Research article

A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks

Holger Conzelmann*[1], Julio Saez-Rodriguez[2], Thomas Sauter[1], Boris N Kholodenko[3] and Ernst D Gilles[1,2]

Address: [1]Institute for System Dynamics and Control Engineering, University of Stuttgart, Pfaffenwaldring 9, 70569 Stuttgart, Germany, [2]Max-Planck-lnstitute for Dynamics of Complex Technical Systems, Sandtorstr. 1, 39106 Magdeburg, Germany and [3]Department of Pathology, Anatomy and Cell Biology, Thomas Jefferson University, 1020 Locust St., Philadelphia, PA 19107, USA

Email: Holger Conzelmann* - Conzelmann@isr.uni-stuttgart.de; Julio Saez-Rodriguez - saezr@mpi-magdeburg.mpg.de; Thomas Sauter - sauter@isr.uni-stuttgart.de; Boris N Kholodenko - Boris.Kholodenko@jefferson.edu; Ernst D Gilles - gilles@mpi-magdeburg.mpg.de

* Corresponding author

Abstract

Background: Receptors and scaffold proteins possess a number of distinct domains and bind multiple partners. A common problem in modeling signaling systems arises from a combinatorial explosion of different states generated by feasible molecular species. The number of possible species grows exponentially with the number of different docking sites and can easily reach several millions. Models accounting for this combinatorial variety become impractical for many applications.

Results: Our results show that under realistic assumptions on domain interactions, the dynamics of signaling pathways can be exactly described by reduced, hierarchically structured models. The method presented here provides a rigorous way to model a large class of signaling networks using macro-states (macroscopic quantities such as the levels of occupancy of the binding domains) instead of micro-states (concentrations of individual species). The method is described using generic multidomain proteins and is applied to the molecule LAT.

Conclusion: The presented method is a systematic and powerful tool to derive reduced model structures describing the dynamics of multiprotein complex formation accurately.

Background

Receptor-mediated signal transduction is the subject of intense research since it plays a crucial role in the regulation of a variety of cellular functions. The ligand binding to a receptor triggers conformational changes that allow for receptor dimerization and phosphorylation of numerous residues. The subsequent formation of multiprotein signaling complexes on these receptors and their scaffolding adaptor proteins initiates a variety of signaling pathways. The number of feasible different multiprotein species grows exponentially with the number of binding domains, and can easily reach thousands or even millions [1,2]. In the past years, a large number of mathematical models have attempted to describe signaling phenomena and to get deeper insights into the dynamics of cellular responses [3-11]. Most of these models did not consider the combinatorial variety of all possible species and interactions [1,2]. The obvious advantage of models neglecting

the combinatorial complexity is there are less ordinary differential equations (ODEs) that have to be considered. For example, if all possible species were included in the model of Schoeberl et al. [6], the number of variables would grow from around 100 to almost 40000 [2]. However, there is no general method to decide a priori which species play a crucial role and which can be neglected. Recent investigations indicate that the structure of reduced models which focus on the most important species, is highly dependent on the kinetic parameters [12]. The procedure described in [12] is based on deleting species from a full network model to find the smallest network that reproduces the dynamics of a set of observed quantities.

In 1998, the stochastic simulation tool StochSim was developed to handle the problem of combinatorial complexity [11,13]. The number of multi-state protein complexes and chemical reactions could reach millions in deterministic models of bacterial chemotaxis, and StochSim was employed to circumvent this problem [11]. The use of StochSim is especially practical for networks in small volumes with low numbers of molecules, where a stochastic algorithm more closely describes the physical reality than a deterministic algorithm. For large networks with hundreds of different proteins the computational cost would be extremly high increasing proportionally to the number of molecules squared. An alternative deterministic approach considers a complete mechanistic description that accounts for all possible species and reactions [1,14]. In the software tool BioNetGen, an algorithm for rule-based modeling is implemented [15]. However, the computational cost also appears to be very high, making it difficult to analyze large networks with a vast number of states.

Recently, a new approach has been introduced where a mechanistic picture of all possible states is substituted by a macro-description that follows the occupancy levels and other characteristics of individual domains, e.g. the phosphorylation states of these sites [16]. These quantitative indicators of the system (levels of occupancy) are referred to as macro-states. A modeling description using macro-states accounts for limitations in the current techniques to measure concentrations of individual multiprotein species. The results of common biological measurements (e.g. immunoprecipitation followed by western blotting) correspond to cumulative quantities like levels of occupancy or degrees of phoshorylation. Thus, the introduction of these and similar quantities into modeling facilitates the comparison of model variables with experimental readouts. This approach adopts the point of view of Pawson and Nash [17] that molecular domains instead of molecules are the fundamental elements of signal transduction. In their review they discuss that the regula-tion of many different cellular processes and functions require the use of protein interaction domains, and that aberrant interactions can induce abnormal cellular behaviour and diseases. We show that besides these considerations this macroscopic description also provides a number of mathematical benefits. In their work Borisov et al. demonstrate that for scaffold proteins with independent binding sites and scaffolds with one controling domain, the dynamics of these macroscopic states can be accurately described by reduced models [16]. However, a methodology to derive the reduced model equations for any scaffold with a more complex pattern of domain interactions is still missing. In this contribution, we will introduce a new systematic approach formalizing and extending the model reduction presented in [16]. Our method not only answers the questions whether a mathematically accurate model reduction is possible in a given system, but also how many and which equations are required. Our approach is based on a state space transformation linking the approaches provided in [15] and [16]. This link is achieved by a hierarchical structure of the state space transformation introducing mesoscopic state variables (describing different levels of detail) in addition to macroscopic and microscopic states defined in [16]. In contrast to many other model reduction methods our approach is independent of exact numerical values. Often, only qualitative biological knowledge about domain interactions is needed to derive the reduced model equations. For instance, data reported for receptor tyrosine kinases provide qualitative knowledge about domain interactions [18]. The ligand binding to these receptors controls their dimerization and the rate of phosphorylation of docking sites, which bind multiple partners. Here, we address the question of how different domain interactions influence the structure of a reduced model and how many states are required in order to describe the system dynamics accurately. We demonstrate which signaling systems can be described using only macro-states, and which additionally require the consideration of mesoscopic states.

The contribution is structured as follows. First, we introduce our method using simple examples reconsidering some of the cases described in [16] in order to show that our method provides the exact same results. For example, if the binding sites are independent, or only one of the domains controls some others (which is the case of a scaffolding adaptor protein phosphorylated by receptor kinase, e.g., insulin receptor substrate phosphorylated by the insulin receptor [19]) it can be shown that the number of states can be strongly reduced. Additionally, we confirm our results via dynamical simulations and discuss the advantages of the reduced models. Second, we generalize these results and also consider an example with a more complex pattern of domain interactions, and finally, we

apply the method to a real scaffold protein, namely the adaptor molecule LAT (Linker for activation of T-cells).

Results

In order to introduce and discuss our method in an descriptive manner, we first want to introduce some definitions and outline the principles shortly. Afterwards, we will consider some simple generic examples (scaffold proteins with three and four binding domains) in more detail. Of course, these examples might also be treated without recourse to our theoretical developments, because the number of states is relatively small. But precisely because some of the results may appear intuitive, this shows the potential of our method which is also applicable to other more complicated problems. Additionally, we generalize two important cases to scaffolds with n binding sites (namely independent binding domains and domains that are controlled by one controlling site). We also provide the generalized transformation matrix for proteins with n binding sites and exemplify our method considering the scaffold molecule LAT (Linker of activation in T-cells).

Definitions

In the following, we will consider a receptor or scaffold protein R with n distinct binding domains i (i.e $i = 1, 2, ..., n$). Each of these domains i can bind m_i different effector proteins E_i^j with $j = 1, ..., m_i$, and hence can be in $m_i + 1$ different states (unoccupied or occupied by one of the m_i effectors). The number of feasible micro-states for such a scaffold protein is

$$\prod_{i=1}^{n} (m_i + 1). \qquad (1)$$

The number of possible reactions that have to be considered is

$$\sum_{i=1}^{n} m_i \prod_{k=1, k \neq i}^{n} (m_k + 1). \qquad (2)$$

E_i^j denotes the j-th effector which can bind to the i-th binding domain i. We introduce labels for the status of the different binding sites such as 0 for an unoccupied binding domain and E_i^j for the domain i occupied with the j-th effector. Alternatively, the different states of a domain can also be enumerated. For instance, the two representations $R[E_1^4 , E_2^1 , 0, E_4^2]$ and $R[4, 1, 0, 2]$ for a scaffold protein with four binding sites are equivalent. In addition, the representation $R[a_1, 0, 0, 0]$ denotes all scaffold pro-

teins R with unoccupied binding domains 2 to 4 independent of the status of domain 1, etc.

