Automatic decomposition of kinetic models of signaling networks minimizing the retroactivity among modules

Julio Saez-Rodriguez†,‡, Stefan Gayer*,‡, Martin Ginkel and Ernst Dieter Gilles
Max-Planck-Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, 39106 Magdeburg, Germany

ABSTRACT

Motivation: The modularity of biochemical networks in general, and signaling networks in particular, has been extensively studied over the past few years. It has been proposed to be a useful property to analyze signaling networks; by decomposing the network into subsystems, more manageable units are obtained that are easier to analyze. While many powerful algorithms are available to identify modules in protein interaction networks, less attention has been paid to signaling networks defined as chemical systems. Such a decomposition would be very useful as most quantitative models are defined using the latter, more detailed formalism.

Results: Here, we introduce a novel method to decompose biochemical networks into modules so that the bidirectional (retroactive) couplings among the modules are minimized. Our approach adapts a method to detect community structures, and applies it to the so-called retroactivity matrix that characterizes the couplings of the network. Only the structure of the network, e.g. in SBML format, is required. Furthermore, the modularized models can be loaded into ProMoT, a modeling tool which supports modular modeling. This allows visualization of the models, exploiting their modularity and easy generation of models of one or several modules for further analysis. The method is applied to several relevant cases, including an entangled model of the EGF-induced MAPK cascade and a comprehensive model of EGF signaling, demonstrating its ability to uncover meaningful modules. Our approach can thus help to analyze large networks, especially when little a priori knowledge on the structure of the network is available.

Availability: The decomposition algorithms implemented in MATLAB (Mathworks, Inc.) are freely available upon request. ProMoT is freely available at http://www.mpi-magdeburg.mpg.de/projects/promot

Contact: gayers@mpi-magdeburg.mpg.de

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

The modularity of biological processes is generally accepted. Specifically, the concept of functional units is widely prevailing (Hartwell et al., 1999; Lauffenburger, 2000); functional units are entities whose interaction with their environment is significantly smaller than their internal interaction, and can thus be seen as semi-autonomous modules. However, a general consensus definition of a module is lacking (Sauro, 2008; Wolf and Arkin, 2003).

For example, some propose modules to be a group of molecules chemically isolated from their environment or clustered according to graph theory methods. Others consider a module to be a set of elements connected in a statistically relevant fashion, or active in a specific place or during the same time scale (Nordling et al., 2007). Alternative criteria are to be evolutionary conserved or robust. According to this large list of definitions, there are an extensive number of efforts attempting to unravel the modularity of biochemical networks from different perspectives.

Since the concept of modularity can be applied to different levels of detail in a hierarchical manner (Saez-Rodriguez et al., 2005), these subunits may be comprised of anything ranging from a single domain of a particular molecule to a whole organism. Usually, the simplest units are referred to as motifs, while larger components are called modules (Wolf and Arkin, 2003). Motifs are normally no more than two or three proteins, appear repeatedly and are characterized by their structure (topology) or dynamic properties (e.g. a switch, amplifier or filter). Statistical analyses have uncovered motifs, which appear significantly often in signaling and gene regulatory networks. These motifs can be correlated with specific signal processing functions (Milo et al., 2002).

Many previous studies have analyzed the properties of large biological networks, among them the modularity, using graph theory techniques. It has been shown that metabolic networks (Ravasz et al., 2002) as well as protein networks in yeast (Jeong et al., 2001; Rives and Galitski, 2003) and human (Rodriguez-Caso et al., 2005) have a modular, hierarchical structure using different clustering methods. A remarkable effort to decompose networks into modules has been developed by Newman and colleagues (Newman, 2006; Newman and Girvan, 2004), whose approach has been applied by others to metabolic (Guimera and Amaral, 2005) and protein–protein networks (Chen and Yuan, 2006). Particularly appealing is their idea of quantifying the degree of modularity, considering it as an objective function which is subsequently optimized (Guimera and Amaral, 2005; Newman, 2006). Their definition is based on the idea that the modularity is high if there are fewer edges than expected between modules, i.e. if there are fewer than in a randomly generated network.

