AUTOMATIC MODEL GENERATION FOR SIGNAL TRANSDUCTION WITH APPLICATIONS TO MAP-KINASE PATHWAYS

Category: Methodologies for system-level understanding of life

Bruce E Shapiro*[1], Andre Levchenko[2], Eric Mjolsness*[3]**
* Jet Propulsion Laboratory, California Institute of Technology
** Divison of Biology and Division of Engineering and Applied Science, California Institute of Technology

ABSTRACT

We describe a general approach to automatic model generation in the description of dynamic regulatory networks. Several potential areas of application of this description are outlined. We then describe how a particular implementation of this approach, Cellerator , can be used in study of the mitogen-activated protein kinase (MAPK) cascade signal transduction modules operating in solution or when bound to a scaffold protein. We show that the results of simulations with the Cellerator –created model are consistent with our previously published report, where an independently written model was described. New results made possible by the use of Cellerator are presented. An important aspect of Cellerator operation – explicit output description at several steps through the model generation is emphasized. This design allows intervention and modification of the model "on the go" leading to increased flexibility of model description and straightforward error correction mechanism. We also outline our future plans in Cellerator development.

INTRODUCTION

In the past few decades rapid gain of information about intracellular signal transduction and genetic networks resulted in a view of regulatory biomolecular circuits as highly structured multi-component systems evolved to optimally perform in very uncertain environments. The emergent complexity of biochemical intracellular regulation necessitates development of new tools for analysis, most notably computer assisted mathematical models. Computer modeling, proved to be of crucial importance in analysis of genomic DNA sequences and molecular dynamics simulations, is likely to become an indispensable tool in biochemical and genetic research. There are presently several attempts of creating platforms for building computer models of cell signaling and gene regulation.

In spite of their promise, these new modeling environments have so far made very limited inroads into the wide biological research community. Arguably, among the reasons for this is a relative inaccessibility of the modeling interface for the typical classically trained geneticist or biochemist. Instead of cartoon representations of signaling pathways, in which activation can be represented simply by an arrow connecting two molecular species, they are often asked to write a specific differential equation or chose among different modeling approximations. For sufficiently complex biomolecular circuits such description involves explicit writing out dozens of equations, a job often difficult, tedious and prone to error even for an experienced computer modeler. It would be of use then to create a modeling interface allowing automatic conversion of a cartoon or reaction based biochemical pathway description into a mathematical representation suitable for solvers inbuilt into various currently existing software packages.

---

[1] Bruce E. Shapiro, Machine Learning Systems Group, Jet Propulsion Laboratory, California Institute of Technology, M/S 264-355, 4800 Oak Grove Drive, Pasadena CA, USA 91109; tel. (818) 393 0980 bshapiro@jpl.nasa.gov.

[2] Andre Levchenko, Division of Biology, California Institute of Technology, Mail Stop 156-29, Pasadena, CA USA 91125; tel. (626) 395 8542 andre@vigeland.paradise.caltech.edu.

[3] Eric Mjolsness, Machine Learning Systems Group, Jet Propulsion Laboratory, California Institute of Technology, M/S 126-347, 4800 Oak Grove Drive, Pasadena, CA USA 91109; tel. (818) 393 5311mjolsness@jpl.nasa.gov.

A tool allowing automatic generation of a mathematical model description has thus the advantage of an easier accessibility for a wider research community. Another important benefit of having such a tool is facilitation of modeling very complex networks or interaction that may be present in the system of interest. For example, in intracellular signal transduction it is not uncommon to find multimolecular complexes of modifiable proteins. One example of such complexes – scaffolds in MAPK cascades – will be studied in detail later in this report. It can be demonstrated that the number of different states a multimolecular complex can be found in increases exponentially with the number of participating molecules. If, as is often the case, the dynamics of each of the states is of interest, one may find that the unpleasant task of writing dozens if not hundreds of equations is necessary. Automatic generation of equations, however, circumvents this difficulty.

In this report we consider a general approach to carrying out automatic model generation in description of dynamic regulatory networks. Several potential areas of application of this description will be outlined. We then will describe how a particular implementation of this approach, Cellerator , can be used in study of the mitogen-activated protein kinase (MAPK) cascade signal transduction modules operating in solution or when bound to a scaffold protein. An important aspect of Cellerator operation – explicit output description at several steps through the model generation will be emphasized. This design allows intervention and modification of the model "on the go" leading to increased flexibility of model description and straightforward error correction mechanism.
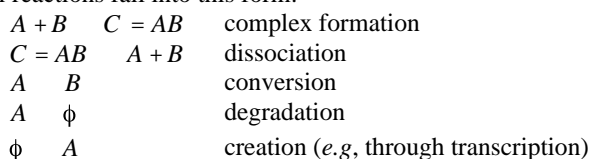
AUTOMATIC MODEL GENERATION

