F}_j ($= \left| \int_{t_1}^{t_2} v_j dt / (t_2 - t_1) \right|$ with units molecules/cell/min) was calculated for each of the reactions over the time interval from t_1 to t_2 . This time interval corresponds to either Phase I ($t_1 = 0$, $t_2 = \text{end of Phase I}$) or Phase II ($t_1 = \text{end of Phase I}$, $t_2 = \text{end of Phase II}$). Simultaneously, the maximum absolute value of the scaled sensitivity coefficient was determined for all the reactions in this time interval. This maximum value was denoted $|W_{ij}|_{\text{max}}$. In the following step, a parameter $|W_{\text{crit}}|$ was chosen for the entire network based on a fitted 'reduction factor', ε and $|W_{ij}|_{\text{max}}$:

$$|W_{\text{crit}}| = \varepsilon \cdot |W_{ij}|_{\text{max}} \quad (6)$$

All reactions with absolute value of scaled sensitivity coefficient greater than $|W_{\text{crit}}|$ were classified to be 'essential reactions' and the remaining was termed 'non-essential'. These essential reactions were identified for each module (A–D); some of the non-essential reactions that had low flux were eliminated during model reduction as follows. In some cases, we observed that none of the reactions in a given module were essential, in which case all reactions in that module were eliminated during model reduction. For modules that had essential reactions, we identified which essential reaction in that module had the lowest mean reaction flux. In this module 90% of this lowest flux value was termed the 'critical flux', F_{crit} according to

$$F_{\text{crit}} = 0.9(\text{smallest } \overline{F}_j \text{ among essential reactions}) \quad (7)$$

All non-essential reactions in the module with absolute flux less than F_{crit} were then eliminated while all other reactions were retained in the reduced model. The degree of model reduction is thus controlled by two parameters, primarily by the reduction factor ε and also by the factor of 0.9 in Equation (7).

For EGF network reduction, calculations for Phases I and II as described above were performed over a range of EGF stimulus dosages. In the final reduced model, only reactions that were eliminated in both phases and for all EGF dosages were deleted.

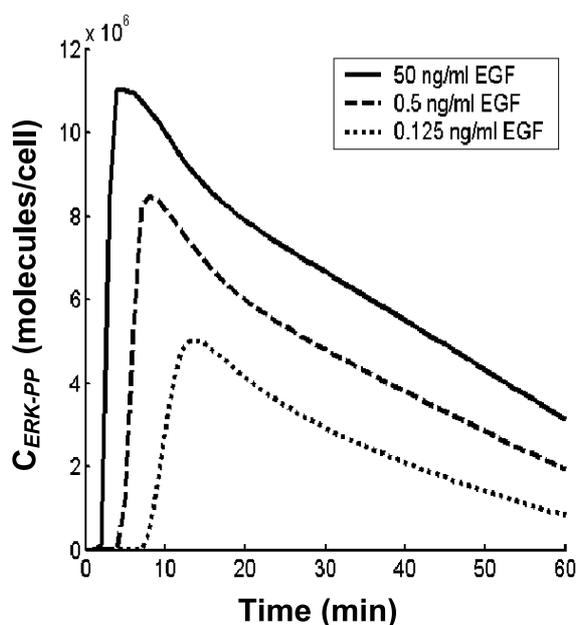


Fig. 2. Time course of ERK-PP: Simulations were carried out at three different EGF stimulus dosages: 50, 0.5 and 0.125 ng/ml. Signal output was monitored in terms of cytoplasmic ERK-PP concentration. Signaling can be conceptually divided into two phases: In Phase I, ERK-PP concentration increases and in Phase II it decreases. The transition between the two phases occurs later for lower EGF concentrations. This transition takes place at 4 min for a stimulus dosage of 50 ng/ml, 8 min for 0.5 ng/ml and 14 min for 0.125 ng/ml.

IMPLEMENTATION AND RESULTS

Time dependent evolution of scaled sensitivity coefficient

Figure 2 illustrates the output signal (i.e. concentration of cytoplasmic ERK-PP, C_{ERK-PP}) following EGF stimulus under our simulation conditions. As seen, a 100-fold reduction in EGF concentration from 50 to 0.5 ng/ml decreases peak output by $\sim 25\%$. Further 4-fold reduction in stimulus to 0.125 ng/ml halves the network output. The time point where ERK-PP concentration is highest, which marks the transition from Phase I to Phase II, is smaller for the higher EGF concentrations. This transition point ranges from 4 min for an EGF dosage of 50 ng/ml to 14 min for 0.125 ng/ml.

Figure 3A–D, derived using the direct differential scheme [Equation (3)], demonstrates that the scaled sensitivity coefficients (W_{ij}) evolve with time in different ways depending on the reaction being considered. The forward reactions illustrated here describe the degradation of internalized EGFR (k_{60}^f), the binding of EGF to EGFR (k_1^f), the dephosphorylation of MEK-PP (k_{49}^f) and the phosphorylation of MEK to form MEK-P (k_{44}^f). While W_{ij} remains zero for some reactions (panel A), it either decreases (panel C) or increases (panel D) in other cases. Negative W_{ij} implies that the signal output decreases with increasing rate constant and vice versa. $|W_{ij}| > 1$ suggests that changes in reaction rate and corresponding reactant concentrations in one region may have an amplified effect on signal output. We note that detailed analysis of signal amplification and dampening in protein kinase mediated signal transduction is discussed elsewhere (Heinrich *et al.*, 2002). While some reactions apparently only affect either Phase I (panel B) or Phase II (panel C),

others display non-zero sensitivity coefficients in both phases (panel D).