The studies mentioned above analyze large biological networks as graphs, and aim to unravel general structural principles. Even though we will make use of some of their ideas (in particular, the approach of Newman), the goal of this work is very different: to decompose a given (kinetic model of a) signaling network into units in order to facilitate its analysis, thus providing a rationale for modular models.

To that extent, our approach is connected to the work of Ederer et al. (2003), where also kinetic models were analyzed by clustering methods. There, however, typical trajectories of the concentrations of proteins were used to cluster compounds into modules, while here the focus is on the network structure. Noteworthy efforts towards
a modular understanding of biochemical networks have also been performed in the field of metabolic control analysis (MCA), which has proved to be successful in the analysis of metabolic networks, and has been extended to signal transduction networks (Kahn and Westerhoff, 1991; Khodolenko et al., 1997).

In this contribution, we present a novel approach to automatically decompose biochemical networks into modules using a criterion we have previously introduced (Saez-Rodriguez et al., 2005), namely the absence of retroactivity among modules (see Fig. 1, Sauro, 2008). Retroactivity, the effect of downstream elements on the state of the upstream element, is an essential and well-studied issue in systems theory. It is also significant in biochemical systems: retroactive connections modify the functioning of a module, which is then dependent on what is downstream of it, hindering the analysis of signaling systems and the design of synthetic biological devices (Del Vecchio et al., 2008). Furthermore, it challenges the interpretation of evolution as a ‘tinkerer’, which connects different blocks in various ways until the desired behavior is obtained (Alon, 2007). Different cases leading to retroactive connections in biochemical systems have been extensively studied elsewhere (Del Vecchio et al., 2008; Saez-Rodriguez et al., 2005; Sauro and Khodolenko, 2004).

Our approach aims to define modules so that the number of retroactive interconnections is minimized, thus obtaining subsystems as uncoupled as possible. This framework, based on network theory, allows us to precisely characterize the connections among components. The method requires only the structure of the network. However, if information on the actual strength of the interactions between elements is available, it can be used to perform a more refined analysis. We begin the article by presenting a mathematical framework to characterize retroactive connections in biochemical systems. We then use the concept of retroactivity to define a modularity coefficient, which we optimize using methods of community structure detection. Finally, we implement the algorithm and test it with different cases of increasing complexity, demonstrating the applicability of our approach on realistic large signaling transduction cases.

2 METHODS

2.1 The absence of retroactivity as a criterion to demarcate modules

Since engineering sciences are used to working in a modular manner, it is tempting to approach the definition of biological modules from a technical perspective. From a system-theoretical point of view, an interesting criterion might be the definition of elements where both the input and the output are unidirectional. This is actually the form in which most technical systems are devised, facilitating their analysis and design (Del Vecchio et al., 2008; Saez-Rodriguez et al., 2005).

Consider a signaling network as a general non-linear dynamical system described by a set of ordinary differential equations (ODEs) of the form

$$\frac{d\vec{c}}{dt} = f(\vec{c}, \vec{u}, \vec{p}),$$  

(1)

where $d\vec{c}/dt \in \mathbb{R}^n$ is the vector of the ‘balances’ of the concentrations $c_i$, $\vec{u}$ the vector of inputs and $\vec{p}$ the vector of parameters. A vector of outputs $y=g(\vec{c})$ may also be defined. The goal would be to decompose $\vec{c}$ into two sub-systems $\vec{c}_1$ and $\vec{c}_2$ so that

$$\vec{c}_1 = f(\vec{c}_1, \vec{u}, \vec{p}),$$

$$\vec{c}_2 = f(\vec{c}_1, \vec{c}_2, \vec{u}, \vec{p}).$$  

(2)

If modules are connected in the same form as $\vec{c}_1$ and $\vec{c}_2$, they fulfill the requisite of independence postulated for functional units (Hartwell et al., 1999): the behavior of $\vec{c}_1$ is only influenced by the input $\vec{u}$ and is independent of what is downstream of it. Importantly, decoupled units can be analyzed in a relatively straightforward manner by means of systems theory’s tools. Actually, the decomposition into decoupled systems of the form of Equation (2) is a well-studied problem in the field of systems theory (Sontag, 1998).

