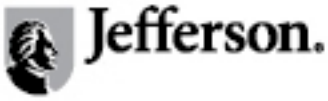



March 2006

# Trading the micro-world of combinatorial complexity for the macro-world of protein interaction domains

Nikolay M. Borisov
*Thomas Jefferson University*, nikolay.borisov@jefferson.edu

Nick I. Markevitch
*Thomas Jefferson University*, nikolai.markevitch@jefferson.edu

Jan B. Hoek
*Thomas Jefferson University*, jah.hoek@jefferson.edu

Boris N. Kholodenko
*Thomas Jefferson University*, Boris.Kholodenko@jefferson.edu

**Trading the micro-world of combinatorial complexity for the macro-world of protein interaction domains.**

Nikolay M. Borisov[1], Nick I. Markevich[1,2], Jan B. Hoek[1] and Boris N. Kholodenko[1*]

[1]Department of Pathology, Anatomy and Cell Biology, Thomas Jefferson University, 1020 Locust St., Philadelphia, PA 19107, USA, [2]Institute of Theoretical and Experimental Biophysics, Pushchino, Moscow reg., 142290

Running title:

**Reducing combinatorial complexity of multi-domain signaling**

**Key words:** scaffold, tyrosine kinase receptor, adapter protein, complex signaling networks, time series, model reduction

[*]Correspondence should be addressed to:
Boris N. Kholodenko
Dept. of Pathology, Anatomy and Cell Biology
Thomas Jefferson University, JAH
1020 Locust Street
Philadelphia, PA 19107
E-mail: Boris.Kholodenko@jefferson.edu, Fax: (215) 923-2218

**Abstract**

Membrane receptors and proteins involved in signal transduction display numerous binding domains and operate as molecular scaffolds generating a variety of parallel reactions and protein complexes. The resulting combinatorial explosion of the number of feasible chemical species and, hence, different states of a network greatly impedes mechanistic modeling of signaling systems. Here we present novel general principles and identify kinetic requirements that allow us to replace a mechanistic picture of all possible micro-states and transitions by a macro-description of states of separate binding sites of network proteins. This domain-oriented approach dramatically reduces computational models of cellular signaling networks by dissecting mechanistic trajectories into the dynamics of macro- and meso-variables. We specify the conditions when the temporal dynamics of micro-states can be exactly or approximately expressed in terms of the product of the relative concentrations of separate domains. We prove that our macro-modeling approach equally applies to signaling systems with low population levels, analyzed by stochastic rather than deterministic equations. Thus, our results greatly facilitate quantitative analysis and computational modeling of multi-protein signaling networks.

**INTRODUCTION.**

Processing of extracellular signals involves covalent modification of amino acid residues on cell-surface receptors and cytoplasmic signaling proteins (Bray, 1998; Hunter, 2000). For instance, receptors that belong to the large family of receptor tyrosine kinases (RTK) modify tyrosine residues on proteins by attaching a phosphate group (Schlessinger, 2000). Tyrosine phosphorylation delivers a message for a plethora of binding partners, including adaptor proteins and effector enzymes, such as protein and lipid kinases and phosphatases (Pawson and Nash, 2003).

Receptors and various signaling proteins contain a number of different domains that display a multitude of phosphorylation states and generate a large variety of heterogeneous multi-protein complexes (Pawson and Scott, 1997). Even for a few initial steps in transduction of extracellular cues, a combinatorial variety of signaling processes involving receptors, scaffolds and adapters may generate hundreds of thousands of molecular species (Faeder et al., 2003; Hlavacek et al., 2003). For instance, following activation, insulin receptor (IR) and insulin-like growth factor receptor-1 (IGF-1R) bind various combinations of downstream targets, including growth factor receptor binding protein-2 (Grb2), the Src homology and collagen domain protein (Shc), the p85 subunit of phosphatidylinositol 3-kinase (PI3K) and scaffolding adaptor proteins known as insulin receptor substrates, IRS1 – IRS6 (Cai et al., 2003; Paz et al., 1996; Saltiel and Pessin, 2002; White, 2002). Likewise, receptor-mediated phosphorylation of multiple tyrosine residues on the IRS molecules creates docking sites for a variety of SH2 domain-containing proteins, such as Grb2, PI3K, soluble tyrosine kinases Src and Fyn, RasGAP, and the protein tyrosine phosphatase SHP-2 (Cai et al., 2003; Myers et al., 1996; White, 1998). In addition, both IR and IGF-1R can bind and phosphorylate the scaffold adaptors known as Grb2-associated binders (GAB-1 and GAB-2) (Ingham et al., 1998; Lehr et al., 2000; Rodrigues et al., 2000). In turn, phosphorylated GABs bind numerous targets at their docking sites, including Shc, Grb2, p85, phospholipase Cγ, SHP-2, and the Crk adapter protein (Nishida and Hirano, 2003; Yamasaki et al., 2003).

Different docking sites initiate separate signaling pathways that propagate distinct cellular responses. For instance, a pathway initiated by the binding of Grb2 to tyrosine phosphorylated GAB and the recruitment of the GDP/GTP exchange factor SOS enables activation of the small GTPase Ras and leads to activation of the mitogen-activated protein kinase (MAPK) cascade, which promotes mitogenesis and differentiation. Independent binding of p85 to GAB initiates the PI3K/AKT pathway implicated in glucose and lipid metabolism and cell survival (Luo et al., 2003; Shepherd et al., 1998). Owing to the multiplication of all different possibilities, the number of feasible molecular species in

the IR and IGF1-R pathways readily reaches hundreds of thousands and even a million, hampering a rigorous quantitative analysis and mathematical modeling of these signaling pathways.

An entire collection of potential molecular species that emerge as different forms of receptors, scaffold proteins and signaling enzymes will be referred to as a set of "micro-states". In a framework of a standard mechanistic description, all these possible micro-states and their chemical transformations are taken into account (Faeder et al., 2003; Hlavacek et al., 2003; Levchenko et al., 2000). Because of exceedingly high numbers of micro-states, previous models of large signaling networks merely ignored a combinatorial variety of feasible state combinations, focusing on experimentally detected protein complexes (Asthagiri and Lauffenburger, 2001; Hatakeyama et al., 2003; Haugh et al., 1999; Haugh et al., 2000; Kholodenko et al., 1999; Moehren et al., 2002; Schoeberl et al., 2002). The potential problem for network modeling that arises from the combinatorial complexity of signal transduction has been recognized (Kholodenko et al., 1999; Morton-Firth and Bray, 1998; Shimizu et al., 2000), and several approaches toward handling it have been suggested. An elegant algorithm to account for all potential species and reactions was developed and implemented in the rule-based software tool BioNetGen (Blinov et al., 2004). However, due to the enormous number of distinct chemical species and a lack of knowledge of the kinetics for each possible transformation, such a detailed "micro-description" rapidly becomes impractical. The structure of reduced models that concentrate on the predominant species is still difficult to predict, and this structure can be highly dependent on the values of kinetic parameters.