Figure 3E–H provide validation of the direct differential analysis scheme. In these panels, finite perturbations of 1, 5 or 50% were applied to individual reaction rate constants, and the effect of this change on cytoplasmic ERK-PP concentration was monitored. Upon comparing Figure 3A with Figure 3E, it is apparent that reactions with zero sensitivity coefficient have no effect on signal output following perturbation. Similarly, if the peak of scaled sensitivity coefficient is ~ 1 (as in Fig. 3B), a 1% change in k_1^f results in a corresponding $\sim 1\%$ change in signal output (Fig. 3F). Larger perturbations of 5 and 50% result in corresponding larger increases in signal output, though the relationship is not strictly linear. Increasing reaction rate constants in cases with negative sensitivity coefficients decreases output signal (Fig. 3C and 3G). Similarly, non-zero sensitivity coefficients in both Phase I and Phase II (Fig. 3D and 3H) affect system output at all times.

Identification of key reactions using sensitivity analysis

Figure 4 shows the maximum and minimum scaled sensitivity coefficients for all reactions in the system for an EGF stimulus dosage of 50 ng/ml. To obtain this data, plots like Figure 3 were generated for each forward and reverse rate, and the maximum and minimum W_{ij} was determined in a given time interval. Panel A presents data for Phase I (0–4 min), while panel B presents the same information for Phase II (beyond 4 min). Data for other EGF concentrations are presented in supplemental material (Supplemental Fig. S2A–F).

As seen in Figure 4A, in module A, the binding of EGF to EGFR (reaction 1) and binding of GAP (GTPase activating protein) to autophosphorylated dimerized EGFR (reaction 8) are key steps controlling the output signal. Of the reactions in module B, reactions 26 and 27, which control the formation of Ras-GTP in module B, have larger W_{ij} compared to other reactions in this module. The reactants in module C have low sensitivity coefficients suggesting that EGF signaling proceeds primarily via an Shc-dependent rather than Shc-independent fashion. In support of this proposition, we note that at the point where the flux from module A bifurcates into module B and C (Fig. 1), the flux from reaction 8 (module A) to reaction 22 (Shc-dependent pathway, module B) is 100–3000 fold greater than that to reaction 16 (Shc-independent pathway, module C) over the range of our simulation conditions. Many of the reactions in module D exhibit high sensitivity coefficients. As examples, reactions 45 and 46, which mediate single and dual phosphorylation of MEK, respectively, exhibit high W_{ij} . Similarly, reactions 28 and 42, which control the phosphorylation and dephosphorylation of Raf, have high W_{ij} . This confirms the central importance of Raf in linking upstream binding events with the MAPK pathway (Morrison and Cutler, 1997). Finally, of the internalization reactions, only 118 and 121, which control internalization via the coated-pit pathway, are relevant in Phase I of the signaling process.

The reactions that affect system output in Phase II (Fig. 4B) differ from those in Phase I. First, it is observed that receptor binding and autophosphorylation (reactions 1 and 8) in module A are relatively unimportant in Phase II, since signaling following EGF binding has passed downstream of this module in Phase II. Second, while reactions that phosphorylate MAP kinase proteins are important during the first phase, it is phosphatase activity that peaks in the second phase. For example, while the sensitivity coefficients of phosphorylating reactions like 45 and 46 are high during Phase I (Fig. 4A),