Unfortunately, a hallmark of biochemical—i.e., signaling—networks is the high degree of coupling. Hence, a clean decomposition in the form of Equation (2) is in most cases not possible. We therefore introduce a subtly different, more relaxed definition, which we shall call the absence of retroactivity, illustrated in Figure 1: two modules $\vec{c}_1$ and $\vec{c}_2$ are connected without retroactivity if there is no pair of elements (components), one in each module, which influence each other, i.e.

$$\vec{c}_1 \not\rightarrow \vec{c}_2 \wedge \vec{c}_2 \not\rightarrow \vec{c}_1.$$

(3)

The key difference with the previous concept of decoupling is that here, instead of the global decoupling between the modules as a whole imposed in Equation (2), we just require a local decoupling between each element of one module and each element of the other module.

Even the relaxed concept of absence of retroactivity (Equation (3)) may not be completely fulfilled by biological systems. Therefore, the algorithm to decompose signaling networks which we shall see later on will rely on methods optimizing the modularity (i.e., finding the set of modules so that the number of retroactive connections among modules is minimized), rather than pursuing a clean separation.

2.2 Retroactivity and network theory

The particularities of biochemical systems should be taken into account. A biochemical system is often described by a set of ordinary differential equations (ODEs) of the form

$$\frac{d\vec{c}}{dt} = \vec{f}(\vec{c}, \vec{u}, \vec{p}),$$

(4)

which is a special form of Equation (1). Here, $\vec{f}(\vec{c}, \vec{u}, \vec{p}) \in \mathbb{R}^n$ is the vector of the $n$ reactions, and $\vec{f}$ is the stoichiometric matrix (Heinrich and Schuster, 1996). This structure of the differential equations will be helpful to cleanly characterize, from a biochemical point of view, the coupling among modules.

The structure of Equation (4) emerges naturally from a description of signaling networks following the network theory (Gilless, 1998), which is also a convenient framework for a modular approach. There, one describes a particular system (in our case a biochemical network) as a combination of two types of elementary units: components, which have storages of physical quantities (here, concentrations) and coupling elements, which describe the...
Table 1. Types of connections between two species as a function of $J^R$ and $N^{CT}$

<table>
<thead>
<tr>
<th>$J^R_{ij}$</th>
<th>$N^{CT}_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

For $\theta = 0$, one would get $J'$ as for the structural retroactivity (Equation (5)). Increasing the values of $\theta$, elements in $J^I$ become 0 which were before 1, and thus the coupling of the network decreases. Finally, using either Equations (5) or Equation (6) to define $J'$ the symmetric matrix $J^R \in [0, 1]^{n \times n}$, with

$$J^R_{ij}^\theta = \begin{cases} 1 & \text{if } J_{ij} \neq 0 \text{ and } J_{ji} \neq 0 \\ 0 & \text{else} \end{cases}$$

defines the presence of a bidirectional influence via potentials between two compounds. However, it is not possible to characterize the different possible couplings only with $J'$ (Table 1), since it provides no specific information about the currents. This information is encoded in the stoichiometric matrix, which can be interpreted as a reaction map in which the storages are the nodes and the reactions are the directed edges (Famili and Palsson, 2003). Taking $-N^{CT}$ yields a compound map in which the reactions are the nodes and the storages are the directed edges. We define the symmetric matrix $N^{CT} \in [0, 1]^{n \times n}$, so that

$$N^{CT}_{ij} = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} N^{CT}_{ij} \neq 0 \\ 0 & \text{else} \end{cases}$$