An appealing stochastic approach to modeling multistate signaling systems involved in bacterial chemotaxis was suggested by Bray and colleagues (Morton-Firth and Bray, 1998; Shimizu et al., 2000). In the computer program StochSim, individual multistate complexes present distinct software objects. Consequently, a combinatorial explosion of the number of micro-states is circumvented by following stochastic changes in the states of individual, distinguishable molecules, the number of which does not increase in the course of simulations (Le Novere and Shimizu, 2001). Still, for large networks of hundreds of different proteins, the StochSim calculation time would be exceedingly slow, increasing proportionally to the number of molecules squared.

Recently, a novel approach has been introduced that replaces a mechanistic picture of all possible states by a macro-model that analyzes the states of individual protein domains/sites, such as the phosphorylation levels and the fractions occupied by binding partners (Borisov et al., 2005; Conzelmann et al., 2005). In the present paper this approach is extended to include macro-models of receptors operating as scaffolds and stochastic simulations of signaling systems with low population

numbers. A general modeling framework proposed here drastically reduces the number of states and differential equations to be solved and, therefore, the computational cost of both deterministic and stochastic simulations. We demonstrate that a necessary prerequisite for the reduction of a mechanistic model is the presence of protein domains/sites that do not influence other sites, allosterically or through interactions with bound partners. Importantly, the existence of additional sites involved in allosteric interactions does not impede the reduction of combinatorial complexity of multicomponent receptor-mediated signal transduction. We show when and how a mechanistic description of a signaling network can be expressed explicitly or approximated in terms of a "domain-oriented", reduced model.

## METHODS

We will analyze typical motifs of cellular signaling networks, including signal transduction through (*i*) a cell-surface receptor and (*ii*) an adapter protein. Both receptor and adapter protein may act as scaffolds and display multiple sites that can bind various partners simultaneously. A key property that will allow us to reduce combinatorial complexity of these networks is the assumption of the absence of allosteric interactions for a subset of domains/docking sites on a scaffold protein.

**1 Receptors with multiple binding sites.** We consider a cell-surface receptor ($R$) that dimerizes and/or undergoes autophosphorylation upon binding a ligand ($L$). Dimerization and/or phosphorylation of the activation loop in the kinase domain leads to a significant increase in kinase activity and enables phosphorylation of a number of docking sites, located in the cytoplasmic tail of the receptor. When phosphorylated, these docking sites can engage several adapter proteins (and their complexes) or can be dephosphorylated by phosphatases ($P$).

Various domains/sites are involved in different aspects of receptor signaling and are differentially controlled. We conceptually divide all binding sites of the receptor into three different groups according to their places in the hierarchical control of receptor signaling and the dependences/influences between the sites within single or distinct groups. At the top level of this hierarchy is the group of sites referred to as *controlling* sites. Their states influence chemical transformations of sites of other groups. For instance, agonist-binding site(s) on a receptor and the residues that are phosphorylated within the activation loop in the kinase domain are controlling sites, since their states control phosphorylation of multiple docking sites on the receptor. A controlling site may also influence the states of the other controlling sites. For instance, binding a ligand may induce receptor dimerization and/or activation of the kinase domain. To describe the system quantitatively, it

is convenient to assign digital flags (numbers) to possible states of binding sites. The states of each of $m$ controlling sites are described by numbers $h_j = 0,\ldots, H_j$ ($j = 1,\ldots,m$). For instance, if controlling site 1 stands for the ligand-binding site on the receptor, $h_1 = 0$ and $h_1 = 1$ correspond to the free and ligand-bound state, respectively. Similarly, $h_2$=0, 1 can stand for receptor monomer and dimer, respectively, $h_3$ =0, 1 describes activation loop phosphorylation, $h_4$ =0,1 can indicate that the phosphatase is or is not bound to a specific site, and so on.

Receptor docking sites are also subdivided into two groups. They are referred to as *independent docking* (or *a*-type) sites and *allosterically interacting docking* (*b*-type) sites. The chemical transformations of any *a*-type site are assumed to be independent of the states of the other *a*-type sites and may depend only on the states of controlling (*h*-type) sites. Possible states of an *a*-type site are denoted by $a_i = 0, 1, \ldots, A_i$, where $a_i = 0$ indicates an unphosphorylated free site, $a_i = 1$ denotes a phosphorylated free site, $a_i = 2$ represents a site occupied by a binding partner ($T_s$), $a_i = 3$ can stand for phosphorylation of this partner ($T_sp$) or binding a new partner ($T_j$), $a_i = 4$ can denote the binding of the complex $T_sp{\cdot}T_j$, and so on. Finally, *b*-type docking sites interact with each other allosterically or through their bound proteins. For instance, a kinase bound to a *b*-type docking site can phosphorylate a neighboring site. We assume that *b*-type docking sites do not influence the transformations of *independent docking* sites and *controlling* sites.

To summarize, micro-states of a receptor molecule can be described by a vector ($a_1,\ldots, a_n, b_1,\ldots, b_q, h_1, \ldots, h_m$), where $a_j = 0,\ldots, A_j$, $j = 1,\ldots, n$ refers to the state $a_j$ of $j$-th *independent docking* site, whereas $b_j = 0, \ldots, B_j$, $j = 1,\ldots, q$ refers to the state $b_j$ of $j$-th *allosterically interacting* site and $h_j = 0,\ldots, H_j$, $j = 1,\ldots,m$ refers to the state $h_j$ of $j$-th *controlling* site. Possible influences and dependences of different types of sites and the requirements for the rate constants of the chemical transformations are summarized in Table 1.