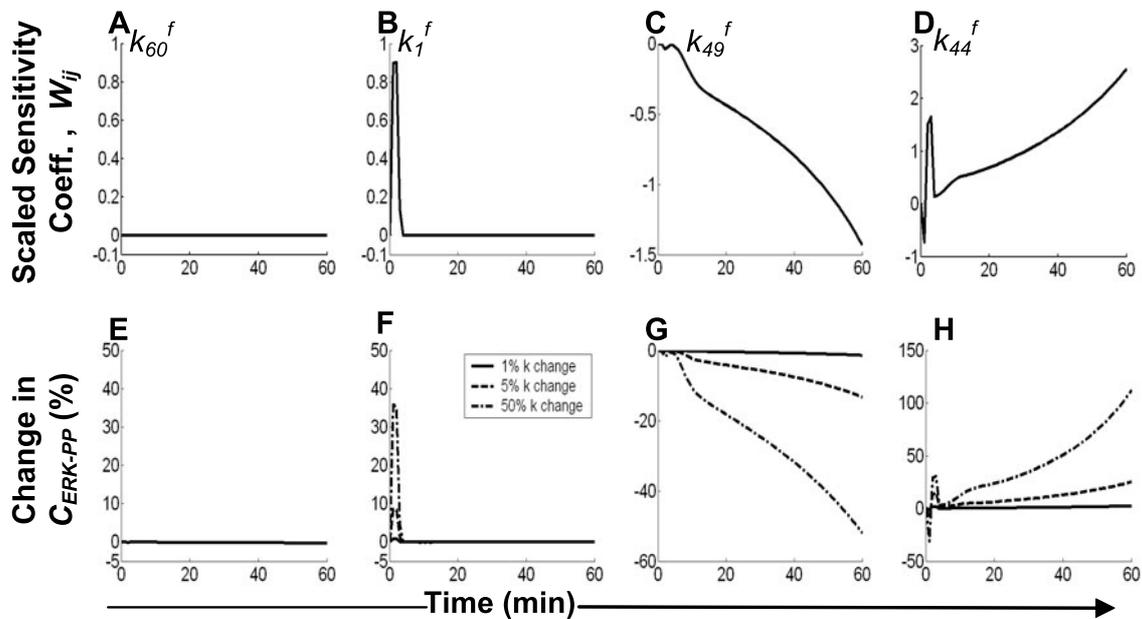


Fig. 3. Time-dependent evolution of scaled sensitivity coefficients: Panels A–D illustrate temporal changes in W_{ij} for four forward rate constants based on the direct differential method. The EGF concentration was 50 ng/ml. Below each of these panels are corresponding figures where the same rate constants were perturbed by a finite amount (1, 5 or 50%), and the effect on ERK-PP signal output was monitored. Panels A and E correspond to k_{60}^f , B and F to k_1^f , C and G to k_{49}^f , and D and H to k_{44}^f .

W_{ij} for phosphatase reactions (numbers 48–51) are relatively high in Phase 2 (Fig. 4B). This result is expected since the ERK-PP concentration increases in the first phase due to phosphorylation and decreases in the second phase due to dephosphorylation. Finally, while endosomal signaling is not important in Phase I, the contribution of these mechanisms to output signal increases in Phase II. In both Figure 4A and 4B, it is primarily the forward, rather than the reverse, rate constants that exhibit high sensitivity coefficients.

EGF–EGFR binding kinetics is important at low stimulus dosage

We performed sensitivity analysis over a range of EGF dosages in order to determine if the identity of the key reactions affecting output signal is altered with stimulus dose. We observed, prominently, that endosomal signaling becomes important at low EGF concentrations since several of the reactions between numbers 63 and 101 exhibited non-zero W_{ij} only at low concentrations (Supplemental Fig. S2). Simultaneously, besides 118 and 121, several other internalization reactions also display non-zero sensitivity coefficients at lower EGF doses (Supplemental Fig. S2). Among the cytoplasmic reactions, the sensitivity coefficient of the EGF binding reaction (number 1) and Raf de-phosphorylation reaction (number 42) are augmented at the lower doses, suggesting that these are important steps regulating EGF signaling (Fig. 5). Finally, we observed that some reactions, in particular the reaction regulating the binding of GAP with autophosphorylated dimerized EGFR (number 8) and the reaction controlling Ras-GTP formation in module B (number 27), exhibited higher W_{ij} at an intermediate concentration (Fig. 5). The behavior of these reactions is analogous to the ultra-sensitivity behavior predicted by others for the MAP kinase pathway (Chi-ying and Ferrell, 1996), though these authors perform their analysis for the steady-state case.

Overall, changing initial EGF concentration does not affect the key regulating reactions in the EGF signaling cascade, though endosomal signaling and internalization processes that did not play a role at high concentrations do make significant contributions at lower stimulus doses and longer times.

PCA complements and extends sensitivity analysis

Sensitivity analysis data results can be weighted using PCA. Key reactions that contribute to this function can then be determined using eigenvalue–eigenvector analysis using Equation (5) (Fig. 6). In this figure, reactions with higher values of PCA parameter, e_j , contribute more markedly to system output. Upon comparison of Figure 6 with data in Figure 4 and Supplemental Figure S2, we observe that the single plot of Figure 6 efficiently captures the essential features of the EGF pathway for the range of concentrations and time scales depicted in the other multi-panel plots.

Model reduction by combining sensitivity analysis with flux analysis

Model reduction involves elimination of reactions that do not contribute significantly to the final output. We used a combination of flux and sensitivity analysis for model reduction as discussed in Methods and illustrated in Figure 7. According to this scheme, all eliminated reactions satisfied two criteria: (i) they exhibited low sensitivity coefficients and (ii) the flux through these reactions was low. These two analysis schemes were combined since sensitivity analysis alone may result in erroneous deletion of reactions with high flux but low sensitivity coefficients; such reactions often occur as intermediates in a sequence of consecutive reactions. Flux analysis alone also does not capture the critical rate controlling features of the reaction network.