The entries of $N^{CT}$ indicate whether two storages are connected via a current or not. Now, using $J'$ and $N^{CT}$, the different possible connections between two storages can be unambiguously characterized (Table 1). A subtle point is the case where $J^R_{ij} = J^R_{ji} = 1$, but $N^{CT}_{ij} = N^{CT}_{ji} = 0$. This case presents a bidirectional coupling in terms of the states [Equation (1)] but not of the elements of the network theory, since compounds are linked to the reactions via unidirectional connections. We will here optimize the retroactive connections among states that are stored in the matrix $J^R$, which we shall call the retroactivity matrix. $J^R$ can be interpreted as an adjacency matrix of a graph where the compounds are the nodes, which are connected if there is a bidirectional coupling in the corresponding model. Thus, by defining modules so that the number of connections among modules are minimized, we will minimize the retroactive connections. The methods of community structure detection are an efficient approach to identify such modules.

### 2.3 Retroactivity matrix

We would now like to characterize mathematically the currents and potentials among storages and reactions. A reaction $v_i(\vec{u}, \vec{p})$ is affected by a storage $c_j$ via a potential $J_{ij} = \partial v_i/\partial c_j \neq 0$. $\delta$ is known as unscaled elasticity in the field of Metabolic Control Analysis (Heinrich and Schuster, 1996). Conversely, a reaction $v_i(\vec{u}, \vec{p})$ affects a storage $c_j$ if $J_{ji} \neq 0$. We note that by multiplying $N$ and $\delta$ we obtain the Jacobian $J = \partial N v_i/\partial c_j$ of the network, which defines the effect of each storage on each other: the concentration of storage $c_j$ depends on the concentration of storage $c_i$ if $J_{ij} \neq 0$. The indicator matrix of the Jacobian $J \in [0, 1]^{n \times n}$, so that

$$J_{ij} = \begin{cases} 1 & \text{if } J_{ij} \neq 0 \\ 0 & \text{else} \end{cases}$$

determines the existence of an effect from $c_j$ to $c_i$. It is important to note that $J$ depends only on the structure of the system [Equation (4)], but does not consider the (possibly large) differences in the magnitude of the interactions. We shall call the retroactivity exclusively related to the structure structural retroactivity. For a more refined decomposition, we can (if available) quantify the retroactivity matrix using, for example, typical simulation conditions. Subsequently, we can define which retroactivities are important as follows: for each time point, we compare the value of all elements $J_{ij}$ (which represents the influence of $c_j$ on $c_i$) and $\Delta_j$, the maximum of $J_{ij}, J_{ji}, J_{ik}, \ldots$ representing the most important influence on $c_i$. We define a certain $J_{ij}$ to be relevant (and thus $J_{ij}^R = 1$), if, at least for one time point, $J_{ij}$ is larger than a certain fraction $\theta$ of the maximum $\Delta_j$,

$$J_{ij}^R = \begin{cases} 0 & \text{if } \frac{J_{ij}}{\Delta_j} \leq \theta \forall t \\ 1 & \text{else} \end{cases}$$

### 2.4 Methods from community structure detection

Community structure detection follows the philosophy that networks possess a natural modular structure that has to be discovered. Newman and colleagues define modules as entities that contain more edges than one would expect. This definition requires the comparison of the network’s real structure with an expected structure. Given the real structure of a network represented by the adjacency matrix $A$, the number of edges we can expect to lie between the nodes $i$ and $j$ equals $k_i k_j / 2m$, where $k_i = \sum_j A_{ij}$ is the number of edges connected to node $i$, and $m$ is the total number of edges in the network $(m = (1/2) \sum_k k_i)$. Therefore, the expected structure of a network can be represented by a matrix $P$ with

$$P_{ij} = k_i k_j / 2m.$$ 

Thus, for a network containing $n$ vertices and $m$ edges the modularity can be defined as (Newman, 2006)

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - P_{ij}) \delta(M_{ij}, M_{ji}).$$

where

$$\delta(M_{ij}, M_{ji}) = \begin{cases} 1 & \text{if node } i \text{ and node } j \text{ are in the same module} \\ 0 & \text{else} \end{cases}$$

The leading factor $1/2m$ is just conventional for compatibility purposes with other definitions of modularity (Newman, 2006) and confines $Q$ to values between $-1$ and $1$. This definition provides a quantitative measure for the modularity, which is subsequently optimized (see Section 2 in
Supplementary Material). Importantly, the optimal number of modules is also a result of the solution of the optimization problem. We will describe how to apply this methodology to our task of determining biochemical modules as uncoupled as possible in the next section.

