MAPK signaling pathways and their recursive modularization

Rajat K. De and Losiana Nayak
Machine Intelligence Unit
Indian Statistical Institute
203 B. T. Road, Kolkata-700108, India
{*rajat,losiana_t*}*@isical.ac.in*

Abstract

Signal transduction pathways are very complex biochemical networks that include a large number of vital biomolecules and chemical compounds whose activities are dependent on member(s) involved in the network under consideration as well as other networks. Modularized study of such networks will be more appropriate for deciphering the underlying mechanisms that control physiological behavior of cell(s). Here we have proposed an algorithm for modularizing signal transduction network which is applied to MAPK pathway of D. melanogaster and H. sapiens. The idea "a member whose activity is controlled by maximum number of other members in the network under consideration, tends to be the center of a module" is used to divide a network into different modules. These modules are created by taking a certain level of complexity c into account, i.e., in an expanding module, a member having more than c relations that lie outside the module, are ignored. Recursive modularization can create modules at different stages of various complexities that finally lead to convergence of the entire network into a single module.

1. Introduction

MAPK pathway is one of the most ubiquitous signal transduction systems [11]. It is characterized by the following general path, "$Stimulus > MAPKKK > MAPKK > MAPK > Response$", where MAPKK is the kinase of MAPK and MAPKKK is the kinase of MAPKK. The symbol "$A > B$" stands for "A stimulating B". In most of the cases, MAPKKK is activated by small G proteins such as Ras and Rap1 [9, 10]. MAPK pathway is conserved in all eucaryotes and play a key role in regulation of gene expression as well as cytoplasmic activities. They transduce a large variety of external signals, leading to a wide range of cellular responses, including mating, filamentation, high osmolarity responses, cell wall remodelling, sporulation (*S. cerevisiae*), cell growth, differentiation, stress response, T-cell development, inflammation and apoptosis (mammals), morphogenesis, spatial patterning (*D. amoebae*), eye development (*D. melanogaster*), vulva induction (*C. elegans*) [9]. In this article, we have developed an algorithm that is used for creating modules from MAPK signal transduction networks. The algorithm views an entire biochemical pathway as a graph having gene products and chemical compounds as vertices and edges being different kind of interactions as shown in Fig. 1(a). An edge can be a protein-protein interaction or protein-compound interaction or a link to another map. For simplicity, here we have not taken the links to other maps into account. But how is a module defined? As there is no single definition to satisfy, we have followed certain criteria to create the modules. Here we have assumed a module is a subset of the original biochemical network which tend to be self-sufficient and have minimal dependency on the rest part of the network. Now the question arises why create modules? Since the complexity of each module is much less than that of entire pathway, we can easily analyze it. Thus analyzing all the modules generated from a pathway separately, we can have better operational view of the entire network. When all modules of the pathway are generated by the algorithm, the network can be pictorially represented with modules as nodes and interactions among modules as edges. If the network so generated is still quite complex to study, we can apply the same algorithm again on it to get a further reduced network. Thus the proposed algorithm can be recursively applied to signal transduction pathways to get reduced networks up to a certain level as desired by the user. The methodology has been applied to a sample pathway [Fig. 1(a)] and two MAPK pathways (of *D. melanogaster* [Fig. 2] and *H. sapiens*[Fig. 3]).

2. Related work

Though signal transduction units in prokaryotes and eukaryotes perform similar tasks such as switching on or off

Proceedings of the 15th International Conference on Computing (CIC'06)
0-7695-2708-6/06 $20.00 © 2006 IEEE

a required process or amplifying a certain signal, they have different underlying biochemical mechanisms. Hence properties of these systems as a whole is difficult to grasp. A sound strategy to study these complex networks is to decompose them into smaller units or modules. Such modules facilitate study of biological processes by deconstructing complex biological networks into conceptually simple entities. Definition of network modules is often based on intuitive reasoning. Methods were developed for defining biochemical network modules in an unbiased fashion. These unbiased network modules were mathematically derived from structure of the whole network under consideration [6]. One way to organize the signaling reactions, might be to separate modules with clearly defined input and output, based on pathway and cellular compartments where relationship between modules may depend on the biological state and cellular context [2]. Studying complex models of signaling pathways can be accomplished sometimes by creating operational boundaries, which do not exist in a cell. It should be noted that these modules may not correspond to conventional cell biological boundaries such as various membranes. Boundaries of such modules are often defined by functional input- output relationships. Modules may also reflect spatial locations in cytoplasm, as defined by protein scaffolds and anchors [5]. A novel criterion for defining modules can be based also on absence of retroactivity [8]. Division of a biological reaction network into smaller units highly facilitates its investigation. Ederer et. al. have proposed an algorithm to divide an ordinary differential equation (ODE) model of a biological reaction network hierarchically into functional units. They defined an activity function dependent on concentration or concentration change rate, for every compound. After performing suitable simulations, distances between compounds have been computed by comparing activities along the trajectories of the simulation. A dendrogram generated using distance information revealed the internal structure of reaction network. The original paper described detection of functional units in two models of different networks: catabolite repression in *Escherichia coli* and epidermal growth factor (EGF) signal transduction in mammalian cells [1]. Sometimes cartographic representation of complex biological networks involves identification of functional modules in first step that enables the user to view a coarse grained simplified description of the network [3].