Canonical Forms for Cell Simulation

We can loosely classify the components needed to perform cell simulation in order of their biological complexity: simple chemical reactions including degradation, enzymatic reactions in solution, multi-molecular complexes with a non-trivial number of states (e.g., scaffold proteins), multiple interacting and non-overlapping pathways, transcription, translation, intracellular components, transport processes and morphogenesis. We will examine these processes and attempt to derive general *canonical forms* that can be used to describe these processes in the following paragraphs. These canonical forms can be either *input forms*, such as chemical reactions, or *output forms*, such as differential equations that are automatically generated by the program. It is crucial to identify these canonical forms so that an efficient mapping from the input forms to the output forms can be implemented. Specific examples of how these forms may be implemented in a computer program are given in the following section.

Biochemistry is frequently referred to as the language of biology, in much the same way that mathematics has been called the language of physics. Cellular activity is generally expressed in terms of the biochemical cascades that occur. These chemical reactions constitute the core of our input forms; the corresponding differential equations constitute the core of our output forms. (Differential equations can be thought of as output because they are passed on solver and/or optimizer modules to handle). A fundamental library of simple chemical reactions can be quickly developed; such reactions take the form

$$X_i \xrightarrow{k} Y_i \qquad (1)$$
$$X_i \subseteq S, \quad S' \qquad Y_i \subseteq S, \quad S''$$
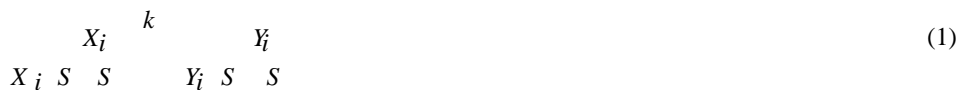
where S is a set of reactants and S' and S'' are (possible empty and possibly non-distinct) subsets of S and k is a representation of the rate at which the reaction proceeds. In general there are rarely more than two elements in either S' or S'' but it is possible for there to be more. Fore example, all of the following chemical reactions fall into this form:

$A + B \to C = AB$	complex formation
$C = AB \to A + B$	dissociation
$A \to B$	conversion
$A \to \phi$	degradation
$\phi \to A$	creation (*e.g*, through transcription)

Enzyme kinetic reactions, which are usually written as

$$S + E \quad P + E \tag{2}$$

where E is an enzyme that facilitates the conversion of the substrate S into the product P, would also fall into this class. More generally, equation 2 is a simplification of the cascade

$$S + E \quad SE \quad S + P \tag{3}$$
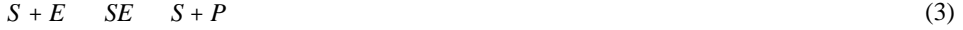
where the double arrow is used to indicate that the first reaction is reversible, i.e., it is equivalent to the pair of reactions

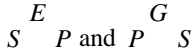$$S + E \quad SE, \, SE \quad S + E$$
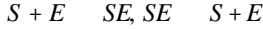
We use the double-arrow notation to indicate set of three reactions in (3),

$$S \overset{E}{\quad} P \tag{4}$$

to indicate the conversion of S to P that is catalyzed by the enzyme E. We further use the double-double arrow notation

$$S \overset{E}{\underset{G}{\quad}} P \tag{5}$$

to indicate the pair of enzymatic reactions

$$S \overset{E}{\quad} P \text{ and } P \overset{G}{\quad} S$$

Observe that equation (5) therefore represents a total of six elementary reactions, each of which is of the form given by equation (1). We therefore take equation (1) as our input canonical form for chemical reactions. The corresponding output canonical form is given by the set of differential equations

$$\tau_i \dot{X}_i = \sum_\alpha c_{i\alpha} \prod_j X_j^{n_{i\alpha j}} \tag{6}$$

where the τ_i and $c_{i\alpha}$ are constants that are related to the rate constants , the signs of the $c_{i\alpha}$ are determined from which side of equation (1) the terms in equation (6) correspond to, and the $n_{i\alpha j}$ represent the cooperativity of the reaction. The summation is taken over all equations in which X_i appears. Multimolecular reactions (e.g., binding to a scaffold protein) and multiple interacting and overlapping pathways are described in much the same way - there are just more reactions that must be included in our model. Every one of these reactions can still be described by the canonical forms (1) and (6).