Assuming mass-action kinetics for receptor activation, the temporal evolution of receptor concentrations, $r(a_1,\ldots, a_n, b_1,\ldots, b_q, h_1, \ldots, h_m, t)$ is governed by the following system of $(A_1+1)\cdot\ldots\cdot(A_n+1)\cdot(B_1+1)\cdot\ldots\cdot(B_q+1)\cdot(H_1+1)\cdot\ldots\cdot(H_m+1)$ ordinary differential equations, which describe the chemical transformations of all feasible species in terms of transitions between (micro-)states of the receptor (to simplify designations, we will omit the variable $t$ where this does not result in misunderstanding),

$$\frac{dr(\mathbf{a},\mathbf{b},\mathbf{h})}{dt} = -\sum_{j=1}^{n}\sum_{\substack{\tilde{a}_j=0 \\ \tilde{a}_j\neq a_j}}^{A_j} k_{aj}(a_j \to \tilde{a}_j,\mathbf{h})r(a_1,...,a_{j\text{-}1},a_j,a_{j+1},...,a_n,\mathbf{b},\mathbf{h}) +$$

$$\sum_{j=1}^{n}\sum_{\substack{\tilde{a}_j=0 \\ \tilde{a}_j\neq a_j}}^{A_j} k_{aj}(\tilde{a}_j \to a_j,\mathbf{h})r(a_1,...,a_{j\text{-}1},\tilde{a}_j,a_{j+1},...,a_n,\mathbf{b},\mathbf{h}) -$$

$$\sum_{j=1}^{q}\sum_{\substack{\tilde{b}_j=0 \\ \tilde{b}_j\neq b_j}}^{B_j} k_{bj}(b_j \to \tilde{b}_j,\mathbf{b},\mathbf{h})r(\mathbf{a},b_1,...,b_{j\text{-}1},b_j,b_{j+1},...,b_q,\mathbf{h}) +$$

$$\sum_{j=1}^{q}\sum_{\substack{\tilde{b}_j=0 \\ \tilde{b}_j\neq b_j}}^{B_j} k_{bj}(\tilde{b}_j \to b_j,\mathbf{b},\mathbf{h})r(\mathbf{a},b_1,...,b_{j\text{-}1},\tilde{b}_j,b_{j+1},...,b_q,\mathbf{h}) -$$

$$\sum_{j=1}^{m}\sum_{\substack{\tilde{h}_j=0 \\ \tilde{h}_j\neq h_j}}^{H_j} k_{hj}(h_j \to \tilde{h}_j,\mathbf{h})r(\mathbf{a},\mathbf{b},h_1,...,h_{j\text{-}1},h_j,h_{j+1},...,h_m) +$$

$$\sum_{j=1}^{m}\sum_{\substack{\tilde{h}_j=0 \\ \tilde{h}_j\neq h_j}}^{H_j} k_{hj}(\tilde{h}_j \to h_j,\mathbf{h})r(\mathbf{a},\mathbf{b},h_1,...,h_{j\text{-}1},\tilde{h}_j,h_{j+1},...,h_m). \qquad (1)$$

Here $k_{aj}(\xi\to\eta)$, $k_{bj}$ $(\xi\to\eta)$, and $k_{hj}$ $(\xi\to\eta)$ stand for (pseudo-)first order rate constants of the transformation of the $j$-th $a$-site, $b$-site, and $h$-site, respectively from the state $\xi$ to the state $\eta$. If this is a binding reaction, the $k_{aj}$, $k_{bj}$ , $k_{hj}$ include the concentration of a binding protein as a multiplier.

Eq. 1 is illustrated by a simplified model of IR or IGF-1R activation. Both these receptors exist as preformed dimers containing two α and two β subunits linked by disulfide bridge (Munshi et al., 2003; Ottensmeyer et al., 2000). For the sake of simplicity, we assume that each IR dimer binds only one ligand molecule, since strong negative cooperativity for ligand binding to IR is reported {Pang, 1984 #71}. Upon ligand binding, IR undergoes autophosphorylation at three neighboring tyrosine residues in the activation loop of the intracellular kinase domain, which causes a 36-fold increase in its kinase activity (Ablooglu and Kohanski, 2001). The kinase domain of IR can phosphorylate the docking sites in the cytoplasmic receptor tail (for the sake of simplicity, only a single docking site is considered on each monomer).

Fig. 1 illustrates the minimal model of IR that uses four digital flags. Two controlling sites are described by the variables $h_1$ and $h_2$ ($h_1 = 0$ and $h_1 = 1$ correspond to the free and ligand-bound forms of R, respectively, and $h_2 = 0$ and $h_2 = 1$ stand for unphosphorylated and phosphorylated activation loop, respectively). We assume that phosphorylation of two independent docking sites in the

cytoplasmic receptor tail (described by the variables $a_1$ and $a_2$) occurs only when the activation loop is phosphorylated, whereas the dephosphorylation of these sites by phosphatases can occur regardless the states of controlling sites (Ablooglu and Kohanski, 2001). To simplify the resulting equations, we do not consider states of docking sites with adapter protein bound, restricting the set of docking site states to 0 (unphosphorylated) and 1 (phosphorylated). Under conditions specified above, the primary steps of IR activation kinetics will be governed by the following differential equation,

$$\begin{aligned}\frac{dr(a_1,a_2,h_1,h_2)}{dt} = &(-1)^{a_1+1}\cdot\left(k_{p1}\cdot\delta_{h_2 1}\cdot r(0,a_2,h_1,1)-k_{-p1}\cdot r(1,a_2,h_1,h_2)\right)+\\ &(-1)^{a_2+1}\cdot\left(k_{p2}\cdot\delta_{h_2 1}\cdot r(a_1,0,h_1,1)-k_{-p2}\cdot r(a_1,1,h_1,h_2)\right)+\\ &(-1)^{h_1+1}\cdot\left(k_0\cdot L\cdot r(a_1,a_2,0,h_2)-k_{-0}\cdot r(a_1,a_2,1,h_2)\right)+\\ &(-1)^{h_2+1}\cdot\left(k_{act}\cdot\delta_{h_1 1}\cdot r(a_1,a_2,1,0)-k_{deact}\cdot r(a_1,a_2,h_1,1)\right).\end{aligned} \quad (2)$$

Here each of the four terms in the right-hand side of Eq. 2 describes transitions between the states $a_1$, $a_2$, $h_1$ and $h_2$, respectively; $k_0$ and $k_{-0}$ are the "on" and "off" rate constants for ligand binding to the receptor, $k_{act}$ and $k_{deact}$ stand for rate constants of the activation loop phosphorylation and dephosphorylation, $k_{pi}$ and $k_{-pi}$ are rate constants of docking site $i$ phosphorylation and dephosphorylation, respectively; $\delta_{ij}$ designates the Kronecker symbol ($\delta_{ij} = 1$, if $i = j$, otherwise $\delta_{ij} = 0$).