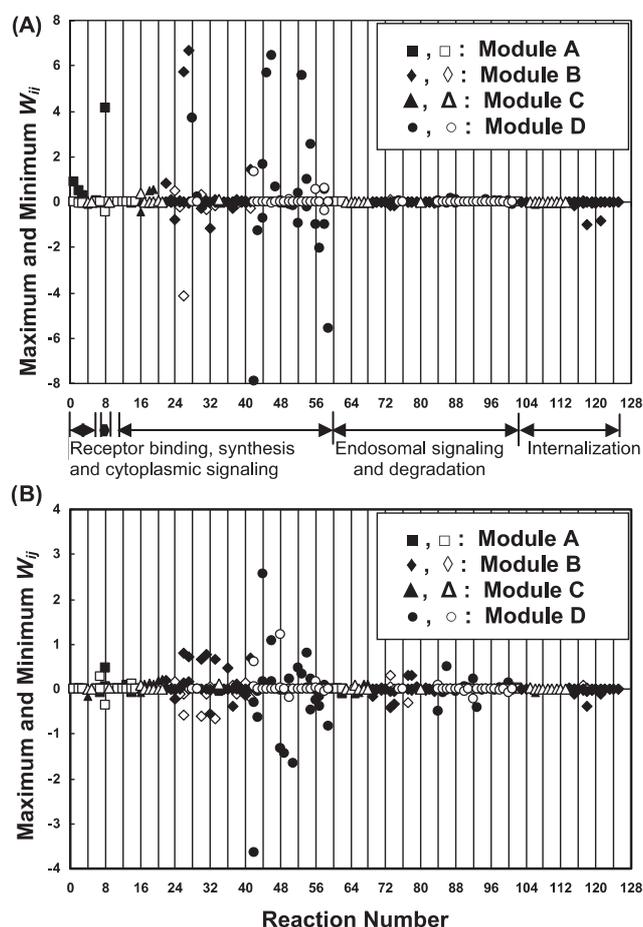


Fig. 4. Sensitivity analysis applied to determine key reactions: Maximum and minimum sensitivity coefficients (W_{ij}) are shown for all 125 reactions in the signaling cascade in Phase I (panel A) and Phase II (panel B) for an EGF concentration of 50 ng/ml. Squares, diamonds, triangles and circles depict W_{ij} in modules A, B, C and D respectively. Filled symbols are used for W_{ij} associated with forward rate constants (k^f), and open symbols for those associated with reverse rate constants (k^r). Reactions corresponding to receptor binding, signaling, degradation and internalization are marked between panels A and B. Reactions 6 and 7 correspond to internalization of EGF receptor or its complex via the smooth-pit pathway, and reactions 10–12 and 14 correspond to signaling via these internalized molecules.

Figure 7 illustrates the relationship between the sensitivity coefficient of individual reactions and mean reaction flux (\bar{F}_j). For generating this plot, calculations were performed for Phase II, the reduction factor ε was set to 0.07 and the EGF concentration was 0.125 ng/ml. As seen, there is no direct relationship between sensitivity coefficient and reaction flux. Rectangles drawn for modules A, B and D in this figure enclose reactions in each module that were not eliminated during model reduction. The left-hand edge of each rectangle corresponds to the critical flux (F_{crit}) for that module, while the right-hand edge corresponds to the reaction with the greatest flux in that module. The upper and lower bounds of the rectangle correspond to the greatest and least scaled sensitivity coefficients in that module. Reactions in module C are not considered during this analysis since the absolute value of the scaled sensitivity coefficient for all reactions in this module is less than $|W_{crit}|$. In order to simplify the

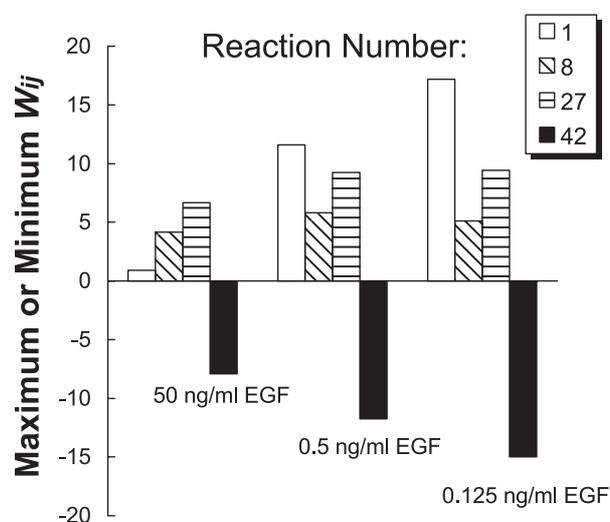


Fig. 5. Effect of EGF dosage on scaled sensitivity coefficients: Detailed analysis of W_{ij} for a range of EGF dosages is presented in Supplemental Figure S2. Selected data from this analysis are presented here for reactions 1, 8, 27 and 42. The figure illustrates that some W_{ij} either increase (reaction 1) or decrease (reaction 42) monotonically with decreasing EGF dosage. Others (reactions 8 and 27) peak at an intermediate EGF concentration.

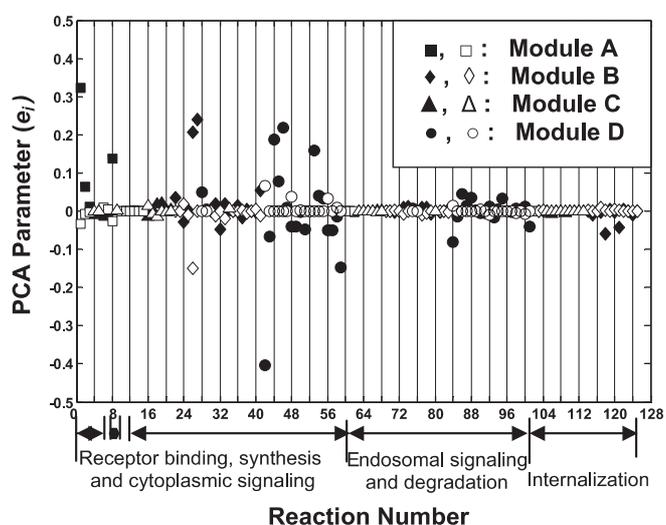


Fig. 6. PCA analysis: A 183×250 sensitivity coefficient matrix (S) was constructed by accounting for W_{ij} at three EGF dosages (50, 0.5 and 0.125 ng/ml) and 61 time points (0–60 min at 1 min intervals). The system parameter consisted of the 250 reaction rate constants that describe the network's 125 reversible reactions. The only system variable considered was C_{ERK-pp} . Eigenvalues and eigenvectors for the product matrix $S^T S$ (250×250) were compared. These were applied to quantify the PCA parameter, e_j , for each of the 250 reactions as described in Methods. The Symbol notations used for this figure are identical to those in Figure 4.