Genetic transcription and translation into proteins can be described by an extension of the form (6) to include terms of the form

$$\tau_i \dot{X}_i = \sum_\beta \frac{c_{i\beta} X_\beta^{n\beta}}{K_{oi\beta} + X_\beta^{n\beta}} \tag{7}$$

In the case of transcription, X_i would be the quantity of mRNA produced, and the sum is taken over all transcription factors present for X_i. If there are any reactions of the form (1) for X_i then the expression on the right side of equation (7) would be added to the right hand side of (6). In the case of translation, X_i would be the protein produced X_β represents mRNA.

Sub-cellular components represent a higher order of biological complexity. If we assume perfect mixing each component can be treated as a separate pool of reactants which we can describe by the reaction

$$X_A \quad X_B \tag{8}$$

This is taken to mean that X in pool A is transported into pool B at some rate. When the concentration changes and distances involved are small such processes can be described by the canonical forms in equation (1). In large or elongated cells with long processes (such as neurons) or when the molecules have a net charge the transport process defined in equation (8) can not be described by the output canonical form (6). Instead we must modify this ordinary differential equation into a partial differential equation to allow for diffusion,

$$\tau_i \frac{\partial X_i}{\partial t} = \cdot \left(D_i \ X_i + C_i D_i \ V \right) + \sum_\alpha c_{i\alpha} \prod_j X_j^{n_{i\alpha j}} \tag{9}$$

where the D_i are (possibly spatially dependent) diffusion constants for species X_i, C_i are charge and temperature dependent constants, and V is the voltage. Other voltage and pressure dependent movement between compartments (especially those with membranes) that are controlled by channels and transport proteins could be described by including additional terms on the right hand side of equation (9) (e.g., Hodgkin-Huxley type expressions).

Implementation

Protein cascades are generally written by biologists in a compact notation with arrows; to translate this into a computable form we can specify it as a multiset

$$C = \{ P, R, IC, I, F \} \tag{10}$$

where P is a set of proteins, R is a set of reactions, IC is a set of initial conditions, I is a set of input functions, and F is a set of output functions. To see the relationship between the two (chemical notation and set notation) consider a simple linear phosphorylation cascade

$$A \quad B \quad C$$

which means that A, when it is phosphorylated, facilitates the phosphorylation of B, which in turn facilitates the phosphorylation of C. In general, a cascade can have any length, so we define the elements of a cascade with a simple indexed notation, *e.g*,

$$K_4 \quad K_3 \quad K_2 \quad K_1,$$

where K is used to indicate that all the members of the cascade induce phosphorylation of their substrates, that is they are kinases. In general, activation can proceed by any specified means.

This indexed notation is always used internally by the program; the user, however, has the option of using either common names or the indexed variables. There is still a great deal of information hidden in this expression, such as how many phosphate groups must be added to make each successive protein active. In the MAPK cascade for example (as explained below), the input signal that starts this cascade is K_4. The output, however, is not K_1, as this notation would suggest, but a doubly phosphorylated version of K_1. Hence for MAPK cascade we introduce a modified notation:

$$
\begin{array}{cccccc}
K_3 & \overset{K_4}{} & K_3 & & & \\
K_2 & \overset{K_3}{} & K_2 & \overset{K_3}{} & K_2 & \\
K_1 & \overset{K_2}{} & K_1 & \overset{K_2}{} & K_1 &
\end{array}
\tag{11}
$$

where each phosphate group that has been added is indicated with an asterisk. From this notation it is clear that the input is K_4 and the output is $K_1{}^{**}$. In general, suppose we have a cascade formed by n proteins K_1, K_2, ..., K_n, and that the ith protein K_i can be phosphorylated a_i times. Denote by $K_i{}^j$ the fact that kinase K_i has be phosphorylated j (possibly zero) times. The set P of all kinases $K_i{}^j$ in an n-component cascade is then

$$P = \{K_i^j \,|\, i = 1,2,..., n,\ j = 0,1,\ldots, a_i\} \tag{12}$$

The *reactions* in the cascade are of the form

$$R = \left. K_i^j \quad \xrightarrow{K_{i+1}^{a_i+1}} \quad K_i^{j+1} \right| i = 1,\ldots,n-1, j = 0,\ldots, a_j - 1 \tag{13}$$

We note at this point that this notation describes a linear cascade, in which each element K_i is only phosphorylated by the active form of K_{i+1}. It does not include other reactions, when, for example, K_3 might, under special circumstances, phosphorylate K_1 directly without the intermediate step of first phosphorylating K_2. Such additional reactions could be added explicitly, but they have been omitted from this presentation to simplify the discussion. We can also add the dephosphorylation enzymes, ot phosphatases, with a double-arrow notation:

$$R = \left. K_i^j \quad \underset{Ph_i}{\overset{K_{i+1}^{a_i+1}}{\rightleftharpoons}} \quad K_i^{j+1} \right| i = 1,\ldots,n-1, j = 0,\ldots, a_j - 1 \tag{14}$$

Feedback is specified by setting up aliases for the variables, e.g., $Ph_1 \circ K_1$ would indicate that Ph_1 is a phosphatase that acts on K_1.