All these definitions shall be clarified by considering a simple scaffold protein R with two domains 1 and 2 ($i = 1, 2$). Assuming it is known that binding domain 1 can bind three different effectors E_1^1, E_1^2 and E_1^3 ($m_1 = 3$ and $j = 1, 2, 3$), and that the second domain 2 can bind two different effectors E_2^1 and E_2^2 ($m_2 = 2$ and $j = 1, 2$). The number of feasible states of domain 1 is $m_1 + 1 = 4$, namely either unoccupied or occupied by one of the three effectors. According to these considerations the number of different states of domain 2 is $m_2 + 1 = 3$. The total number of complexes results from $\prod(m_i + 1) = 4 \cdot 3 = 12$. In order to calculate the number of reactions that may occur in this example we first assume that domain 1 is unoccupied. Now three different effectors may bind to this domain, which corresponds to three different reactions. At the same time binding domain 2 can be unoccupied or occupied by one of the two effectors E_2^j. Hence, one has to consider $3 \cdot 3$ reactions describing the occupation of domain 1. Now we repeat these considerations for domain 2. This results in $2 \cdot 4$ reactions. The total number of reactions in our example is

$$\sum_{i=1}^{n} m_i \prod_{k=1, k \neq i}^{n} (m_k + 1) = 3 \cdot 3 + 2 \cdot 4 = 17. \qquad (13)$$

Reduction method

Our method can be divided into three essential steps. First, we start generating a complete mechanistic description of the considered scaffold or receptor like described in [1,14,15]. This step also includes the incorporation of qualitative information about domain interactions. Second step is the introduction of macro-states (levels of occupancy of each domain) using a state space transformation following a hierarchical pattern. The transformed system is reduced in a third step by eliminating all equations which do not influence the dynamics of the output variables $\vec{y}$. In the following we assume that one of the goals of signaling pathway modeling is to accurately describe the dynamics of the macro-states. Hence, we choose the levels of occupancy as output variables $\vec{y}$. This choice seems to be reasonable since these values correspond to experimentally verifiable quantities as discussed before. Additionally, similar examples can be found in literature where the same or very similar output quantities
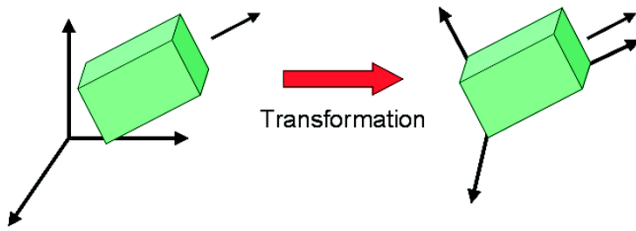
Figure 1
State space transformation. The idea of our method can be described more easily by considering a mechanical example: In order to model the movement of a mass in space one has to choose a certain coordinate system. However, if this coordinate system is not adjusted to the problem (like shown on the left site) the model equations will be more complicate than in a transformed coordinate system.

are defined [6,14,20]. Note that in principle one also may define other quantities of interest instead of the discussed levels of occupancy. Another choice of $\vec{y}$ would not affect the applicability of our method. However, it would modify the number of ODEs to be considered.

Step 1: The reaction networks considered are described by a system of ordinary differential equations (ODEs). All feasible reactions $A + B \leftrightarrows C$ are translated into reaction rates $r = k_{on}c_Ac_B - k_{off}c_C$ using the law of mass action. Here, k_{on} and k_{off} denote the association and dissociation constants, while c_i refers to the concentrations of the components. The ODEs for all feasible micro-states have the form

$$\frac{dc_i}{dt} = \sum r_{production} - \sum r_{comsumption}. \qquad (4)$$

In vector notation, Equation 4 can be written as $\dot{\vec{x}} = \vec{f}(\vec{x}, \vec{u})$, where $\vec{x}$ denotes the vector of all considered micro-states and $\vec{u}$ denotes all possible input signals (e.g. extra-cellular ligand concentration in the case of a receptor at the cell membrane). The macro-states/output variables we are interested in will be denoted as $\vec{y} = \vec{h}(\vec{x})$, which can be calculated using $\vec{x}$. Incorporating qualitative information about domain interactions helps to reduce the number of relevant parameters. For example, if it is known that two binding sites A and B are independent, the on- and off-rate constants do not change upon effector binding to the other domains. Note, that this is not a simplification, but merely a systematic incorporation of real physico-chemical constraints.

Step 2: In order to introduce new coordinates $\vec{z}$ including the macro-states, we perform a linear transformation $\vec{z} = T\vec{x}$, where T is a quadratic, non-singular matrix. The transformed model equations are

$$\dot{\vec{z}} = T\dot{\vec{x}} = T\vec{f}(\vec{x}, \vec{u})\Big|_{\vec{x} = T^{-1}\vec{z}}. \qquad (5)$$

In the following, we will introduce this transformation for a number of simple cases to exemplify our method and discuss the general pattern of this transformation. A strict mathematical and general formulation is also provided.

Step 3: Figure 1 shows that a state space transformation can simplify a model. In many relevant cases the transformed model equations of the considered scaffold proteins can be divided into two modules as shown in the following Equations

$$\vec{z} = \begin{bmatrix} \dot{\vec{z}}_1 \\ \dot{\vec{z}}_2 \end{bmatrix} = \begin{bmatrix} \vec{g}_1(\vec{z}_1, \vec{u}) \\ \vec{g}_2(\vec{z}_1, \vec{z}_2, \vec{u}) \end{bmatrix}, \qquad (6)$$

$$\vec{y} = \vec{h}(\vec{z}_1). \qquad (7)$$

Since T is a quadratic and non-singular (i.e. invertible) matrix these transformed equations still contain exactly the same information as the complete mechanistic model. Equation 6 shows that the macro-states (which are a subset of $\vec{z}_1$ and correspond to the output variables) are not influenced by the states $\vec{z}_2$. This implies that a reduced model only has to account for the ODEs $\dot{\vec{z}}_1 = \vec{g}_1(\vec{z}_1, \vec{u})$, and that this reduced model provides exactly the same output dynamics as the complete mechanistic model. However, note that by eliminating the equations for $\vec{z}_2$ one looses the possibility of using the inverse mapping T^{-1} in order to retrieve the original variables $\vec{x}$.

Generic example with three binding domains
We will start with a scaffold protein with three binding domains. In our example, each domain can bind one distinct effector molecule. Hence, the scaffold protein can exist in eight different micro-states ($R[0, 0, 0]$, $R[1, 0, 0]$, $R[0, 1, 0]$, $R[0, 0, 1]$, $R[1, 1, 0]$, $R[1, 0, 1]$, $R[0, 1, 1]$, $R[1,1,1]$). Even in this simple example 12 elementary reactions have to be considered (see Table 1).

Functionality of a scaffold protein
Using the law of mass action and the reactions defined in Table 1, the model equations can be derived using Equa-

Table 1: Reactions for scaffold with 3 binding sites. A complete mechanistic model of a scaffold protein with 3 binding domains (1,2,3), where each domain can bind one effector protein (E_1, E_2, E_3), has to consider the following 12 reactions. The kinetic parameters for each reaction can be denoted with k_{+i} for the association and k_{-i} for the dissociation reaction

Binding of E_1		Binding of E_2		Binding of E_3	
$R[0, 0, 0]+E_1 \leftrightharpoons R[1, 0, 0]$	(49a)	$R[0, 0, 0]+E_2 \leftrightharpoons R[0, 1, 0]$	(50a)	$R[0, 0, 0]+E_3 \leftrightharpoons R[0, 0, 1]$	(51a)
$R[0, 1, 0]+E_1 \leftrightharpoons R[1, 1, 0]$	(49b)	$R[0, 0, 1]+E_2 \leftrightharpoons R[0, 1, 1]$	(50b)	$R[0, 1, 0]+E_3 \leftrightharpoons R[0, 1, 1]$	(51b)
$R[0, 0, 1]+E_1 \leftrightharpoons R[1, 0, 1]$	(49c)	$R[1, 0, 0]+E_2 \leftrightharpoons R[1, 1, 0]$	(50c)	$R[1, 0, 0]+E_3 \leftrightharpoons R[1, 0, 1]$	(51c)
$R[0, 1, 1]+E_1 \leftrightharpoons R[1, 1, 1]$	(49d)	$R[1, 0, 1]+E_2 \leftrightharpoons R[1, 1, 1]$	(50d)	$R[1, 1, 0]+E_3 \leftrightharpoons R[1, 1, 1]$	(51d)

tion 4. However, besides the model structure the kinetic parameters play a crucial role in determining the system dynamics. A scaffold protein can perform a number of different functionalities, which can be characterized in terms of domain-interactions. One can think of a variety of cases such as non-interacting binding sites or the existence of a controlling domain influencing the others. A scaffold protein with 3 binding domains may provide 13 different functionalities or interaction motifs [21]. Qualitative knowledge about domain interactions have to be included in the modelling step. By defining the occuring domain interactions the number of relevant parameters follows immediately as exemplified in Figure 2. The assumption that, e.g., all three domains are completely independent implies that the kinetic on- and off-rate constants of one domain do not change upon effector binding at another domain. In the following, we want to show that for a number of interesting functionalities of scaffold proteins, we are able to find considerably reduced models using our method. Additionally, our method is also applicable to all other kind of functionalities which shall be

exemplified considering a scaffold protein with four docking sites and a more complicated structure of occuring domain interactions. More detailed information about how the kinetic parameters have to be adjusted to realize a certain functionality can be found in the supplementary material (simulation files provided).