3. Algorithm for modularization

First of all, we have defined some terms which are useful for describing our proposed algorithm.

- E: Set of all non-isolated nodes (representing gene products and chemical compounds) present in a network, i. e. each node must have atleast one relation
- M: Set of nodes present in a module (a part of network)
- k: Extension index (stage of inclusion of immediate neighbors of nodes in a module)
- M^k: Set of nodes present in a module after kth *extension*
- N_S: Set of succeeding nodes of a given node
- n_s: An individual member of N_S
- N_P: Set of preceding nodes of a given node
- n_p: An individual member of N_P
- r: Type of relation that exists between n_p and n_s (activation, binding or association, inhibition, indirect effect)
- N_R: Set of relations
- $n_r = (n_p, n_s, r)$: An individual member of N_R
- R_{np}: Total number of relations that exist with n as the preceding node
- R_{ns}: Total number of relations that exist with n as the succeeding node
- M_P: Set of permanent nodes (nodes having all their relations inside a module)
- Max: A function that detects maximum value among elements present of a given set

The total number of relations with n as either a preceding or succeeding node is given by

$$T_n = R_{np} + R_{ns} \tag{1}$$

Since R_{np} and R_{ns} are the outdegree and indegree, respectively of node n, T_n is equivalent to total degree of node n. T_n represents the total number of relations associated with node n. T_{n^k} stands for the total number of relations of node n^k that gets included in a module during kth *extension*. Like wise T_M represents a set, comprising of T_n values where $n \in M$.

The proposed algorithm starts with detection of a node n having maximum number of relations in node pool E for a given network. Considering the detected node as starting point (the starting member is always a permanent member), an initial module is created for relations r where n is either a *predecessor* or *successor*. Here an eventuality can arise where more than one node may have maximum number of relations. Then any one of the nodes (having maximum number of relations) that is encountered first by the algorithm is taken as the start point by default (followed by the others). Once a module is initialized, the total number of relations (T_n) of every individual member is considered. For a node in a module, if the number of relations lying inside the module is equal to total number of relations associated with the node, the member is considered to be *permanent*. If a node in a module has more than c relations that lie outside the module, it gets *excluded* from the module along with decreasing the previous *non-permanent* nodes' total relations (T_n) by one. These *extension* and *exclusion* continue till there is no new node or no node is under consideration i.e. all the nodes of a module become permanent. Once a member is declared permanent, it is no more in node pool E for