3 ALGORITHM
As outlined in the Methods section, we define the matrix $J^{IR}$ to characterize the bidirectional couplings among storages, and it is used as the adjacency matrix which is required for the calculation of the modularity, i.e. we take $A = J^{IR}$ in Equation (10). Note that we do not impose any restrictions on the unidirectional connections between modules, as we exclusively want to minimize bidirectional interactions. Several methods have been proposed for the optimization of the modularity $Q$. Among them are simulated annealing (Guimera and Amaral, 2005), extremal optimization (Duch and Arenas, 2005) and methods that are based on the evaluation of the eigenvalues and eigenvectors of matrix $B = A − P$ (where $P_{ij} = \delta_{ij}/2m$) (Newman, 2006). These methods, which are described in detail in the Supplementary Information, were implemented and tested for the different cases discussed below.

3.1 Sensitivity analysis of modularity
Some nodes (storages) are highly connected to other members of the same module. Therefore, they are placed in a very central position within a module and others are located in the ‘periphery’ and are thus not as ‘bound’ to the module. Accordingly, moving a central node to another module would produce a large decrease of the modularity, while moving a more peripheral would have less impact. To study this property, we define a sensitivity of the modularity $\Delta Q$ with respect to the shift of a particular node $i$ from a module $M_1$ to a module $M_2$; mathematical details can be found in the Supplementary Material. This sensitivity will be useful to validate or correct the final results of the optimization: we are dealing with a combinatorial problem for which an optimal solution can not be guaranteed, and the solutions might be suboptimal, particularly for large networks. Furthermore, $\Delta Q$ can be used to support some of the optimization algorithms (see Supplementary Material).

If a component in the biochemical network is not coupled bidirectionally to any other element, it has no entries in the adjacency matrix. Therefore, in which module it is located has no effect on the modularity $Q$ ($\Delta Q = 0$, see Equation (10)). To efficiently compute the modules, these nodes have to be removed before the optimization procedure. After the modules have been defined, these nodes can be assigned to a certain module by incorporating them into the module they are most connected to through unidirectional couplings.

4 IMPLEMENTATION
The algorithm was implemented in MATLAB (www.mathworks.com). Models can be loaded in the Systems Biology Markup Language (SBML) format (Hucka et al., 2003). The optimization is run in MATLAB and the algorithm not only provides the modules but also characterizes, for analysis and visualization purposes, all the connections as either (i) unidirectional potential, (ii) unidirectional current or (iii) bidirectional. The results can be exported in a text format compatible with Pajek (Batagelj and Mrvar, 2002), where the results can be visualized. In addition, the file can be read (together with the SBML file describing the model) in ProMoT, a tool supporting modular modeling (Ginkel et al., 2003). ProMoT can also export the modular model in SBML format, using the compartments to demarcate the different modules.

The algorithm was tested with several models of increasing complexity describing epidermal growth factor (EGF)-mediated signaling. We start with a relatively simple model of the EGF-induced MAPK cascade of Brightman and Fell (2000). Next, in increasing order of complexity, we address the EGF-signaling map of Oda et al. (2005), and the EGF-induced MAPK cascade of Schoeberl et al. (2002). The description of the models and corresponding biological systems is rather succinct due to space limitations; for a detailed explanation, the reader is referred to the original publications of the models.

4.1 Simple model of the EGF-induced MAPK of Brightman and Fell
As the first example, we consider a relatively simple model of the EGF-induced MAPK cascade of Brightman and Fell (2000).