Hierarchical transformation
The second step after having formulated a complete mechanistic model is to perform a state space transformation. This transformation introduces new states, including the levels of occupancy of each domain. Choosing a globally invertible and smooth transformation assures that the system's dynamics is preserved and, as long as none of the new equations is eliminated, the original micro-states can be retrieved from the new ones at any time [22].

We choose a hierarchical transformation matrix, consisting of different tiers (Table 2). Each tier represents a certain level of detail. First, we define the 0th tier of our transformation matrix, which includes only one state rep-
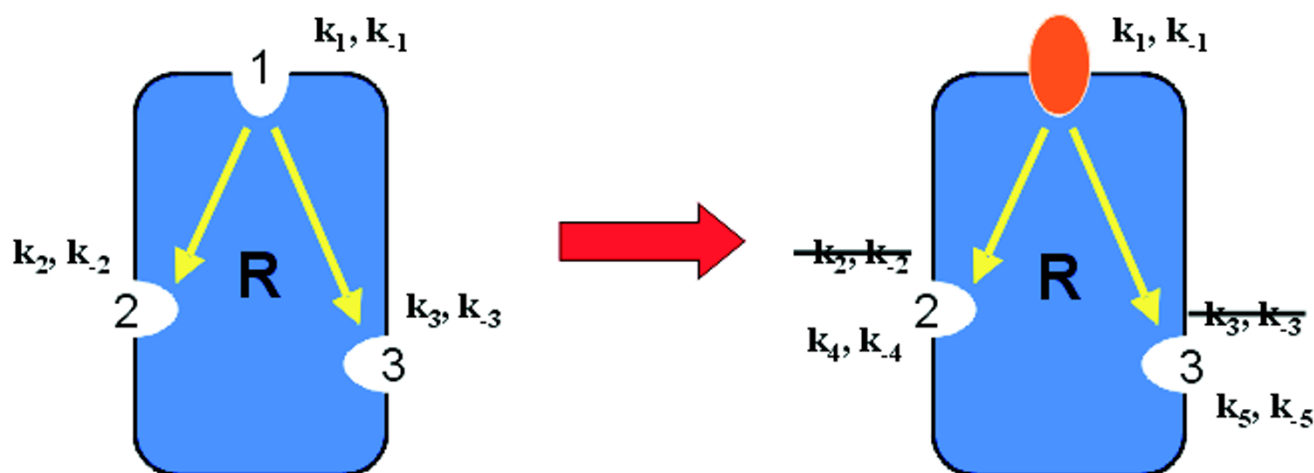

Figure 2
Domain interactions. We assume that binding domain one controls the other two domains like indicated by the arrows. From this assumption the kinetic parameters for the model follow immediately. As soon as binding domain one is occupied, the affinities of the docking sites two and three will change. Since binding domain one is independent of the other binding sites, the on- and off-rate constants of this domain are also independent.

Table 2: State Space Transformation for scaffold with three binding domains. Transformation for a scaffold protein with 3 binding domains. The transformation is hierarchically structured and introduces macroscopic quantities like the overall concentration of *R* and the levels of occupancy of each domain (z_0 to z_3), mesoscopic quantities describing pairs of concurrently occupied domains (z_4 to z_6) and microscopic quantities corresponding to individual multiprotein species (z_7).

$z_0 = R[0, 0, 0] + R[1, 0, 0] + R[0, 1, 0] + R[0, 0, 1] + R[1, 1, 0] + R[1, 0, 1] + R[0, 1, 1] + R[1, 1, 1]$ (52a)

$z_1 = R[1, 0, 0] + R[1, 1, 0] + R[1, 0, 1] + R[1, 1, 1]$ (52b)

$z_2 = R[0, 1, 0] + R[1, 1, 0] + R[0, 1, 1] + R[1, 1, 1]$ (52c)

$z_3 = R[0, 0, 1] + R[1, 0, 1] + R[0, 1, 1] + R[1, 1, 1]$ (52d)

$z_4 = R[1, 1, 0] + R[1, 1, 1]$ (52e)

$z_5 = R[1, 0, 1] + R[1, 1, 1]$ (52f)

$z_6 = R[0, 1, 1] + R[1, 1, 1]$ (52g)

$z_7 = R[1, 1, 1]$ (52h)

resenting the total concentrations of the scaffold protein (Equation 52a). Since there is no production nor degradation of *R* in our example, z_0 will be constant. However, in the general case (including production and degradation) this is also an important dynamic and macroscopic quantity of interest. The 1st tier of our matrix corresponds to the discussed levels of occupancy of each binding domain (Equations 52b to 52d in our example). All states belonging to both of these tiers are called macro-states. The 2nd tier describes the levels of occupancy of all pairs of domains, corresponding to the concentration of scaffold proteins *R* with concurrently occupied binding domains {1,2}, {1,3} and {2,3} (Equations 52e to 52g in our example), which we call mesoscopic states. The 3rd and, in our example, last tier represents the levels of occupancy of all triples of domains corresponding to the micro-state *R*[1, 1, 1] (Equation 52h). The transformation matrix in the general case is structured completely similar. One starts with the overall concentration and the levels of occupancy, followed by all possible pairs, triples and higher tuples of concurrently occupied domains until one reaches the tier describing the micro-states. We define that all new states describing macroscopic properties as overall concentrations or levels of occupancy of distinct docking sites are called macro-states. Others describing pairs, triples or higher tuples of concurrently occupied binding sites are mesoscopic states and those specifying single multiprotein species correspond to the original micro-states. A generalized transformation applicable to scaffold proteins with *n* binding domains as well as a proof that any transformation matrix deduced with the described pattern is invertible is provided in the following. As mentioned above, the functionality of a scaffold protein or a receptor is uniquely determined by the choice of parameters. Hence, the transformation only depends on the number of binding domains and effectors, not on the functionality of the protein. This means that each scaffold protein with n docking sites has to be transformed in exactly the same way no matter which functionality it provides. However, the number of equations of the reduced model will of course depend on the functionality of the protein as will be shown in the following.

Independent binding domains

First, a reduced model describing the example with three independent binding domains shall be introduced. The mechanistic model (Equation 4) is transformed as specified in Table 2. The transformed model equations have a very special structure (see Table 3). The total concentration of the scaffold protein z_0 is constant. Additionally, the output variables (levels of occupancy) which are described by the states z_1 to z_3 in the model appear to be completely decoupled from all other ODEs. This finding implies that a model describing the levels of occupancy does not need to consider the whole set of ODEs. They are *accurately* described by Equations 53b to 53d. From the structure of the equations it can be also seen that all kinetic parameters of the model can only be identified if one has measurements for all three states.

One site controls the others

Now we assume that binding domain 1 controls the domains 2 and 3 (see Figure 3b). Again, this functionality corresponds to a special parameter combination, and the model is transformed using the same transformation as in the example above (see Table 2). In this case the states z_1 to z_5 are completely independent of states z_6 and z_7 (Table 4). As we are only interested in the levels of occupancy (states z_1 to z_3), the Equations for z_6 and z_7 can be excluded. Additionally, the Equations for z_1 to z_5 can be divided into three modules that are completely free of retroactive effects thus facilitating the analysis of the model as well as the application of parameter identification tools [23-25]. In comparison to the previous example two more equations are required in order to describe the dynamics of the macro-states. An explanation of this fact is provided by the following example. Consider the binding domains 1 and 2 of two equal scaffold molecules *R*. Only considering the macro-states ("domain 1 occupied" and "domain 2 occupied"), it is not possible to distinguish between the following two scenarios: [1] one molecule has both sites occupied and the other has none (*R*[1, 1,a_3] and *R*[0, 0,

Table 3: Transformed equations for independent domains. Transformed model equations for a scaffold protein with independent binding domains. The levels of occupancy (z_1 to z_3) do not depend on the states z_4 to z_7.

$$\dot{z}_0 = 0 \quad (53a)$$
$$\dot{z}_1 = k_1 (z_0 - z_1) E_1 - k_{-1} z_1 \quad (53b)$$
$$\dot{z}_2 = k_2 (z_0 - z_2) E_2 - k_{-2} z_2 \quad (53c)$$
$$\dot{z}_3 = k_3 (z_0 - z_3) E_3 - k_{-3} z_3 \quad (53d)$$
$$\dot{z}_4 = k_1 (z_2 - z_4) E_1 - k_{-1} z_4 + k_2 (z_1 - z_4) E_2 - k_{-2} z_4 \quad (53e)$$
$$\dot{z}_5 = k_1 (z_3 - z_5) E_1 - k_{-1} z_5 + k_3 (z_1 - z_5) E_3 - k_{-3} z_5 \quad (53f)$$
$$\dot{z}_6 = k_2 (z_3 - z_6) E_2 - k_{-2} z_6 + k_3 (z_2 - z_6) E_3 - k_{-3} z_6 \quad (53g)$$
$$\dot{z}_7 = k_1 (z_6 - z_7) E_1 - k_{-1} z_7 + k_2 (z_5 - z_7) E_2 - k_{-2} z_7 + k_3 (z_4 - z_7) E_3 - k_{-3} z_7. \quad (53h)$$

a_3]) and [2] one has the first site occupied, but not the second, and the other has the second, but not the first (R[1, 0, a_3] and R[0,1, a_3]). However, this information is essential if binding domain 2 is controlled by domain 1, since in the first case the affinity of R[0, 0, a_3] to the effector E_2^1 has a different value from the second case R[1, 0, a_3].