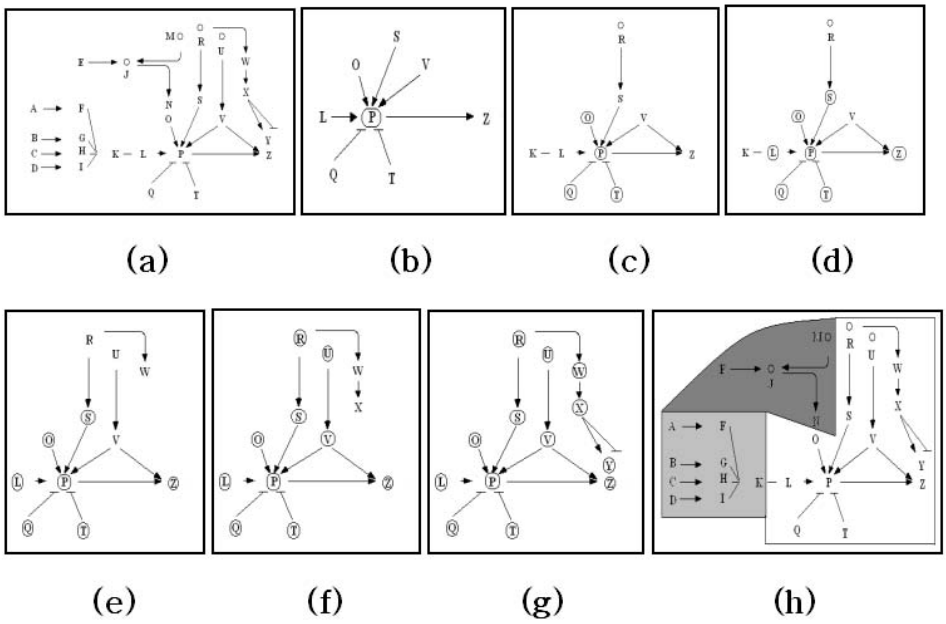

Figure 1. Modularization of a sample pathway

further consideration. Hence a single member can not be included in more than one module. Also, if a member appears more than once in a network, its positional significance is taken into account. That is, if a member X is present four times in a network, it will be considered four times as X1, X2, X3 and X4.

After successfully creating a module, the algorithm will search for another starting point and repeat the above mentioned steps to create a new module from the members of E. With construction of each module, E decreases in size, as some members are getting excluded from it due to permanency. The process continues till exhaustion of all the nodes present in E. Finally after creation of a certain module E becomes a null set and the modularization process stops. Pseudo code of the basic steps followed for modularization is given in Algorithm 1. The obtained modules and their relations with each other form a simplified network of the original complex version. When the proposed algorithm is applied to this relational network, an even simpler network with smaller number of modules is obtained. Thus the process of modularization can be applied recursively to a large network till a very simple network as desired by the user is obtained. Now the proposed algorithm is explained with an example [Fig. 1(a)]. It contains 26 nodes and 26 relations existing among the nodes. The set of relations is N_R = {(A,F,a), (B,G,a), (C,H,a), (D,I,a), (F,K,b), (G,K,b), (H,K,b), (I,K,b), (K,L,b), (L,P,a), (Q,P,i), (T,P,i), (P,Z,a), (E,J,a), (J,N,a), (M,J,a), (O,P,a), (R,S,a), (S,P,a), (R,W,a), (W,X,a), (X,Y,a), (X,Y,i), (U,V,a), (V,P,a), (V,Z,a)}.

We have to calculate total number of relations T_n for all nodes n present in the network in order to choose the node with maximum number of relations as the starting point of an originating module. Here, T_A = 1+0 = 1, T_B = 1+0 = 1, T_C = 1+0 = 1, T_D = 1+0 = 1, T_E = 1+0 = 1, T_F = 1+1 = 2, T_G = 1+1 = 2, T_H = 1+1 = 2, T_I = 1+1 = 2, T_J = 1+2 = 3, T_K = 1+4 = 5, T_L = 1+1 = 2, T_M = 1+0 = 1, T_N = 1+1 = 2, T_O = 1+0 = 1, T_P = 1+6 = 7, T_Q = 1+0 = 1, T_R = 2+0 = 2, T_S = 1+1 = 2, T_T = 1+0 = 1, T_U = 1+0 = 1, T_V = 2+1 = 3, T_W = 1+1 = 2, T_X = 2+1 = 3, T_Y = 2+0 = 2, T_Z = 0+2 = 2. So P is starting point of the module. After first extension, with immediate neighbors, the module resembles Fig. 1(b). Permanent members are denoted with a circle around them. Now we are describing the steps for determining modules from the given network.

1. For T_{Q^1}, T_{T^1} and T_{O^1}, their respective $N_S \cup N_P \subset M^1$. Hence O, Q and T became permanent members.

2. After 2nd extension, for T_{L^1}, T_{S^1} and T_{Z^1}, their respective $N_P \cup N_S \subset M^2$. So they were also considered as permanent members of the module. But $T_{K^2}-$ (number of nodes in M^2 related to K^2) $= 4 - 1 = 3 > 2$ (here c = 2). Node K has more than 2 outrelations that lie outside present module. So K cannot be a member of module P. Here we named the modules after their starting node.