EGF reaction network, analyses similar to that illustrated in Figure 7 were performed for six different conditions: for Phase I and Phase II, and for each of the EGF dosages (50, 0.5 and 0.125 ng/ml). Only reactions that were eliminated under all six conditions were deleted from the reaction network to give the reduced model.

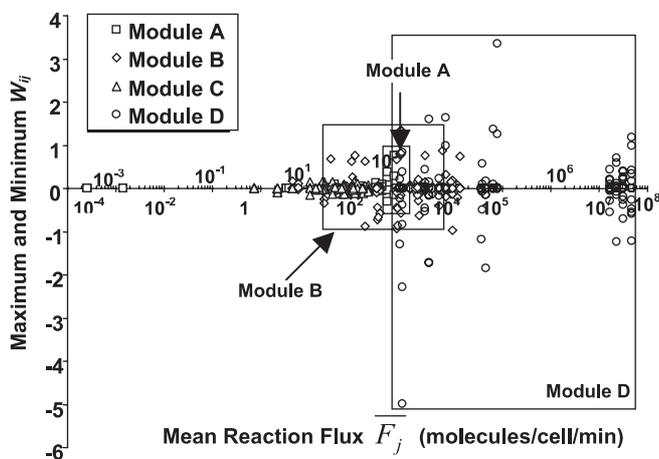


Fig. 7. Sensitivity and flux analysis: Peak scaled sensitivity coefficients (either maximum or minimum) and corresponding flux are plotted for all 125 reactions in modules A (squares), B (diamonds), C (triangles) and D (circles) for EGF stimulus concentrations of 0.125 ng/ml during Phase II of the signaling process. As shown, the fluxes for reactions in module D are greater than in other modules. This demonstrates that signal input is amplified in this biochemical network. No direct relationship is observed between the sensitivity coefficient and the flux of individual reactions. Rectangles are drawn based on our model reduction strategy ($\epsilon = 0.07$) for rate-controlling reactions in modules A, B and D. Only reactions within these enclosures are included in the reduced model shown in Supplemental Figure S3. All reactions in module C were eliminated since the flux and sensitivity coefficients of all reactions in this module were low.

The output signal from the reduced model was compared with the results of the original model for varying values of the reduction factor ϵ at a fixed EGF concentration of 50 ng/ml (Fig. 8A). When ϵ equals zero, all reactions are considered and the reduced model is identical to the original model with 85 reactions in the first three modules (A–C), and 40 in module D. Upon varying ϵ we observed that none of the reactions in module D were eliminated over the range of ϵ tested since the reactions effectively resemble a set of series reactions, with each reaction having comparable magnitudes of flux. When ϵ equals 0.05, some of the reactions in module A and C were eliminated. The number of reactions in modules A–C was reduced from 85 to 64 in this case. Some of the eliminated reactions contributed to internalization via the smooth-pit pathway in module A, while others corresponded to signaling via the Shc-independent pathway in module C. Further, increasing ϵ to 0.07 resulted in complete elimination of all reactions in module C, and a reduction in number of reactions from 85 to 48 in modules A–C. This reduced model, for $\epsilon = 0.07$, is shown in Supplementary Figure S3. When ϵ equals 0.2, many of the internalization and endosomal signaling reactions in module B were eliminated. The number of reactions in modules A–C dropped to 34. As seen in Figure 8B, 44% reduction in the number of reactions ($\epsilon = 0.07$) resulted in only a $\sim 5\%$ alteration in system output over a range of EGF dosages. Our reduced model is also able to mimic the time-course of other components (like MEK-PP in Fig. 8C) with high fidelity.

Knockout experiments yield testable hypothesis

Knockout (KO) models that lack a particular reaction can be simulated, and sensitivity analysis can be performed on the modified

network. Such an analysis can yield hypotheses that are amenable to experimental testing. We illustrate this by selectively deleting some of the reactions shown in Figure 1. As expected, knocking out single reactions with high sensitivity coefficients dramatically alters the system output in comparison to knocking out reactions with lower sensitivity coefficients (Fig. 9A). Here, deleting reaction 27 (peak $W_{ij} = 6.5$ in Fig. 4A) results in a $>95\%$ decrease in peak $C_{\text{ERK-PP}}$ concentration, while only a $\sim 30\%$ decrease in output is observed upon creation of a reaction-22 KO (peak $W_{ij} = 0.8$ for reaction 22 in Fig. 4A). If sensitivity coefficients for normal cells are compared with W_{ij} for reaction-22 KOs, we observe that most of the sensitivity coefficients are identical in both cases except for the rates shown in Figure 9B. Notably, W_{ij} for reaction 27 decreases from 6.7 in normal control to 0.1 in reaction-22 KO. Because of this, double knockouts lacking both reactions 22 and 27 more closely resemble reaction-22 KOs rather than reaction-27 KOs (Fig. 9A). Also, W_{ij} for reaction 19 increases from 0.6 in normal controls to 4.8 in reaction-22 KOs. Thus, deletion of reaction 19 abrogates the system output in reaction-22 KOs, but not in normal controls. For this reason, while reaction 19 is not a necessary part of the reduced model for normal controls (Supplemental Fig. S3), it will be an important component of reduced models generated in reaction-22 KOs. Overall, our analysis suggests that while single KOs may lead to complete loss of signal output, dual KOs can partially restore this lost function. Also, since the sensitivity coefficients evaluated for the gene KO systems are different from normal controls, model reduction must be done independently for each gene KO.