***2. Adapter protein with multiple binding sites.***

The formalism used above for cell-surface receptors (see Eq. 1) can be readily applied to an adapter protein ($S$) that displays several docking sites and acts as a scaffold. The docking sites on $S$ are phosphorylated by a receptor ($R$) and dephosphorylated by a phosphatase ($P$). Assuming the affinities of $S$ for $R$ and $P$ are independent of states of the $S$ docking sites, a mathematical description of this system is similar to the receptor case described by Eq 1. In fact, the states of the docking sites are described identically for the receptor in Eq. 1 and the scaffold protein $S$, using the variables $a_j$ and $b_j$ for independent and allosterically interacting sites, respectively. However, a biological interpretation of controlling $h$-sites is different for $R$ and $S$. The variables $h_1$ and $h_2$ that describe two controlling sites on $S$ indicate if $S$ is bound to ($h_{1,2} = 1$) or dissociated from ($h_{1,2} = 0$) $R$ and $P$, respectively. The kinetic equations for the scaffold protein $S$ are readily obtained from Eq. 1 by the replacement of $r$(**a, b, h)** with $s$(**a**, **b**, $h_1$, $h_2$). We will illustrate this equation using a simple model of signal propagation through a scaffolding adapter protein. A scaffold can be phosphorylated at two docking sites when it is bound to a receptor kinase. Phosphorylated docking sites are dephosphorylated by a soluble phosphatase. We

assume that the affinity of the receptor for the scaffold is not affected by phosphorylation of the docking sites. Fig. 2 explains this scaffold model, which requires three digital flags. For the sake of simplicity, we assume that each of these docking sites can only be in two states, unphosphorylated ($a_i$ = 0, $i$ =1, 2) and phosphorylated ($a_i$ = 1). The third flag indicates the state of the controlling site: $h$ = 0 and $h$ = 1 stand for the scaffold bound and unbound to the receptor, respectively. The kinetics for scaffold activation is described as follows,

$$\begin{aligned}\frac{ds(a_1,a_2,h)}{dt} = &(-1)^{a_1+1}\cdot\left(k_{p1}\cdot\delta_{h1}\cdot s(0,a_2,1)-k_{-p1}\cdot s(1,a_2,h)\right)+\\ &(-1)^{a_2+1}\cdot\left(k_{p2}\cdot\delta_{h1}\cdot s(a_1,0,1)-k_{-p2}\cdot s(a_1,1,h)\right)+\\ &(-1)^{h+1}\cdot\left(k_0\cdot R\cdot s(a_1,a_2,0)-k_{-0}\cdot s(a_1,a_2,1)\right).\end{aligned} \quad (3)$$

Here the rate constants $k_{pi}$ and $k_{-pi}$ are similar to those defined in Eq. 2, $k_0$ and $k_{-0}$ denote the "on" and "off" for binding the receptor to the scaffolding adapter protein.

**RESULTS**

**1. *Dissecting signal transduction by a scaffold protein into signaling by separate docking sites and their groups***

The signaling networks that involve multi-state scaffold proteins (such as cell-surface receptors and adapter proteins) are difficult to analyze quantitatively due to a combinatorial increase in the number of possible states. In this section, we will show how exploiting a domain-oriented approach, numerous micro-states of a network can be combined in terms of macro-states of separate domains or docking sites of signaling proteins. In the methods section, we described signaling states of a receptor-scaffold in terms of an (**a**,**b**,**h**)-formalism. Now, the use of the independence assumption will allow us to follow the fate of an $a$-type site (or a group of such sites) separately of reactions occurring at the other sites.

First, we introduce the "macro-variables" $R_i(a_i,\mathbf{h})$ by adding up the concentrations of all forms of the receptor $R$ that display a particular state ($a_i$) of independent docking site $i$ and states **h** of controlling sites (see Eq. 1),

$$R_i(a_i,\mathbf{h}) = \sum_{a_1=0}^{A_1}\ldots\sum_{a_{i-1}=0}^{A_{i-1}}\sum_{a_{i+1}=0}^{A_{i+1}}\ldots\sum_{a_n=0}^{A_n}\sum_{b_1=0}^{B_1}\ldots\sum_{b_q=0}^{B_q} r(a_1,\ldots,a_n,b_1,\ldots,b_q,\mathbf{h}),\quad i=1,\ldots,n. \quad (4)$$

The macro-variables $R_i(a_i,\ \mathbf{h})$ describes the states of each $a$-type docking site separately, regardless of states **b** of interacting docking sites. Next, we define a "meso-varibale" $R(\mathbf{b};\ \mathbf{h})$ that monitors a combination of states of allosterically interacting sites and controlling sites regardless of the states of $a$-type docking sites,

$$R(\mathbf{b},\mathbf{h}) = \sum_{a_1=0}^{A_1} \ldots \sum_{a_n=0}^{A_n} r(a_1,\ldots,a_n,\mathbf{b},\mathbf{h}). \quad (5)$$

The meso-variable $R(\mathbf{b},\mathbf{h})$ can be interpreted as the concentration of a group of allosterically interacting sites in the vector-state $\mathbf{b}$. Note that the same macro- and meso-variables are defined for the case of an adapter protein *S*.

The concept of interacting docking sites can be generalized for the case of several groups of *b*-sites, $\mathbf{b}_1$, $\mathbf{b}_2$, …, $\mathbf{b}_z$, when there are allosteric interactions *within* each group, but not *between* groups that are independent of each other. In this case, we introduce several mesoscopic variables $R_{\mathbf{b}1}(\mathbf{b}_1,\mathbf{h})$, …, $R_{\mathbf{b}z}(\mathbf{b}_z,\mathbf{h})$. Each of these variables $R_{\mathbf{b}i}$ is the sum of microscopic variables over the states of all *a*-sites and *b*-sites except the sites that belong to $\mathbf{b}_i$ group. Using Eqs 1, 4 and 5, we derive two reduced differential equation systems (models) that describe the time evolution of the macro-variables $R_i(a_i, h)$,

$$\begin{aligned}\frac{dR_i(a_i,\mathbf{h})}{dt} = &-\sum_{\substack{\tilde{a}_j=0 \\ \tilde{a}_j\neq a_j}}^{A_i} k_{aj}(a_j \to \tilde{a}_j,\mathbf{h})R_i(a_i,\mathbf{h}) + \sum_{\substack{\tilde{a}_j=0 \\ \tilde{a}_j\neq a_j}}^{A_i} k_{aj}(\tilde{a}_j \to a_j,\mathbf{h})R_i(\tilde{a}_i,\mathbf{h}) - \\ &\sum_{j=1}^{m}\sum_{\substack{\tilde{h}_j=0 \\ \tilde{h}_j\neq h_j}}^{H_j} k_{hj}(h_j \to \tilde{h}_j,\mathbf{h})R_i(a_i,h_1,\ldots,h_{j\text{-}1},h_j,h_{j+1},\ldots,h_m) + \\ &\sum_{j=1}^{m}\sum_{\substack{\tilde{h}_j=0 \\ \tilde{h}_j\neq h_j}}^{H_j} k_{hj}(\tilde{h}_j \to h_j,\mathbf{h})R_i(a_i,h_1,\ldots,h_{j\text{-}1},\tilde{h}_j,h_{j+1},\ldots,h_m),\end{aligned} \quad (6)$$