In general, it is not necessary to specify explicit conservation laws with this notation because they are built directly into the equations. For example, we do not have to separately specify that the quantities

$$K_i^{Total} = \sum_{j=0}^{a_i} K_i^j \tag{15}$$

because this is implicit in the differential equations that are built using this notations. We do, however, have to specify the initial conditions,

$$IC = \left\{ K_i^j(0) \,|\, i = 1,2,\ldots n,\ j = 0,1,\ldots, a_i \right\} \tag{16}$$

Next, we need to specify how the cascade is initiated. For example if K_4 is not present until some time t_{on} and then is fixed at a level c, would write the set of *input functions* as

$$I = \left\{ K_4(t) = cH(t - t_{on}) \right\} \tag{17}$$

where $H(t) = \begin{cases} 0, t < 0 \\ 1, t \geq 0 \end{cases}$ is the Heaviside step function. In some cases, we are only interested in the total quantity of each substance produced as a function of time, e.g., $K_i^j(t)$. More generally, we would also specify a set of *output functions F*. For example we might have $F = \{f, g\}$ where $f(T)$ is the total accumulated protein concentration after some time T,

$$f(T) = \int_{t_{on}}^{T} K_1^{a1}(t)\,dt \tag{18}$$

and $g(c)$ is the steady state concentration of activated kinase,

$$g(c) = \lim_{t_{on}} \lim_{t} K_1^{a1}(t) \tag{19}$$

where c is the input signal specified I. Then the cascade is then completely specified by the multiset $C = \{P, R, IC, I, F\}$.

If we have an additional regulatory protein, such as scaffold (see below), there are additional reactions to describe binding to the scaffold and phosphorylation within the scaffold. We define the scaffold itself by an object $S_{p_1, p_2, \cdots, p_n}$ where n is as before (the number of kinases that may bind to the scaffold, or alternatively, the number of "slots" in the scaffold) and $p_i \in \{\varepsilon, 0, 1, \ldots, a_i\}$ indicates the state of phosphorylation of the proteins in each slot. Thus if $p_i = \varepsilon$ (or, alternatively, -1) the slot for K_i is empty, if $p_i = 0$, K_i^0 is in the slot, etc. For a three-slot scaffold, for example, we must add to the set P the following set

$$P = \left\{ S_{ijk} \middle| i = \varepsilon, 0, 1, \ldots, a_1, j = \varepsilon, 0, 1, \ldots, a_2, k = \varepsilon, 0, 1, \ldots, a_3 \right\} \tag{20}$$

To describe binding to the scaffold, we also need to add to the set R the reactions

$$R = \left\{ S_{p_1, \cdots, p_i = \varepsilon, \cdots, p_n} + K_i^j \quad S_{p_1, \cdots, p_i = j, \cdots, p_n} \right\} \tag{21}$$

where the indices run over all values in the range

$$p_i = \begin{matrix} \varepsilon, 0, 1, \ldots, a_i, i & j \\ 0, 1, \ldots, a_i, & i = j \end{matrix} \tag{22}$$

For the three member scaffold this would be

$$\begin{aligned}
R = &\left\{ S_{\varepsilon jk} + K_1^i \quad S_{ijk}, i = 0, \ldots, a_1, j = \varepsilon, 0, \ldots, a_2, k = \varepsilon, 0, \ldots, a_3 \right\} \\
&\cup \left\{ S_{i\varepsilon k} + K_2^j \quad S_{ijk}, i = \varepsilon, 0, \ldots, a_1, j = 0, \ldots, a_2, k = \varepsilon, 0, \ldots, a_3 \right\} \\
&\cup \left\{ S_{ij\varepsilon} + K_3^k \quad S_{ijk}, i = \varepsilon, 0, \ldots, a_1, j = \varepsilon, 0, \ldots, a_2, k = 0, \ldots, a_3 \right\}
\end{aligned} \tag{23}$$