DISCUSSION AND CONCLUSION

Mathematical modeling of signal transduction can provide quantitative and mechanistic understanding of diverse cellular functions including cell–cell communication, proliferation, differentiation and adhesion. When combined with high-throughput datasets, it can also contribute to the prediction of yet unidentified biochemical reactions and feedback mechanisms.

Using EGF mediated signaling as an example, we illustrate that various reactions in the EGF pathway have distinct contributions to signal output with varying EGF dosage and time after stimulation. While the binding of EGF to its receptor was important in the first phase, this feature was relatively insignificant at later times. In contrast while receptor internalization, endosomal signaling and degradation were relatively unimportant in Phase I, these features became important at low EGF concentrations and at longer times. These observations are in agreement with other published data (Vieira *et al.*, 1996), which note that internalization reactions and signaling from endosomal compartments may be significant at later times. The current study also extends the observations of others (Schoeberl *et al.*, 2002), by suggesting that EGF–EGFR binding kinetics plays a more prominent role in regulating signal output at the lower dosages than at higher dosages. Within the MAP kinase cascade also, while phosphorylation reactions were important at early times, dephosphorylation and phosphatase activity played a more prominent role at later times.

EGF mediated signaling flowed primarily through the Shc-dependent pathway (module B), rather than the Shc-independent pathway (module C) over the range of EGF dosages and times tested. This proposition is consistent with the results of others (Gong and Zhao, 2003). Immunoprecipitation experiments also demonstrate

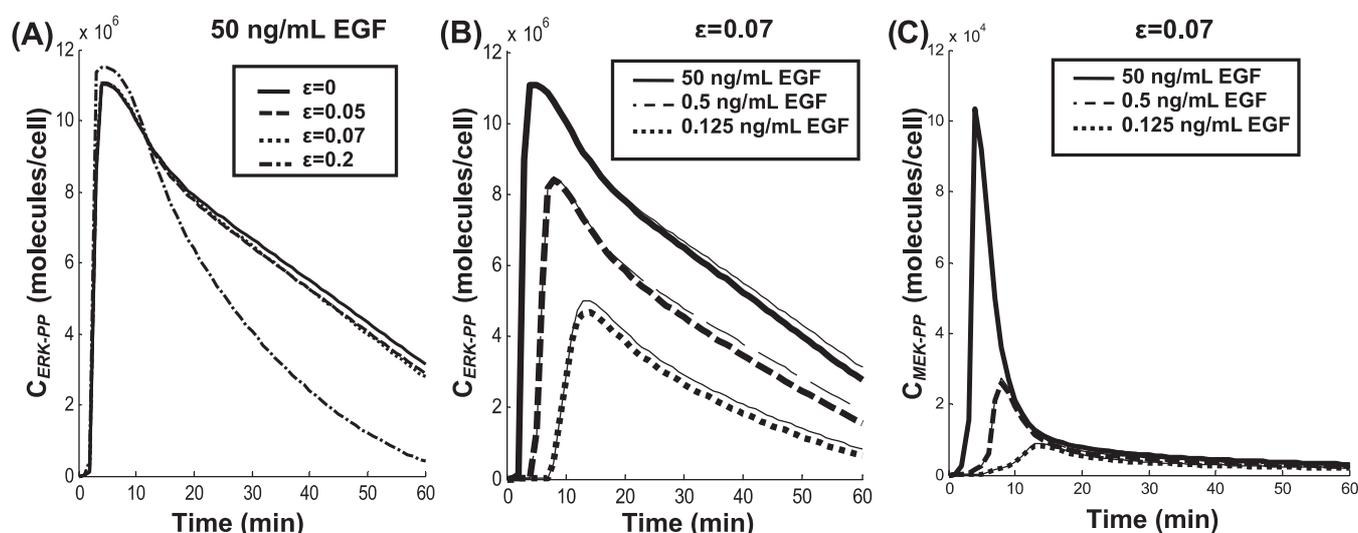


Fig. 8. Reduced versus original model: ERK-PP/MEK-PP concentrations are compared between reduced (Supplemental Fig. S3) and original (Supplemental Fig. S1) models. (A) ERK-PP response for EGF dosage of 50 ng/ml and varying values of reduction factor ϵ . $\epsilon = 0$ corresponds to the original model. (B) ERK-PP at constant $\epsilon = 0.07$ and different EGF dosages. (C) MEK-PP response at constant $\epsilon = 0.07$ and different EGF dosages. In panels B and C, the thick lines are the results from the reduced model, while thin lines correspond to the original model. The differences between the original and the reduced model in panel C are negligible.