and the meso-variables $R(\mathbf{b}, \mathbf{h})$, respectively,

$$\begin{aligned}\frac{dR(\mathbf{b},\mathbf{h})}{dt} = &-\sum_{j=1}^{q}\sum_{\substack{\tilde{b}_j=0 \\ \tilde{b}_j\neq b_j}}^{B_j} k_{bj}(b_j \to \tilde{b}_j,\mathbf{b},\mathbf{h})R_b(b_1,\ldots,b_{j\text{-}1},b_j,b_{j+1},\ldots,b_q,\mathbf{h}) + \\ &\sum_{j=1}^{q}\sum_{\substack{\tilde{b}_j=0 \\ \tilde{b}_j\neq b_j}}^{B_j} k_{bj}(\tilde{b}_j \to b_j,\mathbf{b},\mathbf{h})R_b(b_1,\ldots,b_{j\text{-}1},\tilde{b}_j,b_{j+1},\ldots,b_q,\mathbf{h}) - \\ &\sum_{j=1}^{m}\sum_{\substack{\tilde{h}_j=0 \\ \tilde{h}_j\neq h_j}}^{H_j} k_{hj}(h_j \to \tilde{h}_j,\mathbf{h})R_b(\mathbf{b},h_1,\ldots,h_{j\text{-}1},h_j,h_{j+1},\ldots,h_m) + \\ &\sum_{j=1}^{m}\sum_{\substack{\tilde{h}_j=0 \\ \tilde{h}_j\neq h_j}}^{H_j} k_{hj}(\tilde{h}_j \to h_j,\mathbf{h})R_b(\mathbf{b},h_1,\ldots,h_{j\text{-}1},\tilde{h}_j,h_{j+1},\ldots,h_m).\end{aligned} \quad (7)$$

Eqs 6 and 7 show that the absence of allosteric interactions between docking *a*-sites and between groups of *a*- and *b*-type sites, allows us to analyze a signaling system in terms of macro- and meso-

variables that do not follow the fate of all the molecular species of the scaffold protein simultaneously. A description in terms of macro- and meso- variables is referred to as a macro-presentation of a signaling network. In the example of insulin receptor above (Fig. 1, Eq. 2), the macro-representation will use the variables $R_1(a_1,h_1,h_2) = r(a_1,0,h_1,h_2) + r(a_1,1,h_1,h_2)$ and $R_2(a_2, h_1,h_2) = r(0,a_2,h_1,h_2) + r(1,a_2,h_1,h_2)$. The differential equations for these macro-variables are obtained via summation of Eq. 2 through the variables $a_2$ and $a_1$, respectively:

$$\begin{aligned}\frac{dR_i(a_i,h_1,h_2)}{dt} = (-1)^{a_i+1}\cdot\left(k_{pi}\cdot\delta_{h_2 1}\cdot R_i(0,h_1,1) - k_{-pi}\cdot R_i(1,h_1,h_2)\right) + \\ (-1)^{h_1+1}\cdot\left(k_0\cdot L\cdot R_i(a_i,0,h_2) - k_{-0}\cdot R_i(a_i,1,h_2)\right) + \\ (-1)^{h_2+1}\cdot\left(k_{act}\cdot\delta_{h_1 1}\cdot R_i(a_i,1,0) - k_{deact}\cdot R_i(a_i,h_1,1)\right), \quad i=1,2.\end{aligned} \tag{8}$$

The scaffolding adapter protein (see Fig. 2, Eq. 3) is represented within the macroscopic framework using similar variables, $S_1(a_1,h) = s(a_1,0,h) + s(a_1,1,h)$ and $S_2(a_2, h) = s(0,a_2,h) + s(1,a_2,h)$, the temporal evolution of which satisfies the following equation,

$$\begin{aligned}\frac{dS_i(a_i,h)}{dt} = (-1)^{a_i+1}\cdot\left(k_{pi}\cdot\delta_{h1}\cdot S_i(0,h) - k_{-pi}\cdot S_i(1,h)\right) + \\ (-1)^{h+1}\cdot\left(k_0\cdot R\cdot S_i(a_i,0) - k_{-0}\cdot S_i(a_i,1)\right), \quad i=1,2.\end{aligned} \tag{9}$$

Introduction of macro- and meso-variables, considerably reduces the number of states and equations required for a quantitative analysis of the system behavior. In a domain-oriented framework, there are only $(A_1 +\ldots+ A_n+ n + (B_1+1)\cdot\ldots\cdot(B_q+1))\cdot(H_1+1)\cdot\ldots\cdot(H_m+1)$ states (which is equal to the number of differential equations in Eqs 6 and 7) of the scaffold protein instead of $(A_1+1)\cdot\ldots\cdot(A_n+1)\cdot(B_1+1)\cdot\ldots\cdot(B_q+1)\cdot(H_1+1)\cdot\ldots\cdot(H_m+1)$ micro-states (species). Note that the number of transitions in a transition graph decreases even more drastically. For instance, in a case with one controlling site, two independent docking sites and two allosterically interacting docking sites, each of which can display only two states, the number of microstates equals $2^5 = 32$, whereas the number of macro-states and meso-states is 16. The number of transitions decreases from $(32\cdot5)/2 = 80$ to $(4\cdot2+4\cdot2+8\cdot3)/2 = 20$. In the minimal models of insulin receptor and scaffolding adapter protein (Eqs 2-3 and 8-9), the number of macro-states ($2\cdot2\cdot2 + 2\cdot2\cdot2 = 16$ for the receptor and $2\cdot2 + 2\cdot2 = 8$ for the adapter) will not differ from the number of micro-states ($2^4 = 16$ for the receptor and $2^3 = 8$ for the adapter) because only two states of docking sites (free unphosphorylated and free phosphorylated) are considered. However, even for such a truncated fragment the number of transitions decreases from $(16\cdot4)/2 = 32$ for a micro-description to $(8\cdot3)/2 + (8\cdot3)/2 = 24$ for a macro-description of the receptor,

and from (8·3)/2 = 12 for a micro-description to (4·2)/2 + (4·2)/2 = 8 for a macro-description of the adapter.

**2. *Retrieving a mechanistic description from reduced, macro- and meso-type models***

We found that the time evolution of each *independent docking* (*a*-type) site and the dynamics of a set of *allosterically interacting* (*b*-type) sites can be described by reduced models that separately monitor variables $R_i(a_i,\mathbf{h})$ or $R(\mathbf{b},\mathbf{h})$) without requiring monitoring of the remaining sites on $R$ (Eqs. 6 and 7). However, often the concentration dynamics of a particular micro-state may be of interest, for instance, when the assembly of two or more molecules of certain interacting proteins on a receptor or scaffold is required for activation of a downstream target. We will now show how such a micro-description can be retrieved from reduced macro- and meso-descriptions at arbitrary time $t$.