Finally, we have phosphorylation in the scaffold. This can be done either by a protein that is not bound to the scaffold, *e.g*, for the input signal,

$$R = S_{p_1, \cdots, p_{i-1} = j < a_{i-1}, p_i = a_i, \cdots, p_n} + K \quad S_{p_1, \cdots, p_{i-1} = j+1, p_i = a_i, \cdots, p_n} \tag{24}$$

where the two-sided double arrow (   ) is used as shorthand for the (possibly bidirectional) enzymatic reaction, or by one that is bound to the scaffold,

$$R = S_{p_1, \cdots, p_{i-1} = j < a_{i-1}, p_i = a_i, \cdots, p_n} \quad S_{p_1, \cdots, p_{i-1} = j+1, p_i = a_i, \cdots, p_n} \tag{25}$$

or by some combination of the two, all of which must be added to the reaction list R. For the three-slot scaffold with external signal K_4 that activates K_3, we have

$$R = \left\{ S_{i,a_2 k} \quad S_{i+1,a_2,k}, i = 0,\ldots,a_1 - 1, k = \varepsilon, 0, \ldots, a_3 \right\}$$
$$\cup \left\{ S_{i,j,a_3} \quad S_{i,j+1,a_3}, i = \varepsilon, 0, \ldots, a_1, j = \varepsilon, 0, \ldots, a_2 - 1 \right\} \quad (26)$$

and

$$R = S_{ijk} \overset{K_4}{\underset{Ph_3}{}} S_{i,jk+1}, \quad i = \varepsilon, 0, \ldots, a_1, j = \varepsilon, 0, \ldots, a_2, k = 0, \ldots, a_3 - 1\} \quad (27)$$

Typical a_i values for this type of cascade are $a_1 = a_2 = 2$ and $a_3 = 1$.

As an example, let us continue with the above-mentioned three-member cascade that is initiated with K_4. In what follows, we refer to Cellerator , a Mathematica$^®$ package that implements the above algorithms. In Cellerator we have defined the function

genReacts[*kinase-name***, *n*, {*a$_i$*}, *phosphatase-name***],**
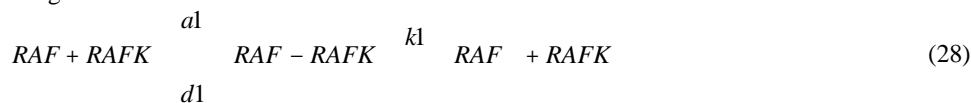
where *kinase-name* and *phosphatase-name* are the names we want to give to the sequences of kinases and phosphatases, respectively, and *n* and *a$_i$* are as before. The following Cellerator command then generates the above set reactions (11),

```
genReacts[K, 3, {2, 2, 1, 1}, kpase]

              K[2,2]              K[2,2]              K[3,1]              K[3,1]              K[4,1]
{K[1, 0] ⇌ K[1, 1] , K[1, 1] ⇌ K[1, 2] , K[2, 0] ⇌ K[2, 1] , K[2, 1] ⇌ K[2, 2] , K[3, 0] ⇌ K[3, 1]}
         kpase[1]             kpase[1]             kpase[2]             kpase[2]             kpase[3]
```
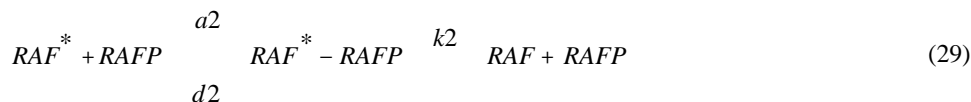
The input is in the first line while the output is the second line. Alternatively, the user could specify the set of reactions explicitly, or copy the output to a later cell to manually add additional reactions. If RAF has been set up as an alias for K_3 then the rate constants are specified by a content-addressable syntax, e.g., as

```
                            RAFK
storeRateConstant[db, RAF ⇌ RAF*, a1, d1, k1, a2, d2, k2];
                            RAFP
```

corresponding to

$$RAF + RAFK \overset{a1}{\underset{d1}{}} RAF - RAFK \overset{k1}{} RAF^* + RAFK \quad (28)$$

and

$$RAF^* + RAFP \overset{a2}{\underset{d2}{}} RAF^* - RAFP \overset{k2}{} RAF + RAFP \quad (29)$$

and so forth, where the numbers over the arrows indicate the rate constants (and not enzymes, as with the double arrow notation). This particular set has 5 high level reactions that are subsequently parsed into 30 low level reactions according to enzyme kinetics description (e.g., formation of each intermediate compound), each with a unique rate constant, and a set of 21 differential equations for 8 kinases, since K_4 is specified by an explicit input function and does not have an ODE of its own, 3 phosphatases, and 10 intermediate compounds. When scaffold proteins are included (discussed below) these numbers increase to 139 high level reactions, 348 low level reactions (300 without kinases), and 101 differential equations (85 without kinases).