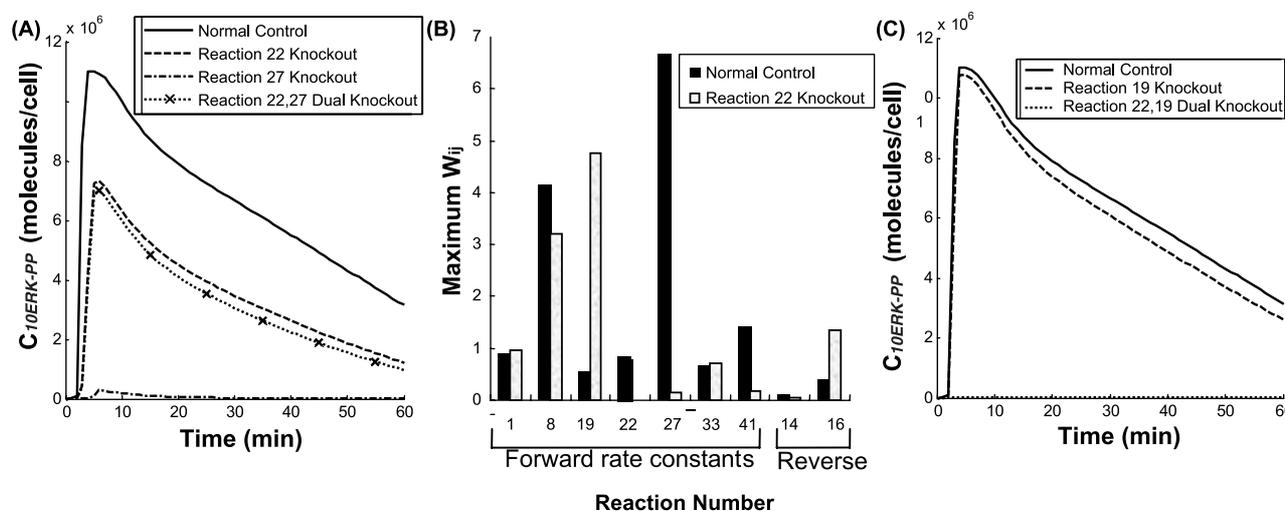


Fig. 9. Gene/reaction knockout: KO of particular reactions was generated by setting both the forward and reverse rate constants for that reaction to zero. The EGF concentration was 50 ng/ml. (A) Effect of knocking out reaction 22 and 27 on system output in comparison to normal control (with all reactions intact). Reaction 27 (which has a high-sensitivity coefficient) has a larger effect on system output as compared to reaction 22. Dual knockout restores lost function due to reaction-22 KO. (B) Maximum scaled sensitivity coefficients are different for reaction-22 KOs in comparison to normal control. For example, while the scaled sensitivity coefficient for reaction 19 is small for normal control, it is dramatically increased in the reaction-27 KOs. (C) Knocking out reaction 19 in normal control has a very small effect on the system output, while knocking out the same reaction in reaction-22 KOs dramatically reduces the system output.

that a larger fraction of Grb2 is bound to Shc rather than to EGFR (Sasaoka *et al.*, 1994; Kholodenko *et al.*, 1999). Further, EGFR binds to Shc with greater affinity than to Grb2 (Sasaoka *et al.*, 1994). Finally, it is reported that Shc binds EGFR through cooperation between two domains: the PI domain which binds to the phosphorylated tyrosine residue at Y1148 and the SH2 domain which binds phosphorylated Y1173 (Batzler *et al.*, 1995). Grb2, on the other hand, binds a single phosphorylated EGFR site at Y1068 through its

SH2 domain. In this model, the multivalent interaction of Shc to EGFR may stabilize the molecular interaction thus contributing to EGF signaling via module B.

We demonstrate that sensitivity analysis, when combined with flux analysis, can aid model reduction and simplification. Overall, as expected, increasing the value of the reduction factor resulted in model simplification at the cost of deviation from the full model. Based on model reduction it appears that the reactions least affecting

signal output correspond to internalization and receptor degradation in module A. Following this, in the order of increasing importance are the Shc-independent reactions in module C, the internalization reactions in module B and finally the reduced model displayed in Supplemental Figure S3. Of the intermediates, the EGF receptor complexed with phosphorylated Shc, Grb2 and Sos simultaneously (e.g. reactants 35 and 36) represent key intermediates that regulate system output. The concentrations of these complexes directly control the rate of Ras-GTP formation via reactions 26 and 27. Internalization of these intermediates also forms an important part of the reduced model. We note that flux analysis can also be combined with PCA to yield a reduced model. Results similar to those from the combined sensitivity and flux analysis above were obtained from such analysis (data not shown).

In this paper, sensitivity analysis is performed using C_{ERK-PP} as the system variable and many of the rate constants are derived from *in vitro* enzymology studies. Similar analyses can be performed using other system variables, and this can allow the identification of reaction intermediates whose concentrations are sensitive to a particular important reaction step but are relatively insensitive to others. These intermediates then represent targets for *in situ* experimental measurement for obtaining improved reaction rate data and new mechanistic insights.

Overall, we demonstrate the use of sensitivity analysis combined with PCA for the analysis of biochemical reaction networks. This method is particularly suited for the analysis of unsteady-state conditions that are typically observed during cell signaling. The direct differential method and PCA allow the systematic identification of rate-controlling steps. Such steps may represent sites for drug development or other intervention. Further, it is apparent that sensitivity and PCA of biochemical networks can be used to generate novel hypotheses that can guide biological experiments.

ACKNOWLEDGEMENT

We acknowledge grant supports, from NIH HL63014, HL76211 and the NYS/GSEU Professional Development Award Program.