***2.1. A scaffold with no controlling sites.*** We start with a simplified model of a scaffold that has no hierarchical control over states of its docking sites (although allosteric interactions between some docking sites may exist). Formally, this model does not incorporate any *h*-site and assumes that the concentration of the scaffold, $S$, in any micro-state is characterized by the time-varying function $s(a_1, \ldots, a_n, \mathbf{b}, t)$. The concentration $S_i(a_i,t)$ of $S$ in a particular macro-state ($a_i$) is defined as the sum of all forms with independent docking site $i$ in state $a_i$ (this macro-variable is given by Eq. 4 after replacement of $r$ and $R$ by $s$ and $S$, respectively, and omission of the vector **h**). The meso-variable $S(\mathbf{b},t)$ describes the overall concentration of all $S$ micro-states with a particular vector-state **b** of allosterically interacting sites and is defined similarly to Eq. 5. The conserved total concentration of the scaffold ($S_{tot}$) can be expressed in terms of micro-, macro- and meso-variables introduced above as follows,

$$S_{tot} = \sum_{a_1=0}^{A_1} \ldots \sum_{a_n=0}^{A_n} \sum_{b_1=0}^{B_1} \ldots \sum_{b_q=0}^{B_q} s(a_1,\ldots,a_n,b_1,\ldots,b_q) = \sum_{a_i=0}^{A_i} S_i(a_i) = \sum_{b_1=0}^{B_1} \ldots \sum_{b_q=0}^{B_q} S(b_1,\ldots,b_q). \quad (10)$$

Mutual independence of sites of the *a*-type and the absence of interactions between *a*-sites and *b*-sites allows us to conjecture that the probability of finding all the sites in states $a_1, \ldots, a_n, b_1, , \ldots, b_q$ simultaneously, equals the product of the probabilities of the corresponding states for each *a*-site and a group of *b*-sites. These probabilities are the fractional concentrations, i.e., the micro-, macro- and meso-variables normalized by the total concentration of the scaffold ($S_{tot}$). Assuming that at the initial time moment $t_0$, the "micro-probabilities" can be expressed as the product of the "macro-" and "meso-probabilities" and using Eqs 1 and 4-7, we can prove that the dynamics of the micro-states $s(a_1, \ldots, a_n; \mathbf{b}; t)$ is expressed in terms of the product of the fractional concentrations of macro- and meso-states for any given $t>t_0$ (Borisov et al., 2005),

$$s(a_1,\ldots,a_n,\mathbf{b}) = \frac{S(\mathbf{b})\cdot\prod_{i=1}^{n} S_i(a_i)}{S_{tot}^{n}}. \qquad (11)$$

***2.2. Constraints imposed by controlling hierarchy.*** In the general case, the control of chemical transformations of protein domains is hierarchical. For instance, the state of docking sites on a receptor is controlled by ligand binding and dimerization (Schlessinger, 2000). Therefore, the above analysis of a scaffold that lacks controlling *h*-sites cannot be applied to a receptor signaling as a scaffold. Even for a receptor (*R*) with several controlling *h*-sites, Eqs. 6 and 7 demonstrate that the time evolution of distinct sites of *a*-type and of the whole group of *b*-sites can be resolved into separate dynamics of the corresponding macro- and meso-variables. However, the temporal patterns of micro-states $r(\mathbf{a},\mathbf{b},\mathbf{h})$ cannot be *exactly* obtained from $R_i(a_i, \mathbf{h})$ and $R(\mathbf{b}, \mathbf{h})$ (Borisov et al., 2005). The culprits are the transitions between the states of controlling sites that impose common constraints on otherwise independent docking sites. For instance, activation of the kinase domain on the receptor leads to phosphorylation of docking sites and, hence, a correlation between their states. This suggests that it may be helpful to analyze cases, where the rates of transitions between *h*-states are much *faster* or much *slower* than the rates of changes of the states of *a*-sites and *b*-sites. In the former case, the ratio of the concentrations of the receptor-scaffold is with the same states of all *a*- and *b*-sites is fixed by the following rapid-equilibrium relations (Heinrich et al., 1977; Kholodenko et al., 1998),

$$\frac{r(\mathbf{a},\mathbf{b},h_1,\ldots,h_m)}{r(\mathbf{a},\mathbf{b},\tilde{h}_1,\ldots,\tilde{h}_m)} \approx \frac{R_i(a_i,h_1,\ldots,h_m)}{R_i(a_i,\tilde{h}_1,\ldots,\tilde{h}_m)} \approx \frac{R(\mathbf{b},h_1,\ldots,h_m)}{R(\mathbf{b},\tilde{h}_1,\ldots,\tilde{h}_m)} \approx \prod_{j=1}^{m}\frac{k_{hj}(\tilde{h}_j \to h_j)}{k_{hj}(h_j \to \tilde{h}_j)}. \qquad (12)$$

The overall concentrations of the molecular species displaying the same states of *a*- and *b*-sites do not change in the fast transitions between the states of *h*-sites. These overall concentrations are commonly referred to as the slow variables (Arnold et al., 1994; Fenichel, 1971; Heinrich et al., 1977; Kholodenko et al., 1998; Mischenko and Rozov, 1980) and are defined as (omitting the variable *t* to simplify designations),

$$\hat{r}(\mathbf{a},\mathbf{b}) = \sum_{h_1=0}^{H_1}\ldots\sum_{h_m=0}^{H_m} r(\mathbf{a},\mathbf{b},h_1,..,h_m), \quad \hat{R}_i(a_i) = \sum_{h_1=0}^{H_1}\ldots\sum_{h_m=0}^{H_m} R_i(a_i,h_1,..,h_m), \quad \hat{R}(\mathbf{b}) = \sum_{h_1=0}^{H_1}\ldots\sum_{h_m=0}^{H_m} R(\mathbf{b},h_1,..,h_m). \qquad (13)$$

Converting Eqs. 1, 6 and 7 into equations that involve the slow variables only, we demonstrate that the overall concentrations of micro-states can be expressed in terms of the overall concentrations of macro- and meso-states in a manner similar to Eq. 11,

$$\hat{r}(a_1,\ldots,a_n,\mathbf{b}) \approx \frac{\hat{R}(\mathbf{b})\cdot\prod_{i=1}^{n}\hat{R}_i(a_i)}{R_{tot}^{n}} \ . \qquad (14)$$