3. After 3rd extension, for T_{V^1}, T_{R^2} and T_{U^2}, their corresponding $N_S \cup N_P \subset M^3$. So except W, every member present in M^3 in the module was permanent, i. e., they were

Algorithm 1 for creation of modules from a network

Ensure: $E \neq \phi$

1: Find start/central node

if $(T_n \leftarrow Max\{T_M\})$ **then**

$n \leftarrow$ start point/central node

$M_P \leftarrow M_P \cup \{n\}, E \leftarrow E - \{n\}$

$k \leftarrow 0$

end if

2: Extend module

for $(k \leftarrow k + 1)$ **do**

select nodes from N_S and N_P of n and put in M^k

end for

3: Check permanency of nodes

if $N_S \cup N_P \subset M^k$ for a node n^k **then**

$E \leftarrow E - \{n^k\}, M_P \leftarrow M_P \cup \{n^k\}$

end if

4: Exclusion of nodes

if $[T_n^k -$ number of nodes in M^k related to $n^k] > c$ **then**

$M^k \leftarrow M^k - \{n^k\}$

for $(n^{(k-1)} \notin M_P)$ **do**

$T_{n^{(k-1)}} \leftarrow [(T_{n^{(k-1)}}) - 1]$

end for

end if

5: Building a complete module

repeat

Step 2-4

until $M^k \subset M_P$

6: Creating another module from E

repeat

Step 1-5

until $E = \phi$

excluded from E.

4. 4th extension made W permanent. Like wise after 5th and 6th extension $M^6 = M_P$. Hence creation of one module was complete. Module P contains 12 permanent members.

5. The whole process is again repeated taking K, i.e., the node with maximum relations from the left over nodes of E.

6. E became Φ, a null set, after creation of 3 modules namely P, K and J.

Fig. 1(b)-(g) shows different stages during construction of module P and the modularized entire network is given in Fig 1(h).

4. MAPK cascade of *D. melanogaster*

Here the effectiveness of the proposed algorithm has been demonstrated on MAPK cascade of *D. melanogaster* (Fig. 2). For this purpose, we have considered KEGG/PATHWAY database information (http://www.genome.jp/kegg/pathway.html) [4]. It is to be mentioned here that the original figure does not include the 3 rectangles representing 3 different modules. The original figure clearly shows the existence of these 3 modules. By applying the proposed algorithm, we have also got the same 3 modules. This fact shows soundness of our proposed algorithm.

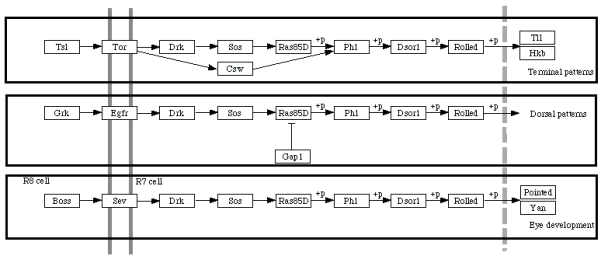

Figure 2. MAPK pathway modules of fruitfly

5. MAPK cascade of *H. sapiens*

We now consider a complex MAPK network belonging to *H. sapiens* for its modularization from KEGG/PATHWAY database. The focused pathway has 136 nodes, i.e., $|E|$ = 136, and 137 relations, i.e., $|N_R|$ = 137. By following the algorithm, 25 modules are created from human MAPK pathway. They are represented by differently shaded portions in Fig. 3. Details about these modules are given in Table 1, where column *size* and *complexity* give number of nodes and relations present in a module respectively. From the table it is clear that for c = 2, 5 modules are significant in terms of size and complexity, namely ERK, JNK, $p38$, Ras, $MEKK1$. Further the modules can be considered as nodes and their relation with each other gives a reduced, less tangled version of the initial complex biochemical network as shown in Fig. 4(a). It contains 25 nodes, each node representing a module of Fig. 3 and 36 relations, where a broken line joining 2 modules indicate indirect effect between them. When Algorithm 1 is applied to this reduced network , we got a simple network of 10 nodes and 17 relations [Fig. 4(b)]. Here each node is a module of modules and dotted lines indicate relation between 2 modules existing via other module(s). If this process is repeated, a still simpler network of 3 modules namely $MEKK1$, $Tak1$ and $IL1R$ is obtained. Finally the algorithm will leave just a single node, representing the entire network as shown in Fig. 4(c). One drawback of the algorithm is that it do not recognize isolated nodes. Hence they cannot be included into modules. Similarly the algorithm is incapable of recognizing isolated modules while