Fig. 2. Decomposition of the EGF-induced MAPK cascade of Brightman and Fell (2000). Left: six modules are identified, connected only through unidirectional connections. Green lines represent control by a potential, blue lines denote control by a current and black lines bidirectional (potential + current) coupling. All nodes of the same color represent a distinct module. The left figure was created with Pajek (Batagelj and Mrvar, 2002). The right panel shows the results of sensitivity analysis, where $\Delta Q$ is mapped to the darkness in the heat map. The heat map illustrates for each compound how much of the optimal modularity $Q$ would decrease if the compound was assigned to another module. White squares denote $\Delta Q = 0$, i.e. the optimal module for the corresponding compound, whereas dark squares indicate a strong decrease of $Q$. It can be thus seen that ShcGS (Shc bound to Grb2 and Sos) and Ras bound to GTP (RasGTP) are hubs of their respective modules. R2i, p38 and Raf1 have no sensitivity as they are not coupled with bidirectional connections to any element.
The model, comprising 25 states (compounds), describes ligand binding to the EGF receptor, binding of the proteins Grb2, Sos and Shc to the receptor, which activates Ras, leading in turn to the activation of the Raf/MEK/ERK MAPK cascade. All reactions describing phosphorylation and dephosphorylation processes follow Michaelis–Menten kinetics (providing unidirectional connections suitable to define the borders of the modules), whereas all other reactions are modeled using mass action kinetics.

This model includes three compounds that have no retroactive connections to any other compound: (i) the internalized EGFR dimer (R2i), (ii) Grb2 bound to phosphorylated Sos (GSP) and (iii) the active form of Raf (Raf*). We therefore proceeded as described in the Algorithm section and performed the modularity analysis in a reduced model without them. These elements were assigned to the modules with which they had the most (non-retroactive) connections. The algorithm uncovers, using either of the methods available, six modules, corresponding to specific signaling events of the system: (i) binding of ligand to receptor, (ii) receptor dimerization, (iii) formation of the intermediate compounds with Grb2, Sos and Shc, (iv) activation of Ras and Raf, and activation of the kinases (v) MEK and (vi) ERK. All modules are connected exclusively by unidirectional interactions, meaning that the solution is optimal. Sensitivity analysis showed that ShcGS (Shc bound to Grb2 and Sos) and Ras bound to GTP (RasGTP) are central members of their modules, as would be expected since they are the elements bound to other storages with the highest number of retroactive connections (Fig. 2).

4.2 Large model of the EGF-signaling network of Oda et al.

We now consider the comprehensive map of EGF signaling by Oda et al. (2005), which is probably the largest model of a signaling network within a kinetic (stoichiometric) framework. The network describes the signaling events triggered by EGF as well as 12 other related ligands, which activate the EGF receptor family, and leads to activation of multiple signaling intermediates and eventually several transcription factors. The enzymatic reactions are described as the conversion of substrate into product, with the enzymes serving as modifiers of the reaction. This can be interpreted as an abstraction of Michaelis–Menten kinetics: the enzyme is not sequestered by the reaction, but no specific law is given. This hinders simulations of the model, but contains enough information to study its structural retroactivity. Furthermore, most reactions are defined as irreversible. These reactions are typically phosphorylation events, and for virtually any phosphorylation a counteracting dephosphorylation occurs. We therefore added these backward reactions. The methods which performed at best in this case was the extreme optimization combined with the sensitivity analysis (see Supplementary Information). In spite of the remarkable size and complexity of the network (comprising over 200 reactions and 300 species), the algorithm is able to decompose it into 55 modules, see Figure 3.

The results were correct in the sense that coupled elements were assigned to the same module but, in many cases, completely decoupled elements were grouped together (e.g. module 20 lumped
together MKK4 with Stat1 and Stat3). However, the separation of decoupled elements into two modules increases the degree of modularity. Thus, these results are most likely due to the fact that the optimization algorithms were not able to find the global optimum, and instead remained in a local optimum. As in the previous case, all nodes that had no retroactive connections were removed from the adjacency matrix before modularization and assigned automatically to those modules to which they had the most links.