Here $R_{tot}$ is the conserved total receptor concentration

$$R_{tot} = \sum_{a_1=0}^{A_1}\ldots\sum_{a_n=0}^{A_n}\sum_{b_1=0}^{B_1}\ldots\sum_{b_n=0}^{B_n}\sum_{h_1=0}^{H_1}\ldots\sum_{h_m=0}^{H_m} r(a_1,\ldots,a_n,b_1,\ldots,b_q,h_1,\ldots,h_m) =$$

$$\sum_{a_i=0}^{A_i}\sum_{h_1=0}^{H_1}\ldots\sum_{h_m=0}^{H_m} R_i(a_i,h_1,\ldots,h_m) = \sum_{b_1=0}^{B_1}\ldots\sum_{b_n=0}^{B_n}\sum_{h_1=0}^{H_1}\ldots\sum_{h_m=0}^{H_m} R(b_1,\ldots,b_q,h_1,\ldots,h_m), i=1,\ldots,n. \qquad (15)$$

Using Eqs. 12-15, the concentrations of micro-states with a given vector (**h**) of states of *h*-sites are expressed as follows,

$$r(a_1,\ldots,a_n,\mathbf{b},\mathbf{h}) \approx \frac{R(\mathbf{b},\mathbf{h})\cdot\prod_{i=1}^{n}R_i(a_i,\mathbf{h})}{R_{tot}^{n}(\mathbf{h})} \ ,$$

$$R_{tot}(\mathbf{h}) = \sum_{a_1=0}^{A_1}\ldots\sum_{a_n=0}^{A_n}\sum_{b_1=0}^{B_1}\ldots\sum_{b_q=0}^{B_q} r(a_1,\ldots,a_n,b_1,\ldots,b_q,\mathbf{h}) = \sum_{a_i=0}^{A_i} R_i(a_i,\mathbf{h}) = \sum_{b_1=0}^{B_1}\ldots\sum_{b_q=0}^{B_q} R(b_1,\ldots,b_q,\mathbf{h}). \qquad (16)$$

Importantly, Eq. 16 suggests an alternative scaling of the fractional concentrations by exploiting the normalizing factor, $R_{tot}(\mathbf{h})$, which is the total concentration of all molecular forms with a certain vector-state **h** of controlling sites. In striking contrast to Eqs. 10 and 15, this total concentration $R_{tot}(\mathbf{h})$ depends on time. However, even for this alternative scaling of concentrations, Eq. 16 cannot be considered an exact relationship for all feasible values of the kinetic constants due to correlations between *a*- and *b*-sites imposed by the their dependence upon *h*-sites. Interestingly, in the other extreme case, when transitions between states of controlling *h*-sites are much *slower* than processes that change states of *a*- and b-sites, Eq. 16 continues to apply. In fact, because changes in **h** are very slow, one may consider the pseudo-equilibrium for the micro-states with a certain **h** separately from the pseudo-equilibrium for the states with another $\tilde{\mathbf{h}}$. Eqs. 11 and 14 imply that at constant **h**, the equilibrium values of $r(a_1, \ldots, a_n, \mathbf{b}, \mathbf{h})/R_{tot}(\mathbf{h})$ equal the product of $R_i(a_i, \mathbf{h})/R_{tot}(\mathbf{h})$ and $R(\mathbf{b})/R_{tot}$. Because of the rapid-equilibrium condition, the exact values of micro-state concentrations will not differ significantly from the *equilibrium* concentrations at all times.

Numerical experiments demonstrate that the concentrations of network micro-states are well approximated by Eq. 16 over a wide range of parameters. Here, we illustrate this by analyzing a simplified model of insulin receptor activation (Fig. 1 and Eq. 2). The precise concentrations of all

micro- and macro-states were calculated according to Eqs. 2 and 8. The approximate values of micro-state concentrations were obtained as the product of macro-state concentrations: $r_{est}(a_1,a_2,h_1,h_2) = R_1(a_1,h_1,h_2)\cdot R_2(a_2,h_1,h_2)/R_{tot}(h_1,h_2)$ according to Eq. 16. Fig. 3 shows the exact and approximate solutions for micro-states phosphorylated at both docking sites ($a_1 = 1$ and $a_2 = 1$) for all four possible states of controlling sites $h_1$ and $h_2$. As seen in Fig. 3, the approximate solution does not approach the exact solution even at steady state. To understand why this is so, recall that docking sites on IR cannot be phosphorylated when $h_2 = 0$ (the activation loop is dephosphorylated). Therefore, at any time, including $t\to\infty$ (steady state), there are non-zero fluxes between the micro-states with the same $a_1$ and $a_2$ but different $h_2$. In fact, these uncompensated fluxes are responsible for the difference between the exact and approximate solutions (see also (Borisov et al., 2005)). In the general case, the deviation of the approximate from the exact solution is difficult to estimate analytically. Numerical experiments suggest that the quality of approximation of micro-variables by the product of macro-variables is better when the activation loop is phosphorylated (panel B), than when it is unphosphorylated (panel A). To explain this phenomenon, we consider a limit case when (de)phosphorylation of the activation loop is much slower than transitions between states of docking sites. In this limit case, the pseudo-equilibrium solution for the micro-states with $a_1$, $a_2 = 1$ and $h_2 = 0$ equals zero, leading to a huge discrepancy with the approximate solution in terms of macro-states (Borisov et al., 2005). Even when the rate constants for (de)phosphorylation of the activation loop and docking sites on the receptor are comparable, the discrepancy between exact and approximate solutions is higher when $h_2 = 0$.

***3. Application of a domain-oriented approach for stochastic simulation of cell signaling.***

For small sub-cellular volumes where few molecules are involved, the deterministic approach may not be valid due to intrinsically random nature of signaling events that occur with each molecule in a network. Signaling systems comprising only a few molecules should be analyzed using stochastic methods (Morton-Firth and Bray, 1998; Rao et al., 2002). A Gillespie-type exact stochastic algorithm is related to the master equation and calculates the probability $p$ for the process $\mu$ to be the *next* reaction event (since the moment $t = 0$) that occurs during the time interval ($t$, $t + \mathrm{d}t$), given a certain molecular composition of the system at the moment $t = 0$ (Gillespie, 1977),

$$p(t,\mu) = h_\mu c_\mu \exp\left(-\sum_{\lambda=1}^{M} h_\lambda c_\lambda t\right). \qquad (17)$$

Here the summation goes through all the possible processes ($\lambda$); $h_\lambda$ and $c_\lambda$ are the number of distinct molecular reactant combinations and the rate parameter, respectively, for the process $\lambda$. For

unimolecular reactions, the number of reactant combinations equals the number $x_\lambda$ of molecules in the (micro-)state initial for the process λ, and the reaction parameter equals the first-order rate constant $k_\lambda$ of ordinary differential equations. For bimolecular reactions, the number of reactant combinations is the product of the initial numbers of two kinds of molecules $(x_\lambda)_1$ and $(x_\lambda)_2$ that participate in the process λ ($h_\nu = (x_\lambda)_1 \cdot (x_\lambda)_2$), and the reaction parameter equals the second-order rate constant of ordinary differential equations divided by the reaction volume ($c_\lambda = k_\lambda/V$) (Gillespie, 1977). Thus, probability $p(t,\mu)$ used in the stochastic Gillespie method and the related master equation uniquely correspond to ordinary differential equations, which are derived for the numbers of molecules in given molecular states.