Analysis and dynamic simulations

In order to illustrate the advantages of our method, the previous example will be discussed (see Figure 3b) in more detail, including dynamic simulations. Using the parameters listed in Table 5, we create three different models for this case: (i) a complete mechanistic description, (ii) an 'heuristically' reduced model, and (iii) an exactly reduced model according to the method described herein. Simulation results of the models are compared, and the advantages and disadvantages of the models are discussed.

Model 1: We create a complete mechanistic model accounting for all molecular species and all possible reactions (compare Table 1). Creating such a model one has to address the problem of combinatorial complexity. In this simple example the combinatorial variety of all feasible molecules includes only 11 different chemical species (8 complexes and 3 effectors). After incorporating conservation relations, the model only consists of 7 ODEs. However, most real problems would lead to models with tens of thousands or even millions of equations and the computational cost in these cases is extremely high. However, an advantage of such a detailed model is that it accurately represents the real reaction network. Now we want to compare the predictions of this complete mechanistic model with two different reduced models.

Model 2: We already mentioned that most heuristic models do not account for all molecular species. However, the equations of these models still describe the system at a microscopic level. The complete mechanistic network structure is substituted by a reduced structure focusing on
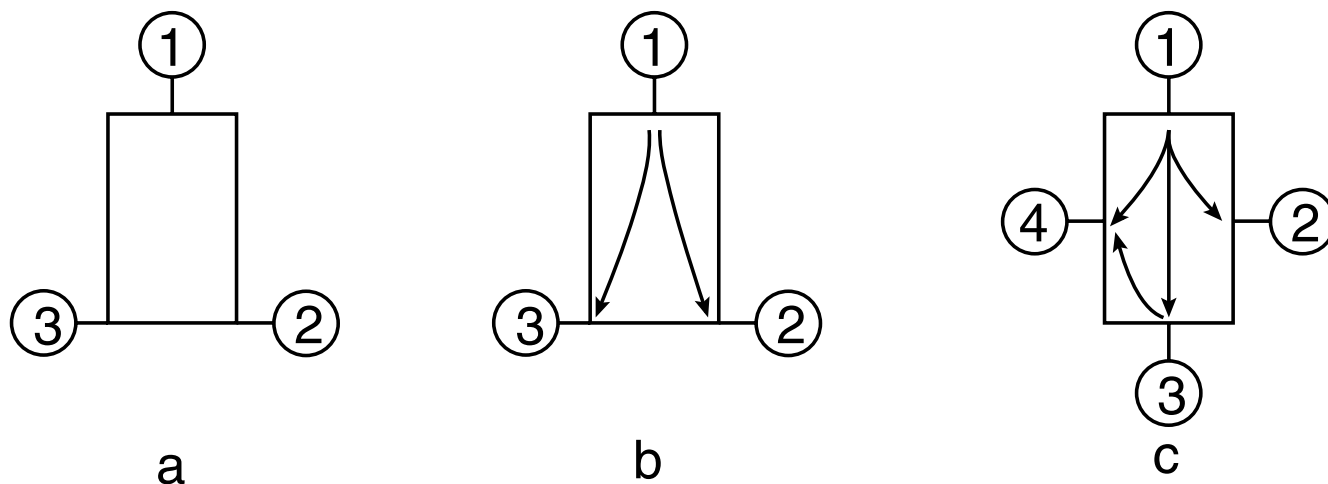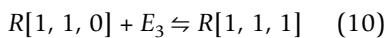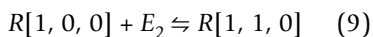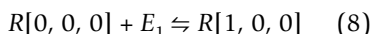

Figure 3
Interaction motifs. Generic examples of scaffold proteins with 3 or 4 docking sites. (a) A scaffold protein with 3 distinct docking sites which do not interact. (b) Another pattern of domain interactions for the same scaffold protein. Here binding domain 1 controls the other two domains. (c) A scaffold protein with 4 docking sites and a more complex pattern of domain interactions. Each pattern of domain interactions can be translated into special kinetic properties (like exemplified in Figure 2).

Table 4: Transformed ODEs for scaffold with on controlling domain. Transformed model equations for a scaffold protein with one controlling domain. The levels of occupancy (z_1 to z_3) are only influenced by the states z_4 and z_5 but not by z_6 and z_7.

$$\dot{z}_0 = 0 \quad (54a)$$

$$\dot{z}_1 = k_1 (z_0 - z_1) E_1 - k_{-1} z_1 \quad (54b)$$

$$\dot{z}_2 = k_2 (z_0 - z_1 - z_2 + z_4) E_2 - k_{-2} (z_2 - z_4) + k_3 (z_1 - z_4) E_2 - k_{-3} z_4 \quad (54c)$$

$$\dot{z}_3 = k_4 (z_0 - z_1 - z_3 + z_5) E_3 - k_{-4} (z_3 - z_5) + k_5 (z_1 - z_5) E_3 - k_{-5} z_5 \quad (54d)$$

$$\dot{z}_4 = k_1 E_1 (z_2 - z_4) - k_{-1} z_4 + k_3 E_2 (z_1 - z_4) - k_{-3} z_4 \quad (54e)$$

$$\dot{z}_5 = k_1 E_1 (z_3 - z_5) - k_{-1} z_5 + k_5 E_3 (z_1 - z_5) - k_{-5} z_5 \quad (54f)$$

$$\dot{z}_6 = k_2 (z_3 - z_5 - z_6 + z_7) E_2 - k_{-2} (z_6 - z_7) + k_4 (z_2 - z_4 - z_6 + z_7) E_3 - k_{-4} (z_6 - z_7) + k_3 (z_5 - z_7) E_2 + k_5 (z_4 - z_7) E_3 - (k_{-3} + k_{-5}) z_7 \quad (54g)$$

$$\dot{z}_7 = k_1 (z_6 - z_7) E_1 - k_{-1} z_7 + k_3 (z_5 - z_7) E_3 - k_{-3} z_7 + k_5 (z_4 - z_7) - k_{-5} z_7 \quad (54h)$$

a reduced number of species and a reduced number of reactions. As shown by Faeder et al. [12], which species and which reactions play an important role is highly dependent on the kinetic parameters, and it can hardly be found without the consideration of dynamic simulations of a complete mechanistic model. To illustrate the problems associated with this heuristic approach, we will show that even in our simple example a number of simplifications that may appear reasonable lead to a wrong model. The parameter values (see Table 5) show that if binding domain 1 is unoccupied the affinity of the other domains is extremely low. Hence, it may seem reasonable to neglect these reactions. Thus, the unoccupied receptor $R[0, 0, 0]$ first has to bind the effector E_1. Since the resulting affinity as well as the resulting on-rate of binding domain 2 is approximately several hundred-fold higher than the affinity or the on-rate of domain 3, we assume that the effector E_2 in the majority of cases will bind before E_3. Therefore, the reduced model only includes the following three reactions

$$R[0, 0, 0] + E_1 \leftrightharpoons R[1, 0, 0] \quad (8)$$

$$R[1, 0, 0] + E_2 \leftrightharpoons R[1, 1, 0] \quad (9)$$

$$R[1, 1, 0] + E_3 \leftrightharpoons R[1, 1, 1] \quad (10)$$

which are parametrized with the known kinetic constants. This case represents a commonly performed simplification. In [6], e.g. the two effectors GAP and Shc bind consecutively to the receptor, although the EGF receptor provides two distinct binding domains for these effectors similar to the scaffold in our example. In order to compare the predictions of this reduced model with the complete one we again consider the levels of occupancy of each domain. A comparison of the simulation results shows that the predictions of the reduced model, for these parameter values, are incorrect. Certainly, it would be possible to improve the predictions by fitting the parameter values using the curves provided by the correct model, but then the parameters would be phenomenological parameters and would not correspond to the kinetic parameters of the reactions. The general problem of models with such a linear chain of reactions is that the resulting levels of occupancy are always higher than one would expect considering the equilibrium constants of the reactions. The reason is quite simple. The equilibrium constant for reaction 8 determines the equilibrium between $R[0, 0, 0]$, E_1 and $R[1, 0, 0]$. However, in equilibrium only a small fraction of scaffold proteins which have bound E_1 are in the state $R[1, 0, 0]$. Most of them have also bound E_2 and E_3, and the structure of Model 2 does not allow E_1 to dissociate after E_2 and E_3 have bound to the scaffold protein.