MAPK PATHWAY WITH SCAFFOLDS: EXPERIMENTAL BACKGROUND

The mitogen-activated protein kinase (MAPK) cascades (Fig. 1) are a conserved feature of a variety of receptor mediated signal transduction pathways ((Garrington and Johnson, 1999; Widmann et al., 1999; Gustin et al., 1998)). In humans they have been implicated in transduction of signals from growth factor, insulin and cytokine receptors, T cell receptor, heterotrimeric G proteins and in response to various kinds of stress ((Garrington and Johnson, 1999; Putz et al., 1999; Sternberg and Alberola-Ila, 1998; Crabtree and Clipstone, 1994; Kyriakis, 1999)). A MAPK cascade consists of three sequentially acting kinases. The last member of the cascade, MAPK is activated by dual phosphorylation at tyrosine and threonine residues by the second member of the cascade: MAPKK. MAPKK is activated by phosphorylation at threonine and serine by the first member of the cascade: MAPKKK. Activation of MAPKKK apparently proceeds through different mechanisms in different systems. For instance, MAPKKK Raf-1 is thought to be activated by translocation to the cell membrane, where it is phosphorylated by an unknown kinase. All the reactions in the cascade occur in the cytosol with the activated MAPK translocating to the nucleus, where it may activate a battery of transcription factors by phosphorylation.
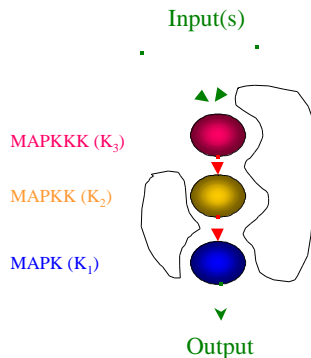

Figure 1. The topology of MAPK signaling cascade. Each red arrow represents activation through dual phosphorylation. Two and three-member scaffolds have been identified experimentally and are depicted here.

MAPK cascades have been implicated in a variety of intercellular processes including regulation of the cell cycle, apoptosis, cell growth and responses to stress. These molecules are crucially important in development of memory and wound healing. Abnormal changes in MAPK pathway regulation often mediate various pathologies, most notably cancer. This central role of MAPK mediated signal transduction in most regulatory processes makes it an especially attractive research and modeling object.

Signal transduction through MAPK cascade can be very inefficient unless additional regulatory proteins called scaffolds are present in the cytosol. Scaffold proteins nucleate signaling by binding two or more MAP kinases into a single multimolecular complex. It has been reported previously that scaffolds can both increase and decrease the efficiency of signaling in a concentration dependent manner (Levchenko et al., 2000). In addition they can reduce the non-linear activation characteristics of the cascade. These properties may be crucial for global and local activation of MAPK as scaffold proteins may selectively translocated to small subcellular compartments thus locally facilitating or inhibiting MAPK activation. We have presented a model of scaffold mediated signal transduction previously (Levchenko et al., 2000). In this report we show how the use of Cellerator software package has allowed us to substantially improve this model and study sensitivity to a parameter not investigated in the preceding report.

MAPK PATHWAY WITH SCAFFOLDS: RESULTS

As described above, addition of scaffold proteins into the MAPK reaction system results in markedly increased number of states and equations describing transitions between them. Here the benefits provided by Cellerator can really be appreciated, as a simple sequence of commands can lead to automatic description of all reactions involving scaffold-kinase complexes (see Fig. 2).

In our simulations the first goal was to verify the automatic model generation for scaffold-medicated MAPK cascade as implemented in Cellerator . As a basis for the comparison we referred to our previous report describing a quantitative model of the effect scaffold proteins can play in MAPK mediated signal transduction. When all the assumptions of that model were made again exactly the same solution for the three-member scaffold case was obtained. This convergence of results verified the model generated by Cellerator . In addition, the difficulty of manual generation of all the necessary equations, a limiting factor of the previous study, has now been removed. We thus attempted to study a more detailed model, in which some of the previous assumptions were relaxed.

```
phosphorylationReactions = genScafPhosReacts[S, {2, 2, 1, 1}, K]

{S[0, 2, -1] →S[1, 2, -1], S[0, 2, 0] →S[1, 2, 0], S[0, 2, 1] →S[1, 2, 1],
 S[1, 2, -1] →S[2, 2, -1], S[1, 2, 0] →S[2, 2, 0], S[1, 2, 1] →S[2, 2, 1],
 S[-1, 0, 1] →S[-1, 1, 1], S[-1, 1, 1] →S[-1, 2, 1], S[0, 0, 1] →S[0, 1, 1],
 S[0, 1, 1] →S[0, 2, 1], S[1, 0, 1] →S[1, 1, 1], S[1, 1, 1] →S[1, 2, 1],
                                                              K[4,1]
 S[2, 0, 1] →S[2, 1, 1], S[2, 1, 1] →S[2, 2, 1], S[-1, -1, 0] ⇌S[-1, -1, 1],
              K[4,1]                K[4,1]                  K[4,1]
 S[-1, 0, 0] ⇌S[-1, 0, 1], S[-1, 1, 0] ⇌S[-1, 1, 1], S[-1, 2, 0] ⇌S[-1, 2, 1],
              K[4,1]               K[4,1]                 K[4,1]
 S[0, -1, 0] ⇌S[0, -1, 1], S[0, 0, 0] ⇌S[0, 0, 1], S[0, 1, 0] ⇌S[0, 1, 1],
             K[4,1]                 K[4,1]                 K[4,1]
 S[0, 2, 0] ⇌S[0, 2, 1], S[1, -1, 0] ⇌S[1, -1, 1], S[1, 0, 0] ⇌S[1, 0, 1],
             K[4,1]                 K[4,1]                 K[4,1]
 S[1, 1, 0] ⇌S[1, 1, 1], S[1, 2, 0] ⇌S[1, 2, 1], S[2, -1, 0] ⇌S[2, -1, 1],
             K[4,1]                 K[4,1]                 K[4,1]
 S[2, 0, 0] ⇌S[2, 0, 1], S[2, 1, 0] ⇌S[2, 1, 1], S[2, 2, 0] ⇌S[2, 2, 1]}
```