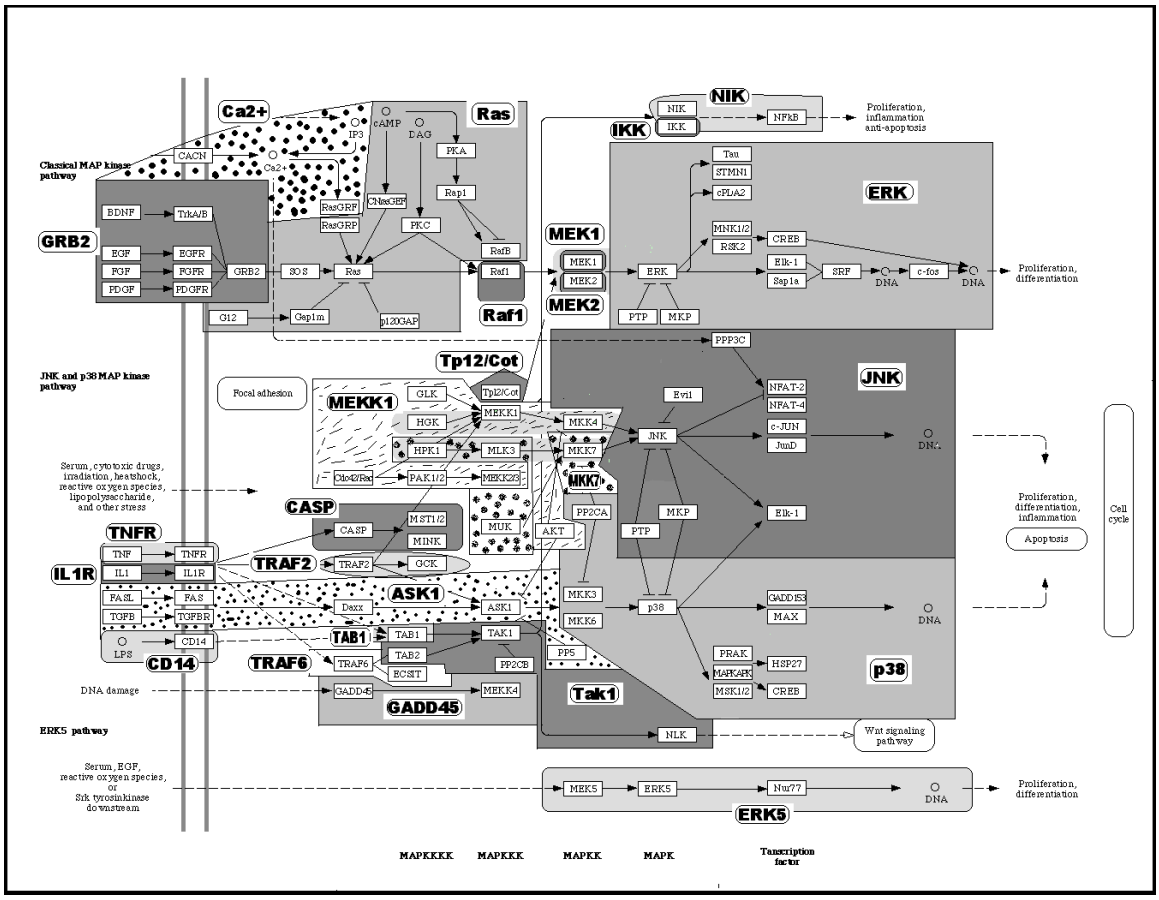

Figure 3. Modules in human MAPK signaling cascade

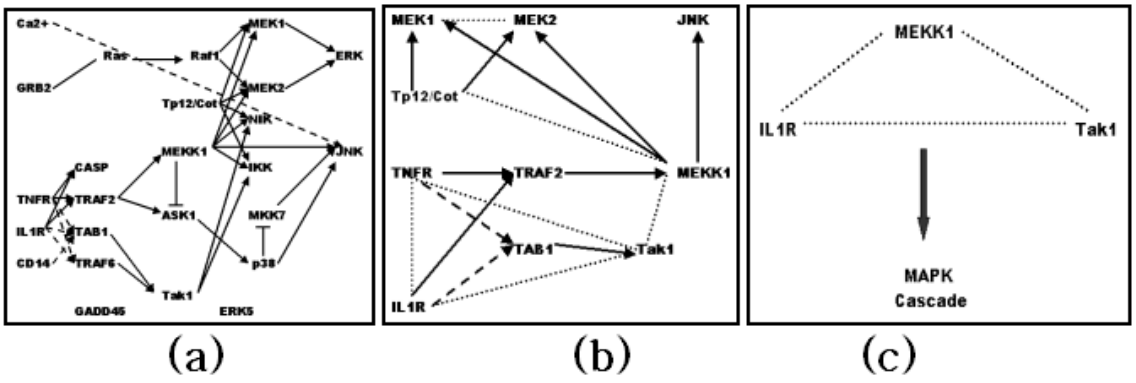

Figure 4. Recusive modularization of human MAPK pathway

Table 1. List of modules in MAPK pathway

Sl. no	Name	Size	Complexity
1	ASK1	7	6
2	Ca2+	4	3
3	CASP	3	2
4	CD14	2	1
5	ERK	15	16
6	ERK5	4	3
7	GADD45	2	1
8	GRB2	9	8
9	IKK	1	Nil
10	IL1R	2	1
11	JNK	11	11
12	MEK1	1	Nil
13	MEK2	1	Nil
14	MEKK1	8	7
15	MKK7	4	3
16	NIK	2	1
17	p38	12	13
18	Raf1	1	Nil
19	Ras	12	13
20	TAB1	1	Nil
21	TAK1	4	3
22	TNFR	2	1
23	Tp12/Cot	1	Nil
24	TRAF2	2	1
25	TRAF6	2	1

creating a simpler network from a relation network of modules.

A major question regarding biological significance arises after creation of modules in a network. We know MAPKs are widely expressed serine-threonine kinases that mediate important regulatory signals in cells. Three major groups of MAPKs exist: p38 Map kinase family, extracellular signal-regulated kinase (Erk) family, and c-Jun NH2- terminal kinase (JNK) family. Members of different MAPK groups participate in generation of various cellular responses, including gene transcription, induction of cell death or maintenance of cell survival, malignant transformation and regulation of cell-cycle progression. Over last few years, extensive work by several groups has established that MAPK pathways play critical roles in pathogenesis of various hematologic malignancies, providing new molecular targets for future therapeutic approaches [7]. After first run of our algorithm we are getting 5 major modules (ERK, JNK, p38, Ras, MEKK1) that is in accordance with above mentioned literature. But there are still 20 more modules to be accounted for. These 20 modules may have significance in analyzing MAPK pathway of *H. sapiens*.

6. Conclusions

The present article has described a new algorithm for modularizing signal transduction networks. The algorithm can recursively create modules at different levels with different complexities as desired by the user. For creating biologically significant modules, it is very important to know what will be the complexity level of a node, to be included into a module and where to stop the algorithm precisely. Effectiveness of the