Model 3: At last we want to consider the reduced model, derived using our method (compare equations for z_1 to z_5 in Table 4). The dynamic simulations of the two models analyzed above show that both have disadvatages. The complete model is an accurate model describing all species and reactions, but the number of ODEs grows exponentially with the number of binding domains. While the second approach has the advantage that the number of equations is manageable, the error is not known. Our method combines the merits of both approaches: the number of equations is lower than in the complete model, but the predictions of both models are exactly the same

Table 5: Kinetic parameters for dynamic simulation

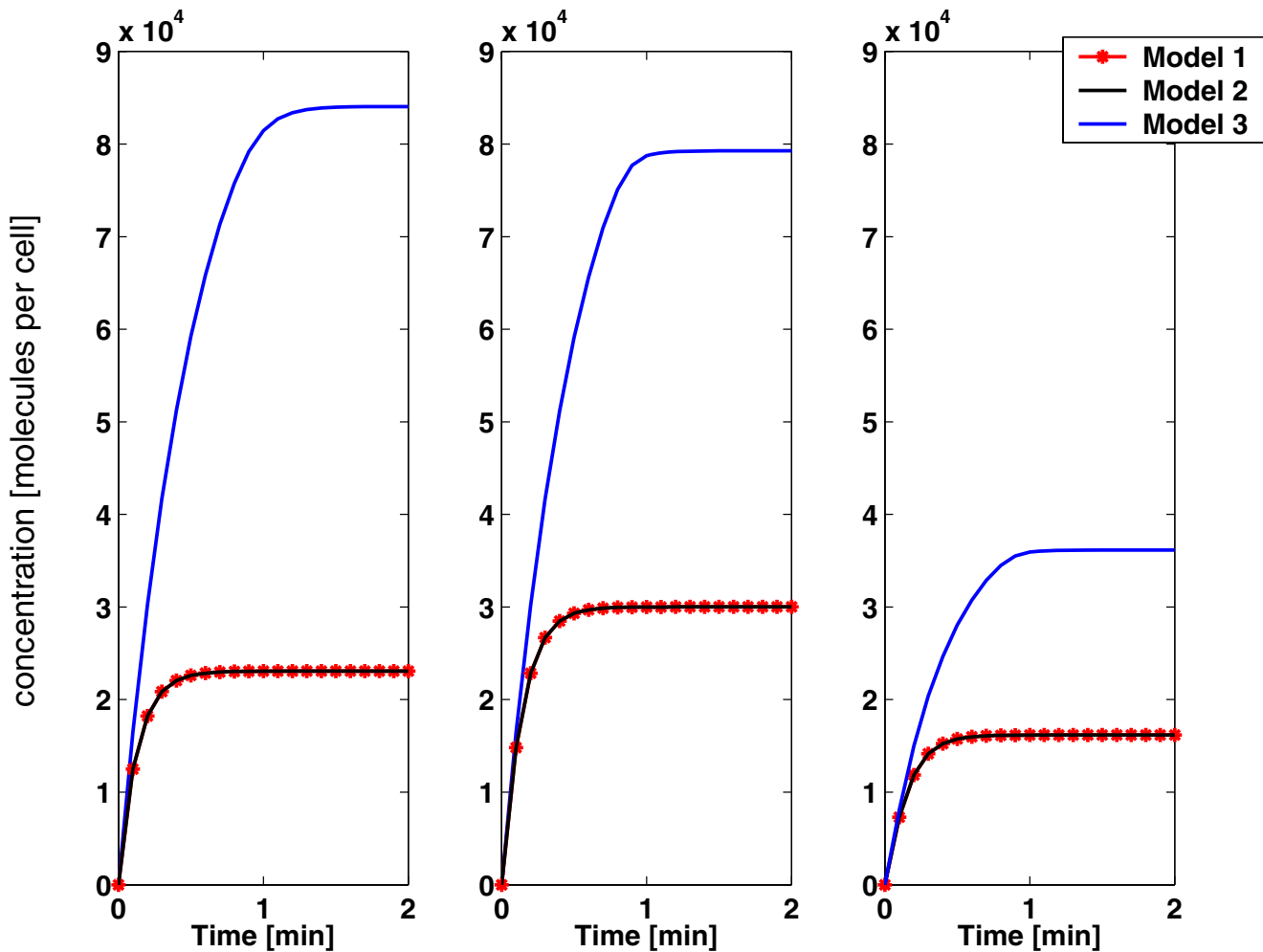| Affinity of domain	k_{on} [$M^{-1} min^{-1}$]	k_{off} [min^{-1}]	Equilibrium K_d [M^{-1}]
1 (always)	$k_1 = 3 \cdot 10^5$	$k_{-1} = 6$	$5 \cdot 10^4$
2 (domain 1 unoccupied)	$k_2 = 1$	$k_{-2} = 18$	$5.6 \cdot 10^{-2}$
2 (domain 1 occupied)	$k_3 = 5 \cdot 10^7$	$k_{-3} = 24$	$2.1 \cdot 10^6$
3 (domain 1 unoccupied)	$k_4 = 1$	$k_{-4} = 12$	$8.3 \cdot 10^{-2}$
3 (domain 1 occupied)	$k_5 = 1 \cdot 10^5$	$k_{-5} = 60$	$1.7 \cdot 10^3$

Figure 4
Dynamic simulations. Dynamic simulations of the example shown in Figure 3b using the parameter values presented in Table 5. Here we compare the levels of occupancy of the three protein domains. The left picture shows the level of occupancy of domain 1, the second picture shows the levels of occupancy of domain 2 and the right picture shows that of domain 3. The results show that the reduced model we obtained by applying our method (model 3) accurately describes the real time course represented by a complete mechanistic model (model 1). The other model which follows from a number of reasonable simplifications which can also be found in literature provides completely different results.

(compare Figure 4). Additionally, our method gives rise to a very useful modular structure, which simplifies greatly the model analysis. In our example the model consists of three modules. The first module only includes the equation for z_1, whereas the second consists of the two states z_2 and z_4 and the third of z_3 and z_5. The states of the second module do not influence the states of the third module and vice versa. The dynamics of z_1 is neither influenced by the states of the second nor the third module, but z_1 influences all the other states. Therefore, the model is not only modular but also hierarchically structured. This motivates a modular approach in order to analyze the model dynamics. First, one can analyze the dynamics of the first module. Afterwards the other two modules can also be

analyzed separately. A similar approach is possible for the parameter identification since our transformation also leads to a modularization of the parameters. The only parameters of the first module are k_1 and k_{-1}. The parameters k_2, k_{-2}, k_3 and k_{-3} can be found only in the second module, and the remaining parameters are present only in the third module. This allows one to identify the parameters step by step, which is a much simpler task than to identify all parameters at once. In addition, it also shows which parameters can be identified by certain measurements: for example, a measured time course of z_2 (level of occupancy of domain 2) does not allow the identification of the parameters k_4, k_{-4}, k_5 and k_{-5}. The same also holds true for the complete model (model 1), but it is not intuitive to

untangle this feature in the structure of the ODEs of the complete model. Because of all these advantages we think that the method proposed here offers a useful framework to handle multiprotein complex formation in signaling and regulation networks.

Example with more than one controlling domain

In the previous chapter we analyzed scaffold proteins with completely independent binding domains or with only one controlling domain. However, our method can also be applied to a more general case. This will be exemplified by considering a scaffold protein with 4 binding domains, where each domain can be free or occupied by an effector. We assume that binding domain 1 controls domains 2, 3 and 4. Additionally, binding domain 3 also interacts with binding domain 4 (see Figure 3c). The number of feasible micro-states in this example is $2^4 = 16$. Applying our method to this example (details can be found in the Appendix), one finds that 9 equations are sufficient to describe the complete dynamics of the macro-states. The states that are required are the levels of occupancy of all 4 binding domains, the pairs of concurrently occupied binding domains {1,2}, {1,3}, {1,4} and {3,4}, as well as the triple of concurrently occupied binding domains {1,3,4}. This result shows that the more binding domains interact, more ODEs are required in order to describe the dynamics of the macro-states. Indeed, if all binding domains interact with each other, one really has to consider the whole combinatorial variety.