Figure 2. The implementation of automatic generation of the MAP kinases activation reactions (through phosphorylation) in the scaffold in the Cellerator environment. All the possible scaffold states (species) are generated as are the transition reactions between them. The indexes in the parentheses indicate the phosphorylation status of the kinase in the corresponding position, with –1 corresponding to the absence of the kinase from the scaffold complex. K[4,1] represents the external kinase activating the first MAP kinase (MAPKKK) in the cascade.

The use of Cellerator software has allowed us to perform systematic investigation of sensitivity to the assumptions made in our previous report describing the role of scaffold proteins in regulation of MAPK cascades (Levchenko et al., 2000). In particular, in the previous report we described dual MAPKK and MAPK phosphorylation within the scaffold to proceed in effect through a single step (processive activation). This is substantially different from a two-step dual phosphorylation occurring in solution. In this, distributive, activation the first phosphorylation event is followed by complete dissociation from the activating kinase and then the second phosphorylation reaction. The assumption of processive phosphorylation in the scaffold has some experimental basis. Mathematically, it is equivalent to assuming that the rate of the second phosphorylation reaction is fast compared to the first reaction. Although this assumption was partially relaxed in our previous report, no systematic study of relaxation of this assumption has been performed. Using Cellerator we performed systematic investigation of the role of increase or decrease of the rate of the second phosphorylation within the scaffold compared to reactions in solution. The results for the case when the two rates are equal are presented in Fig. 3. It is clear that relaxation of this assumption results in a substantial decrease of efficiency of signal propagation.

Similar simulations were performed to investigate the effect of allowing formation of a complex between MAPKKK in the scaffold and MAPKKK-activating kinase, as well as the effect of allowing phosphatases to dephosphorylate scaffold-bound kinases. In all cases the parameter values used in simulation are equal to those used for corresponding reactions in solution (for the full list of parameters see Levchenko et al., 2000). The results are presented in Fig. 3. Again, new assumptions resulted in substantial down-regulation of efficiency of signal propagation. It is of interest that the position of the optimum

scaffold concentration (at which the maximum signaling is achieved) is insensitive to making these new assumptions. This agrees with the analysis in (Levchenko et al., 2000), which suggested that the position of the optimum is determined only by the total concentrations of the kinases and their mutual interaction with the scaffold.
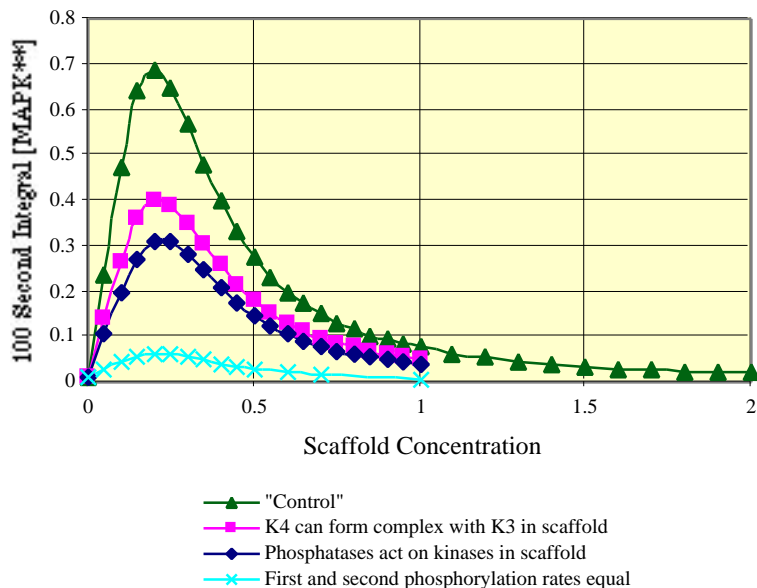