algorithm has been validated on a simple MAPK pathway of *D. melanogaster* and a complex one like that of *H. sapiens*. Some of the modules of MAPK pathway of *H. sapiens*, viz., ERK, JNK and p38, are already known to be biologically significant. Plus points of the algorithm lie in the fact that no 2 modules ever overlap with each other and inspite of not adding any biological constraints to the algorithm, interestingly we are getting some meaningful modules of MAPK signaling pathway.

References

[1] M. Ederer, T. Sauter, E. Bullinger, E.-D. Gilles, and F. Allgwer. An approach for dividing models of biological reaction networks into functional units. *SIMULATION*, 79(12):703–716, December 2003.

[2] N. J. Eungdamrong and R. Iyengar. Modeling cell signaling networks. *Biology of the Cell*, 96(5):355–362, June 2004.

[3] R. Guimera and L. A. N. Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, February 2005.

[4] M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa. From genomics to chemical genomics: new developments in kegg. *Nucleic Acids Research*, 34(Database Issue):D354D357, January 2006.

[5] R. N. Neves and R. Iyengar. Modeling of signaling networks. *BioEssays*, 24(12):1110–1117, December 2002.

[6] J. A. Papin, J. L. Reed, and B. O. Palsson. Hierarchical thinking in network biology: the unbiased modularization of biochemical networks. *Trends in Biochemical Sciences*, 29(12):641–647, December 2004.

[7] L. M. Platanias. Map kinase signaling pathways and hematologic malignancies. *Blood*, 101(12):4667–4679, June 2003.

[8] J. Saez-rodriguez, A. Kremling, H. Conzelmann, K. Bettenbrock, and E. D. Gilles. Modular analysis of signal transduction networks. *Control Systems Magazine, IEEE*, 24(4):35–52, August 2004.

[9] H. J. Schaeffer and M. J. Webber. Mitogen-activated protein kinases: Specific messages from ubiquitous messengers. *Molecular And Cellular Biology*, 19(4):2435–2444, April 1999.

[10] R. Seger and E. G. Krebs. The mapk signaling cascade. *The FASEB Journal*, 9(9):726–735, June 1995.

[11] M. G. Wilkinson and J. B. A. Millar. Control of the eukaryotic cell cycle by map kinase signaling pathways. *The FASEB Journal*, 14(14):2147–2157, November 2000.