Scaffold proteins with n binding sites
Generalized transformation matrix

Here we want to formalize the transformation matrix for any scaffold protein with n binding sites i. Additionally, we assume that a number of m_i effectors compete for the binding domain a_i. Hence, each binding domain i can exist in $m_i + 1$ different states. The transformation matrix is independent of the intramolecular domain interactions. The 0th tier of our transformation matrix consists of only one state describing the overall concentration of the scaffold protein. The new state z_0 corresponds to the sum of all feasible micro-states

$$z_0 = \sum_{a_1=0}^{m_1} \dots \sum_{a_n=0}^{m_n} R[a_1, \dots, a_n]. \tag{11}$$

The levels of occupancy of each binding domain are described in the 1st tier. The status of the binding domain i is fixed at the status k (with $k \in \{1, \dots, m_i\}$) and one sums up all micro-states whose binding domain i is occupied by the effector E_i^k. Mathematically, this tier can be described by

$$z_1 = \sum_{a_2=0}^{m_2} \dots \sum_{a_n=0}^{m_n} R[1, a_2, \dots, a_n] \tag{12}$$

$$\vdots$$

$$z_{m_1} = \sum_{a_2=0}^{m_2} \dots \sum_{a_n=0}^{m_n} R[m_1, a_2, \dots, a_n] \tag{13}$$

$$z_{m_1+1} = \sum_{a_1=0}^{m_1} \sum_{a_3=0}^{m_3} \dots \sum_{a_n=0}^{m_n} R[a_1, 1, \dots, a_n] \tag{14}$$

$$\vdots$$

$$z_{\xi_1} = \sum_{a_1=0}^{m_1} \dots \sum_{a_{n-1}=0}^{m_{n-1}} R[a_1, \dots, a_{n-1}, m_n], \tag{15}$$

with $\xi_1 = \sum_{i=1}^{n} m_i$. The 2nd tier, which represents all pairs of concurrently occupied binding sites, corresponds to

$$z_{\xi_1+1} = \sum_{a_3=0}^{m_3} \dots \sum_{a_n=0}^{m_n} R[1, 1, a_3, \dots, a_n] \tag{16}$$

$$\vdots$$

$$z_{\xi_2} = \sum_{a_1=0}^{m_1} \dots \sum_{a_{j_1-1}=0}^{m_{j_1-1}} \sum_{a_{j_1+1}=0}^{m_{j_1+1}} \dots \sum_{a_{j_2-1}=0}^{m_{j_2-1}} \sum_{a_{j_2+1}=0}^{m_{j_2+1}} \dots \sum_{a_n=0}^{m_n} R[a_1, \dots, a_{j_1-1}, k, a_{j_1+1}, \dots a_{j_2-1}, l, a_{j_2+1}, \dots a_n] \tag{17}$$

$$\vdots$$

with $\xi_2 = \sum_{i_1=1}^{n} m_{i_1} + \sum_{i_1}^{j_1-1} \sum_{i_2=0}^{j_2} m_{i_1} m_{i_2} + kl$.

The following tiers represent all possible tuples, and the last tier of the transformation matrix (the $n + 1$-th tier) contains the micro-states with all binding sites being occupied and can be written as

$$z_{\xi_3} = R[1, \dots, 1] \tag{18}$$

$$\vdots$$

$$z_{\xi_4} = R[m_1, m_2, \dots, m_n] \tag{19}$$

with $\qquad \xi_3 = \prod_{i=1}^{n}(m_i+1) - \prod_{i=1}^{n} m_i \qquad$ and

$\xi_4 = \prod_{i=1}^{n}(m_i+1) - 1$. Using this pattern to derive the transformation matrix one can handle each possible scaffold protein.

Independent binding domains

Now we want to consider a scaffold protein with n binding sites and we assume that all binding domains are independent like we already did for a scaffold protein with three domains (see above). The parameters of the reaction network have to be adjusted to the case of independent binding domains as discussed above. The transformation

matrix in this case also follows the hierarchical pattern as discussed in the section above. It can be shown that the whole dynamical behavior can be described by $\sum m_i$ equations (i.e., states) instead of $\prod (m_i + 1)$ equations. A proof that in this case the macro-states are always sufficient to describe the system can also be found in [16].

One site controls the others

We also want to generalize the case that binding domain 1 controls all other $(n - 1)$ binding domains, in a manner such that the affinity of each binding domain i to its respective effectors E_i^k only depends on the status of the binding domain 1. As already mentioned, the transformation matrix remains the same as for a protein with n independent binding domains. However, since the kinetic parameters are different, in this case more equations are required to describe the dynamics of the macro-states. In addition to the macro-states one requires all states describing pairs of concurrently occupied binding domain 1 and i with $i \neq 1$ (i.e. all states describing allosterically interacting domains). The total number of equations in this case is $2 \sum m_i - m_1$ instead of $\prod (m_i + 1)$.

Linker for activation of T cells (LAT)

LAT (Linker for Activation of T cells) is a scaffold molecule that plays a pivotal role in T cell signaling [26]. LAT has 9 conserved cytoplasmatic tyrosines, of which the four membrane-distal tyrosines (at residues 132/171/191/226 in human LAT) are essential and are phosphorylated upon ligand binding to the T cell receptor [26]. Different signaling molecules, such as PLCγ1, Grb2 and Gads can bind to the different residues. Grb2 recruits Sos, which in turn activates Ras, and subsequently the Raf/MEK/ERK MAP Kinase cascade. On the other hand, binding of PLCγ1 and Gads (bound to the adaptor SLP76 that additionally recruits Itk), allows the activation of PLCγ1, leading to the cleavage of phosphatidyl-inositol-4,5 bisphosphate (PIP_2) and the generation of dyacilglycerol (DAG) and inositol trisphosphate IP_3. DAG activates RasGRP, which in turn activates Ras, as well as PKC, while IP_3 regulates Calcium signaling [27].

PLCγ1 binds at the Y132 tyrosine, Grb2 at Y171, Y191 and Y226, and Gads at Y171 and Y191 (see Figure 4) [26]. The number of different protein complexes occurring in this simple example is already $2 \cdot 3 \cdot 3 \cdot 2 = 36$, and the number of reactions that have to be considered is 86. In the following we will show how one can precisely describe the levels of occupancy without considering all 36 complexes. First, it is assumed that the binding domains do not interact with each other. In order to reduce the model, the system has to be transformed using the transformation pattern discussed above. The transformed model equations show

that only six differential equations are sufficient to completely describe the dynamics of the quantities of interest, namely the levels of occupancy of Y132, Y171, Y191 and Y226 with PLCγ1, Grb2 and Gads. Additionally, the equation describing PLCγ1 binding is decoupled from the other equations, while these are interlinked because Grb2 can bind to Y171, Y191 and Y226 and compete with Gads for the binding domains Y171 and Y191.

Recent experimental data from LAT mutation studies indicate that the binding domains can influence one another, which contradicts the assumption of the complete independence [28]. Binding of Grb2 to Y226 appears to help the binding of Gads to LAT [28]. This effect can be readily incorporated into the model by changing the kinetic parameters for Gads binding if the binding site Y226 is occupied by Grb2. Transforming the model equations shows that now ten ODEs are required to exactly describe the system dynamics (see supplementary information). The four additional states that are required here describe the number of LAT molecules, with concurrently occupied residues Y226 and either Y171 or Y191, while Y171 and Y191 can be occupied either by Grb2 or Gads. Although the model includes 4 additional states compared to the previously discussed one, the model can still be notably reduced from over thirty equations to ten. Hence, our method is capable of reducing signal transduction models including scaffold proteins notably. The modular structure of the derived model equations also strongly facilitates the model analysis as well as parameter estimation [25].

Discussion

We have presented an approach which allows one to create reduced models of multiprotein complex formation processes often occuring in signal transduction cascades, but also in regulation mechanisms (e.g. cell cycle regulation [29]). Since each model reduction results in the loss of information, it is essential to define quantities of interest (output variables $\vec{y}$) whose dynamics should be preserved. As discussed in the background section, a reasonable choice are macroscopic values, such as levels of occupancy or phosphorylation degrees. Besides the determination of the output variables, the incorporation of qualitative knowledge about domain interactions is a key element in our approach. Both the choice of output variables as well as the considered domain interactions determine the number of ODEs in the reduced model. Hence, it is clear that for some patterns of domain interactions a reduction of the model may not be possible (e.g. if each domain interacts with all other domains). However, our preliminary results indicate that in many of these cases good approximate solutions can be found (data not

shown). Since each possible pattern of domain interactions can be realized in the modeling step, and the state space transformation which has to be performed is completely independent of this interaction pattern, our method is generally applicable to all kind of molecules offering several binding sites. The only limitation is the possibility that no exact model reduction is possible which, however, is a general mathematical limitation and not an insufficiency of the method. Since our approach is very systematic, independent of exact numerical values for kinetic parameters and can be easily implemented as a computer algorithm, our hope is that this method will help to standardize, improve and simplify modeling and the analysis of important signaling networks.