REFERENCES

- Alter, O., Brown, P.O. and Botstein, D. (2000) Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl Acad. Sci. USA*, **97**, 10101–10106.
- Batzer, A.G., Blaikie, P., Nelson, K., Schlessinger, J. and Margolis, B. (1995) The phosphotyrosine interaction domain of Shc binds an LXNPXY motif on the epidermal growth factor receptor. *Mol. Cell Biol.*, **15**, 4403–4409.
- Bhalla, U.S. and Iyengar, R. (1999) Emergent properties of networks of biological signaling pathways. *Science*, **283**, 381–387.
- Chi-ying, F. and Ferrell, J.E.J. (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl Acad. Sci. USA*, **93**, 10078–10083.
- 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.
- Fell, D. (1997) Metabolic control analysis. In *Understanding the Control of Metabolism*. Portland Press, London, pp. 101–132.
- Fussenegger, M., Bailey, J.E. and Varner, J. (2000) A mathematical model of caspase function in apoptosis. *Nat. Biotechnol.*, **18**, 768–774.
- Gardner, T.S., di Bernardo, D., Lorenz, D. and Collins, J.J. (2003) Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, **301**, 102–105.
- Gong, Y. and Zhao, X. (2003) Shc-dependent pathway is redundant but dominant in MAPK cascade activation by EGF receptors: a modeling inference. *FEBS Lett.*, **554**, 467–472.
- Heinrich, R., Neel, B.G. and Rapoport, T.A. (2002) Mathematical models of protein kinase signal transduction. *Mol. Cell*, **9**, 957–970.
- Heinrich, R. and Rapoport, T.A. (1974) A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur. J. Biochem.*, **42**, 89–95.
- Holter, N.S., Mitra, M., Maritan, A., Cieplak, M., Banavar, J.R. and Fedoroff, N.V. (2000) Fundamental patterns underlying gene expression profiles: simplicity from complexity. *Proc. Natl Acad. Sci. USA*, **97**, 8409–8414.
- Kacser, H. and Burns, J.A. (1973) The control of flux. *Symp. Soc. Exp. Biol.*, **27**, 65–104.
- Kauffman, K.J., Pajeroski, J.D., Jamshidi, N., Palsson, B.O. and Edwards, J.S. (2002) Description and analysis of metabolic connectivity and dynamics in the human red blood cell. *Biophys. J.*, **83**, 646–662.
- 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., 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.
- Kikuchi, S., Tominaga, D., Arita, M., Takahashi, K. and Tomita, M. (2003) Dynamic modeling of genetic networks using genetic algorithm and S-system. *Bioinformatics*, **19**, 643–650.
- Kuharsky, A.L. and Fogelson, A.L. (2001) Surface-mediated control of blood coagulation: the role of binding site densities and platelet deposition. *Biophys. J.*, **80**, 1050–1074.
- Lee, K.H., Dinner, A.R., Tu, C., Campi, G., Raychaudhuri, S., Varma, R., Sims, T.N., Burack, W.R., Wu, H., Warg, J. et al. (2003) The immunological synapse balances T cell receptor signaling and degradation. *Science*, **302**, 1218–1222.
- Morrison, D.K. and Cutler, R.E. (1997) The complexity of Raf-1 regulation. *Curr. Opin. Cell Biol.*, **9**, 174–179.
- Rapoport, T.A., Otto, M. and Heinrich, R. (1977) An extended model of the glycolysis in erythrocytes. *Acta. Biol. Med. Ger.*, **36**, 461–468.
- Resat, H., Ewald, J.A., Dixon, D.A. and Wiley, H.S. (2003) An integrated model of epidermal growth factor receptor trafficking and signal transduction. *Biophys. J.*, **85**, 730–743.
- Sachs, K., Gifford, D., Jaakkola, T., Sorger, P. and Lauffenburger, D.A. (2002) Bayesian network approach to cell signaling pathway modeling. *Sci. STKE*, **2002**, PE38.
- Sasaoka, T., Langlois, W.J., Leitner, J.W., Draznin, B. and Olefsky, J.M. (1994) The signaling pathway coupling epidermal growth factor receptors to activation of p21ras. *J. Biol. Chem.*, **269**, 32621–32625.
- Sastry, L., Lin, W., Wong, W.T., Di Fiore, P.P., Scoppa, C.A. and King, C.R. (1995) Quantitative analysis of Grb2-Sos1 interaction: the N-terminal SH3 domain of Grb2 mediates affinity. *Oncogene*, **11**, 1107–1112.
- Savageau, M.A. (1976) *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Addison-Wesley, Reading, MA.
- 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.
- Starbuck, C. and Lauffenburger, D.A. (1992) Mathematical model for the effects of epidermal growth factor receptor trafficking dynamics on fibroblast proliferation responses. *Biotechnol. Prog.*, **8**, 132–143.
- Stephanopoulos, G.N., Aristidou, A.A. and Nielsen, J. (1998) *Metabolic Engineering, Principles and Methodologies*. Academic Press, New York, pp. 313–350.
- Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T.S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tarida, S., Yugi, K., Venter, J.C. and Hutchison, 3rd, C.A. (1999) E-CELL: software environment for whole-cell simulation. *Bioinformatics*, **15**, 72–84.
- Varma, A., Morbidelli, M. and Wu, H. (1999) *Parametric Sensitivity in Chemical Systems*. Cambridge University Press, New York.
- Vaseghi, S., Baumeister, A., Rizzi, M. and Reuss, M. (1999) *In vivo* dynamics of the pentose phosphate pathway in *Saccharomyces cerevisiae*. *Metab. Eng.*, **1**, 128–140.
- Vieira, A.V., Lamaze, C. and Schmid, S.L. (1996) Control of EGF receptor signaling by clathrin-mediated endocytosis. *Science*, **274**, 2086–2089.