4.3 Entangled model of the EGF-induced MAPK of Schoeberl et al.

The model of the EGF-induced MAPK cascade of Schoeberl et al. (2002) describes the biological processes described in the work by Brightman and Fell (2000). In addition, it includes internalization, a process by which the receptors are retrieved from the cell surface and moved into special compartments known as endosomes. Thus, internalization increases the complexity of the model. This extension, together with the description of enzymatic reactions following mass action kinetics (and thus including enzyme-substrate complexes neglected in the simplified description of the previous cases), gives rise to a total of 94 states.

Even though this model comprises only a subset of the biology described in the Oda et al. (2005) network, and contains fewer compounds, we consider this case last because it creates the biggest obstacle for the algorithm. The difficulty arises from the high grade of entanglement of the model (Fig. 4) and because all reactions are modeled with mass action kinetics.

The modules have been demarcated by hand elsewhere (Saez-Rodriguez et al., 2005). Remarkably, our algorithm proposes as a slightly suboptimal solution (using the leading eigenvector method, see Supplementary material) the same number of modules and, furthermore, the distribution of the nodes in the modules is very similar in both cases. By applying a more advanced optimization algorithm (multiple eigenvector method or extreme optimization), a slightly higher modularity could be achieved based on the decomposition of the Ras–Raf module into two modules, corresponding to the surface- and endosome-related pools of Ras and Raf.

This model is fully parametrized, allowing for simulation. We therefore used this model as an example to perform a decomposition based on the actual values of the interactions through the simulations, as described in the Methods section. We used the standard conditions of the original work Schoeberl et al. (2002), namely a stepwise input of EGF ligand of 10 nM. By increasing the threshold \( \theta \), which defines whether an interaction is significant, the coupling of the network decreases and the number of modules increases. For instance, \( \theta = 0.001 \) gives rise to nine modules, while \( \theta = 0.01 \) gives rise to 11 modules. Intriguingly, the resulting modules could not be interpreted intuitively based on the structure of the network: the modules did not necessarily break where less number of connections are present. This is likely due to the presence of different dynamic modes within the modules, with some interactions significantly stronger than others. This result suggests that structure is essential but not everything (Ingram et al., 2006), and much information is encoded in the actual quantitative magnitudes describing signaling events.

5 DISCUSSION AND CONCLUSION

This contribution presents an approach, inspired by systems theory, that provides a theoretical framework to analyze signaling networks in a modular manner. Most studies consider these systems as protein interaction networks. Here, however, we rely on the signaling system being described as a biochemical network. Thus, our approach uses more refined information and acts on the kinetic formalism of most mathematical models. Our methodology is very closely related to and based on the same concept (retroactivity) as that of Del Vecchio et al. (2008), but is complementary in the sense that here the focus is on how to partition a network, while there it is to study and understand the effects of retroactive connections. It is also different in that we use the bipartite formalism of network theory, and therefore can characterize more precisely in a chemical sense the nature of bidirectional and unidirectional connections. This allows us to differentiate between an enzymatic effect and an
irreversible reaction, both cases of unidirectional effects between states.

Our methodology can work at two levels of detail: either at the level of the pure structure of the model (with the advantage that much less information is required) or at a more detailed level, where dynamics can provide subtle additional information. The procedure was tested with three case studies, focusing on the Conflict of Interest für Bildung und Forschung (BMBF). Forschungsgemeinschaft (DFG), FOR521, and Bundesministerium Funding the article.

REFERENCES

The integration of the decomposition method presented here into BioModels provides a unique platform for a modular analysis. One can take a model, e.g. in SBML from a database such as BioModels (Le Novere et al., 2006), "blindly" decompose it into sensible subunits, and easily create models of isolated subunits and combinations of them for analysis.

ACKNOWLEDGEMENTS

Thanks to Eduardo Sonntag for useful discussions and to Jooho Chung, Laura Sonntag-Kleiman and William Chen for proofreading the article.

Funding: The authors thank support from the Deutsche Forschungsgemeinschaft (DFG), FOR521, and Bundesministerium für Bildung und Forschung (BMBF).

Conflict of Interest: none declared.

REFERENCES