Conclusion
We have shown that a complete mechanistic model of scaffold proteins or receptors as discussed in [1,14,15], which describes the system at a microscopical level, can be linked in a analytical manner with the macroscopic and reduced description introduced in [16]. Our method is based on a linear state space transformation with a hierarchical structure (substituting the original mechanistic description by macroscopic, mesoscopic and a special set of microscopic states). It is a systematic and powerful tool to derive reduced model structures to describe the dynamics of multiprotein complex formation accurately.

Appendix
Proof
In order to prove that the transformation matrix is invertible, we will first show that the transformation matrix is quadratic, which is a necessary condition. Second, we will show that this matrix can be written as a upper triangular matrix, which is a sufficient condition for invertibility.

For each protein domain/site i, we introduce a set $\{a_i, 1, ..., m_i\}$, where the numbers $1, ..., m_i$ denote possible states of site i and the character a_i replaces 0, which was used to designate the basal state of site i. Having n such sets for all n domains ($i = 1, ..., n$), we select one element from each set and denote the resulting combination by $\tilde{z}_k$. Thus, any $\tilde{z}_k$ is a set of n elements where each element corresponds to one of the n domains and is either a number or a character (a_i). There are $\prod_{i=1}^{n}(m_i + 1)$ different combinations $\tilde{z}_k$, exactly equaling the number of micro-states (columns in the transformation matrix T). The transformation of combination $\tilde{z}_k$ into the variable z_k is straightforward: summing up all micro-states that (1) correspond to each character entry a_i (from $i = 0$ to m_i) and

(2) have the states of the other domain given by the numbers in $\tilde{z}_k$, we obtain the variable $\tilde{z}_k$. We conclude that the number of rows in the matrix T (the variables Z_k) equals the number of columns (the micro-states).

The next step is to proof that all columns of the transformation matrix are linearly independent by complete induction. If we look consecutively at all the new defined variables, starting with the last one in our listing above, one can show that each state is a sum of already defined states plus *one* additional new state. Hence, the transformation matrix is an upper triangular matrix, and an upper triangular matrix has linearly independent columns and is hence invertible. Considering the last tier of our transformation matrix, which corresponds to a number of different micro-states, it is obvious that in each column a new micro-state is introduced. The i-th tier is a sum of micro-states, which have i - 1 binding domains with a well-defined status. In contrast, the status of the remaining binding domains can vary. One possible state is that all n - i + 1 undetermined binding domains are unoccupied. This state is introduced here the first time, since all previous tiers consist of micro-states with a higher number of binding domains with well-defined status (excluding the unoccupied status). Hence, the number of unoccupied binding domains in the previous tiers is lower and cannot contain the described micro-state.

Example with more than one controlling domain
In this example we assume that the considered scaffold molecule has four binding domains (1, 2, 3 and 4), which can bind four distinct effectors E_1, E_2, E_3 and E_4. The domain-domain interactions are depicted in Figure 1c (Binding domain 1 controls all other domains and binding domain 3 additionally influences binding domain 4). A mechanistic model accounting for all possible reactions of this scaffold molecule R with the effectors consists of 16 ordinary differential equations and incorporates 32 reactions. We provide a Mathematica-File, including all reactions, the mechanistic model equations as well as the transformation and the resulting reduced model. As already denoted, the reduced model consist of 9 ordinary differential equation, namely

$$\dot{z}_0 = 0 \quad (20)$$

$$\dot{z}_1 = k_1(z_0 - z_1)E_1 - k_{-1}z_1 \quad (21)$$

$$\dot{z}_2 = k_2(z_0 - z_1 - z_2 + z_5)E_2 - k_{-2}(z_2 - z_5) + k_3(z_1 - z_5)E_2 - k_{-3}z_5 \quad (22)$$

$$\dot{z}_3 = k_4(z_0 - z_1 - z_3 + z_6)E_3 - k_{-4}(z_3 - z_6) + k_5(z_1 - z_6)E_3 - k_{-5}z_6 \quad (23)$$

$\dot{z}_4 = k_6(z_0 - z_1 + z_{10} - z_{13} - z_3 - z_4 + z_6 + z_7)E_4 - k_{-6}(z_4 - z_7 - z_{10} + z_{13})$ (24)

$k_7(z_1 - z_6 - z_7 + z_{13})E_4 - k_{-7}(z_7 - z_{13}) + k_8(z_3 - z_6 - z_{10} + z_{13})E_4$

$- k_{-8}(z_{10} - z_{13}) + k_9(z_6 - z_{13})E_4 - k_{-9}z_{13}$ (25)

$\dot{z}_5 = k_1(z_2 - z_5)E_1 - k_{-1}z_5 + k_3(z_1 - z_5) - k_{-3}z_5$ (26)

$\dot{z}_6 = k_1(z_3 - z_6)E_1 - k_{-1}z_6 + k_5(z_1 - z_6) - k_{-5}z_6$ (27)

$\dot{z}_7 = k_1(z_4 - z_7)E_1 - k_{-1}z_7 + k_9(z_6 - z_{13})E_4 - k_{-9}z_{13}$ (28)

$+ k_7(z_1 - z_6 - z_7 + z_{13})E_4 - k_{-7}(z_7 - z_{13})$

$\dot{z}_{10} = k_4(z_4 - z_7 - z_{10} + z_{13})E_3 - k_{-4}(z_{10} - z_{13}) + k_5(z_7 - z_{13})E_3 - k_{-5}z_{13}$ (29)

$+ k_8(z_3 - z_6 - z_{10} + z_{13})E_4 - k_{-8}(z_{10} - z_{13}) + k_9(z_6 - z_{13})E_4 - k_{-9}z_{13}$

$\dot{z}_{13} = k_1(z_{10} - z_{13})E_1 - k_{-1}z_{13} + k_5(z_7 - z_{13}) - k_{-5}z_{13}$ (30)

$+ k_9(z_6 - z_{13})E_4 - k_{-9}z_{13}.$

In these differential equations z_0 denotes the overall concentration of the scaffold protein. The states z_1 to z_4 represent the levels of occupancy of the four distinct binding domains. The remaining states describe the number of scaffold proteins with two or three concurrently occupied binding domains (z_5 corresponds to concurrently occupied binding domains {1,2}, z_6 equals {1,3}, z_7 {1,4}, z_{10} {3,4} and z_{13} {1,3,4}).

Linker for activation of T cells (LAT)
Applying our method to the adaptor protein LAT (Linker for activation of T cells), we derived two different reduced models (dependent of the assumptions that are made). First, we assume that the binding residues Y132, Y171, Y191 and Y226 are completely independent (see Figure 2). Hence, the reduced model only includes six differential equations. The derivation of these equations as well as the reactions and the mechanistic model equations are provided as a Mathematica-File. The reduced model equations are

$\dot{z}_0 = 0$ (31)

$\dot{z}_1 = k_1(z_0 - z_1)PLC - k_{-1}z_1$ (32)

$\dot{z}_2 = k_2(z_0 - z_2 - z_3)Grb2 - k_{-2}z_2$ (33)

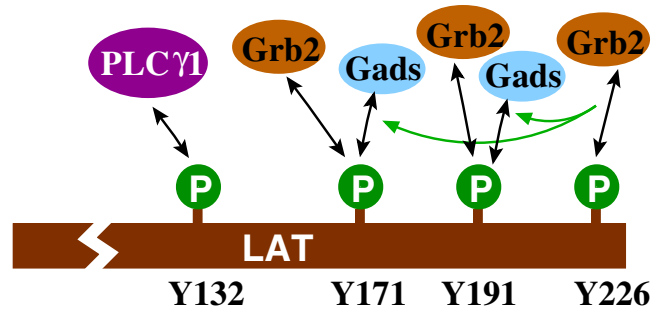$\dot{z}_3 = k_5(z_0 - z_2 - z_3)Gads - k_{-5}z_3$ (34)

Figure 5
Linker for activation of T-cells. The four distal tyrosine rests on LAT and the binding possibilities, according to [26, 28].

$\dot{z}_4 = k_3(z_0 - z_4 - z_5)Grb2 - k_{-3}z_4$ (35)

$\dot{z}_5 = k_6(z_0 - z_4 - z_5)Gads - k_{-6}z_5$ (36)

$\dot{z}_6 = k_4(z_0 - z_6)Grb2 - k_{-4}z_6.$ (37)

The state z_0 denotes the overall concentration of LAT molecules, which is assumed to be constant. The state z_1 equals the level of occupancy of Y132 with PLCγ1. z_2 and z_3 describe the levels of occupancy of Y171 either with Grb2 or Gads, z_4 and z_5 are the equivalent values for the binding domain Y191, and z_6 denotes the level of occupancy of Y226 with Grb2.