Figure 3 . The effect of relaxing several assumptions made in the previous report. The time integral of free dually phosphorylated MAPK over first 100 sec. is plotted vs. scaffold concentration. The "control" curve reproduces the data with all the assumptions made previously, whereas the other curves represent the results of relaxation of these assumptions as described in the legend. All data are obtained using the **Cellerator**[©] package and are plotted in Microsoft Excel.

DISCUSSION AND FUTURE DIRECTIONS

As shown in this report, automatic model generation can simplify the transition step between informal, cartoon based description of reactions in a pathway or a network of pathways to a series of equations based on rigorous description of enzymatic kinetics and other biochemical processes. In addition to facilitating the potentially burdensome task of correctly writing out all the equations necessary for description of a signaling pathway, this methodology provides an explicit and flexible way of controlling successive stages of model creation. In particular, a user intervention is possible at the stage of conversion an informal pathway description into a set of chemical reactions as well as at the later stage of mapping these reactions to the corresponding mathematical forms. This flexibility is likely to increase the ability of the user to participate in building and modifying model at the level limited only by his or her expertise.

Automatic model generation will prove especially useful in describing complex biochemical reactions involving formation of multimolecular complexes. Such complexes may exist in numerous states, each requiring a corresponding equation for description of its dynamics. In many situations writing out all these equations is next to impossible. As we demonstrated for the case of a generic three member scaffold in MAPK cascade mediated signaling, the use of a particular implementation of automatic model generator, Cellerator , we were capable of correctly generating and solving 101 differential equations, a task not achieved in the previous detailed study of the effect of scaffolds.

We intend to pursue the research into role of scaffolds in signal transduction regulation using this new tool. In particular we intend to use extended indexing to specify reactions occurring in various sub-cellular compartments. This will facilitate the study of the effect of scaffold translocation to the cell membrane observed in gradient sensing and other important regulatory processed. In addition we will attempt to develop our algorithm to allow for scaffold dimerization, an experimentally observed phenomenon.

Currently, Cellerator is "tailor-made" for modeling events in a linear pathway mediated by sequential covalent modification. It is within our immediate plans to make the code more universal to include other canonical forms and variable structure systems. In particular, we are in the process of adapting Cellerator to two test cases: NF-κB and PKA pathways. Consideration of these pathways will necessitate implementation of the elementary reactions describing transcription, translation and protein degradation. In addition complex formation will be considered a high level reaction leading to an activation step within the pathway.

ACKNOWLEDGEMENTS

BIBLIOGRAPHY

Crabtree, G.R. and Clipstone, N.A. (1994). Signal transmission between the plasma membrane and nucleus of T lymphocytes. Annu.Rev.Biochem. *63*, 1045-1083.

Garrington, T.P. and Johnson, G.L. (1999). Organization and regulation of mitogen-activated protein kinase signaling pathways. Curr.Opin.Cell Biol. *11*, 211-218.

Gustin, M.C., Albertyn, J., Alexander, M., and Davenport, K. (1998). MAP Kinase Pathways in the Yeast Saccharomyces cerevisiae. Microbiol.Mol.Biol.Rev. *62*, 1264-1300.

Kyriakis, J.M. (1999). Making the connection: coupling of stress-activated ERK/MAPK (extracellular-signal-regulated kinase/mitogen-activated protein kinase) core signalling modules to extracellular stimuli and biological responses. Biochem.Soc.Symp. *64*, 29-48.

Levchenko, A., Bruck, J., Sternberg, P.W., (2000) Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. Proc. Natl. Acad. Sci. U S A. 97(11):5818-23.

Putz, T., Culig, Z., Eder, I.E., Nessler-Menardi, C., Bartsch, G., Grunicke, H., Uberall, F., and Klocker, H. (1999). Epidermal growth factor (EGF) receptor blockade inhibits the action of EGF, insulin-like growth factor I, and a protein kinase A activator on the mitogen-activated protein kinase pathway in prostate cancer cell lines. Cancer Res. *59*, 227-233.

Sternberg, P.W. and Alberola-Ila, J. (1998). Conspiracy theory: RAS and RAF do not act alone. Cell *95*, 447-450.

Widmann, C., Gibson, S., Jarpe, M.B., and Johnson, G.L. (1999). Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. Physiol.Rev. *79*, 143-180.