However, recent experimental results show that the binding domains of LAT influence each other. In a second example we assume that Y226 interacts with the two domains Y171 and Y191. The whole calculation can be found in the provided Mathematica-File. In this case the reduced model consists of ten ordinary differential equations, namely

$\dot{z}_1 = k_1(z_0 - z_1)PLC - k_{-1}z_1$ (38)

$\dot{z}_2 = k_2(z_0 - z_2 - z_3)Grb2 - k_{-2}z_2$ (39)

$\dot{z}_3 = k_5(z_0 - z_2 - z_3 - z_6 + z_{14} + z_{17})Gads - k_{-5}(z_3 - z_{17})$ (40)

$+ k_6(z_6 - z_{14} - z_{17})Gads - k_{-6}z_{17}$

$\dot{z}_4 = k_3(z_0 - z_4 - z_5)Grb2 - k_{-3}z_4$ (41)

$\dot{z}_5 = k_7(z_0 - z_4 - z_5 - z_6 + z_{18} + z_{19})Gads - k_{-7}(z_5 - z_{19})$ (42)

$+ k_8(z_6 - z_{18} - z_{19})Gads - k_{-6}z_{19}$

$\dot{z}_6 = k_4(z_0 - z_6)Grb2 - k_{-4}z_6$ (43)

$$\dot{z}_{14} = k_2(z_6 - z_{14} - z_{17})Grb2 - k_{-2}z_{14} + k_4(z_2 - z_{14})Grb2 - k_{-4}z_{14}$$
$$(44)$$

$$\dot{z}_{17} = k_4(z_3 - z_{17})Grb2 - k_{-4}z_{17} + k_6(z_6 - z_{14} - z_{17})Gads - k_{-6}z_{17}$$
$$(45)$$

$$\dot{z}_{18} = k_3(z_6 - z_{18} - z_{19})Grb2 - k_{-3}z_{18} + k_4(z_4 - z_{18})Grb2 - k_{-4}z_{18}$$
$$(46)$$

$$\dot{z}_{19} = k_4(z_5 - z_{19})Grb2 - k_{-4}z_{19} + k_8(z_6 - z_{18} - z_{19})Gads - k_{-8}z_{19}.$$
$$(47)$$

$$(48)$$

Again the states z_0 to z_6 correspond to the same quantities as described above. The state z_{14} equals to all LAT molecules with Grb2 being concurrently bound to the residues Y171 and Y226, z_{17} describes the molecules with Gads being bound to Y171 and Grb2 bound to Y226. The states z_{18} and z_{19} are equivalent quantities describing the same binding motifs for Y191 and Y226.

Authors' contributions

HC conceived the original mathematical idea. HC, JSR and BNK developed the systematic method.

Model building and numerical analysis were carried out by HC, JSR and TS. EDG initiated, supervised and coordinated the project. All authors wrote the manuscript and approved the final version.

Additional material

Additional File 1

Additional Files can be downloaded from http://www.sysbio.de/projects/ modred/BMC. All Files are created with Mathematica Version 5.0. The five files provide all information about the discussed examples. • Generic example with three binding domains (no domain interactions). • Generic example with three binding domains (one controlling domain). • Dynamic simulations of three distinct models. • Generic example with four binding domains. • LAT example (Linker for activation of T-Cells) Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-7-34-S1.tex]

Acknowledgements

References

1. Hlavacek WS, Faeder JR, Blinov ML, Perelson AS, Goldstein B: **The complexity of complexes in signal transduction.** *Biotechnol Bioeng* 2004, **84(7):**783-794.
2. Conzelmann H, Saez-Rodriguez J, Sauter T, Bullinger E, Allgöwer F, Gilles ED: **Reduction of Mathematical Models of Signal Transduction Networks: Simulation-Based Approach Applied to EGF Receptor Signaling.** *IEE Systems Biology* 2004, **1:**159-169.
3. Eungdamrong NJ, Iyengar R: **Modeling cell signaling networks.** *Biology of the Cell* 2003, **96(5):**355-362.
4. Goldstein B, Faeder JR, Hlavacek WS: **Mathematical and computational models of immune-receptor signaling.** *Nat Rev Immunol* 2004, **4(6):**445-456.
5. Sauro HM, Kholodenko BN: **Quantitative analysis of signaling networks.** *Prog Biophys Mol Biol* 2004, **86:**5-43.
6. Schoeberl B, Eichler-Jonsson C, Gilles ED, Müller G: **Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors.** *Nat Biotechnol* 2002, **20(4):**370-375.
7. Wiley HS, Shvartsman SY, Lauffenburger DA: **Computational modeling of the EGF-receptor system: a paradigm for systems biology.** *Trends Cell Biol* 2003, **13:**43-50.
8. Kholodenko BN, Demin OV, Moehren G, Hoek JB: **Quantification of Short Term Signaling by the Epidermal Growth Factor Receptor.** *J Biol Chem* 1999, **274(42):**30169-30181.
9. Bhalla US, Iyengar R: **Emergent Properties of Networks of Biological Signaling Pathways.** *Science* 1999, **283(5400):**381-387.
10. Hatakeyama M, Kimura S, Naka T, Kawasaki T, Yumoto N, Ichikawa M, Kim JH, Saito K, Saeki M, Shirouzu M, Yokoyama S, Konagaya A: **A computational model on the modulation of mitogen-activated protein kinase (MAPK) and Akt pathways in heregulin-induced ErbB signalling.** *Biochem J* 2003, **373(Pt 2):**451-463.
11. Morton-Firth CJ, Bray D: **Predicting temporal fluctuations in an intracellular signalling pathway.** *J Theor Biol* 1998, **192:**117-128.
12. Faeder JR, Blinov ML, Goldstein B, Hlavacek WS: **Combinatorial complexity and dynamical restriction of network flows in signal transduction.** *IEE Systems Biology* 2005, **2:**5-15.
13. Shimizu TS, Novere NL, Levin MD, Beavil AJ, Sutton BJ, Bray D: **Molecular model of a lattice of signalling proteins involved in bacterial chemotaxis.** *Nat Cell Biol* 2000, **2(11):**792-796.
14. Faeder JR, Hlavacek WS, Reischl I, Blinov ML, Metzger H, Redondo A, Wofsy C, Goldstein B: **Investigation of early events in Fc(epsilon) RI-mediated signaling using a detailed mathematical model.** *J Immunol* 2003, **170(7):**3769-81.
15. Blinov ML, Faeder JR, Goldstein B, Hlavacek WS: **BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains.** *Bioinformatics* 2004.
16. Borisov NM, Markevich NI, Hoek JB, Kholodenko BN: **Signaling through receptors and scaffolds: independent interactions reduce combinatorial complexity.** *Biophys J* 2005, **89(2):**951-966.
17. Pawson T, Nash P: **Assembly of cell regulatory systems through protein interaction domains.** *Science* 2003, **300(5618):**445-452.
18. Schlessinger J: **Cell signaling by receptor tyrosine kinases.** *Cell* 2000, **103(2):**211-225.
19. Johnston AM, Pirola L, Obberghen EV: **Molecular mechanisms of insulin receptor substrate protein-mediated modulation of insulin signalling.** *FEBS Lett* 2003, **546:**32-36.
20. Blinov ML, Faeder JR, Goldstein B, Hlavacek WS: **A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity.** *Biosystems* 2005 in press.
21. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U: **Network motifs: simple building blocks of complex networks.** *Science* 2002, **298(5594):**824-7.
22. Isidori A: *Nonlinear Control Systems* 3rd edition. Springer. Global Decomposition of Control Systems; 1995:77-105.
23. Lauffenburger DA: **Cell signaling pathways as control modules: complexity for simplicity?** *Proc Natl Acad Sci U S A* 2000, **97(10):**5031-5033.
24. Saez-Rodriguez J, Kremling A, Gilles ED: **Dissecting the Puzzle of Life: Modularization of Signal Transduction Networks.** *Comput Chem Eng* 2005, **29(3):**619-629.

25. Saez-Rodriguez J, Kremling A, Conzelmann H, Bettenbrock K, Gilles ED: **Modular Analysis of Signal Transduction Networks.** *IEEE Contr Syst Mag* 2004, **24(4):**35-52.
26. Lindquist JA, Simeoni L, Schraven B: **Transmembrane adapters: attractants for cytoplasmic effectors.** *Immunol Rev* 2003, **191:**165-82.
27. Togni M, Lindquist J, Gerber A, Kolsch U, Hamm-Baarke A, Kliche S, Schraven B: **The role of adaptor proteins in lymphocyte activation.** *Mol Immunol* 2004, **41(6–7):**615-30.
28. Zhu M, Janssen E, Zhang W: **Minimal requirement of tyrosine residues of linker for activation of T cells in TCR signaling and thymocyte development.** *J Immunol* 2003, **170:**325-333.
29. Kohn KW: **Molecular Interaction Map of the Mammalian Cell Cycle Control and DNA Repair Systems.** *Mol Biol Cell* 1999, **10(8):**2703-2734.