In the models analyzed here, the chemical transformations of receptors and scaffold proteins are described in terms of (pseudo-) first-order processes (see Methods). Given ($a_1$, …, $a_{i\text{-}1}$, $a_i$, $a_{i+1}$ …, $a_n$) is the initial state (at $t = 0$), the probability that the first reaction event is the transition to the state ($a_1$, …, $a_{i\text{-}1}$, $\tilde{a}_i$, $a_{i+1}$,…, $a_n$) occurring during the time interval ($t$, $t + \mathrm{d}t$) is the following (for simplicity, we analyze here the case of completely independent sites on a scaffold that lacks controlling and allosterically interacting sites),

$$p(t,a_1,...,a_i,...,a_n \to a_1,...,\tilde{a}_i,...,a_n) =$$

$$x(a_1,...,a_i,...,a_n,t)k_{ai}(a_i \to \tilde{a}_i)\exp\left(-\sum_{j=1}^{n}\sum_{\alpha_j=0}^{A_j}\sum_{\substack{\tilde{\alpha}_j=0 \\ \tilde{\alpha}_j\neq\alpha_j}}^{A_j} k_{aj}(\alpha_j \to \tilde{\alpha}_j)\sum_{\alpha_1=0}^{A_1}...\sum_{\alpha_{j-1}=0}^{A_{j-1}}\sum_{\alpha_{j+1}=0}^{A_{j+1}}...\sum_{\alpha_n=0}^{A_n} x(\alpha_1,...,\alpha_j,...,\alpha_n,t)\cdot t\right). \quad (18)$$

The number of molecules in a certain macro-state is expressed in terms of macro-variables as follows,

$$X_i(a_i) = \sum_{a_1=0}^{A_1}...\sum_{a_{i-1}=0}^{A_{i-1}}\sum_{a_{i+1}=0}^{A_{i+1}}...\sum_{a_n=0}^{A_n} x(a_1,...,a_n). \quad (19)$$

The exponent in Eq. 18 can be rewritten in terms of macro-variables $X_i$ as follows,

$$p(t,a_1,...,a_i,...,a_n \to a_1,...,\tilde{a}_i,...,a_n) =$$

$$x(a_1,...,a_i,...,a_n,t)k_{ai}(a_i \to \tilde{a}_i)\exp\left(-\sum_{j=1}^{n}\sum_{\alpha_j=0}^{A_j}\sum_{\substack{\tilde{\alpha}_j=0 \\ \tilde{\alpha}_j\neq\alpha_j}}^{A_j} k_{aj}(\alpha_j \to \tilde{\alpha}_j)X_j(\alpha_j,t)\cdot t\right). \quad (20)$$

Summing up the left- and right-hand-sides of Eq. 20 over states of all but $i$-th docking site, we arrive at the following equation for the probability that the next event is the conversion of the macro-state $a_i$ to the other state $\tilde{a}_i$:

$$P_i(t, a_i \to \tilde{a}_i) = X_i(a_i, t) k_{ai}(a_i \to \tilde{a}_i) \exp\left( - \sum_{j=1}^{n} \sum_{\alpha_j=0}^{A_j} \sum_{\substack{\tilde{\alpha}_j=0 \\ \tilde{\alpha}_j \neq \alpha_j}}^{A_j} k_{aj}(\alpha_j \to \tilde{\alpha}_j) X_j(\alpha_j, t) \cdot t \right). \quad (21)$$

Here the probability of a macroscopic transition is the sum of probabilities of microscopic events over states of all docking sites except site *i*. Note that this macroscopic transition includes those microscopic events that only change the state of docking site *i* (from $a_i$ to $\tilde{a}_i$),

$$P_i(t, a_i \to \tilde{a}_i) = \sum_{a_1=0}^{A_1} \ldots \sum_{a_{i-1}=0}^{A_{i-1}} \sum_{a_{i+1}=0}^{A_{i+1}} \ldots \sum_{a_n=0}^{A_n} p(t, a_1, \ldots, a_i, \ldots, a_n \to a_1, \ldots, \tilde{a}_i, \ldots, a_n). \quad (22)$$

This equation shows that even when the pathway kinetics is simulated according to exact stochastic methods derived from the master equation, the use of a macroscopic, domain-oriented approach significantly reduces the number of distinct states and reactions to be accounted for, and, thus, the required calculation time

The presence of controlling and allosterically interacting sites on a scaffold protein does not prevent a domain-oriented approach to be applied to stochastic simulations of signaling networks. The states of controlling sites are taken into account for each of macro-variables, $X_i(a_i,\mathbf{h})$; whereas the fate of *b*-type sites are followed by mesoscopic variable, $X(\mathbf{b},\mathbf{h})$ defined similarly to the macro- and mesoscopic concentration variables, introduced above (Eqs 4 and 5).

Remarkably, for stochastic kinetic equations, microscopic variables are retrieved in terms of macro- and mesoscopic states in a manner analogous to the deterministic case (see above, Eq 16). Numerical experiments below show this for an example of signaling by a scaffolding adapter protein (*S*) with two independent docking sites and one controlling site (see Fig. 2 and Eq. 3). Fig. 4 covers three cases of different sub-cellular volumes of $10^{-13}$ l (panels A1 and A2), $10^{-14}$ l (panels B1 and B2) and $10^{-15}$ l (panels C1 and C2). The active receptor and total scaffold concentrations at $t = 0$ are assumed to be 100 nM (Kholodenko et al., 1999), which results in the total numbers of molecules of 6000, 600 and 60 for cases A, B and C, respectively. Panels A1, B1 and C1 show the number of molecules, $x(1,1,1)$ and $x(1,1,0)$ (which correspond to the receptor-bound and unbound scaffold with phosphorylated docking sites), and their estimations in terms of macro-states, $x_{est}(1,1,1)$ and $x_{est}(1,1,0)$. These values were calculated deterministically by multiplying the exact solutions of Eq. 3 and their estimates in terms of macro-states, $s_{est}(a_1,a_2;h) = S_1(a_1;h) \cdot S_2(a_2;h)/S_{tot}(h)$, by the reaction volume. As expected, the curves of panels A1, B1 and C1 are similar to one another, differing only in the numbers of molecules in different volumes. Panels A2, B2 and C2 show the numbers of molecules for the same

micro-statesand their estimates (calculated similarly to Eq. 16) in terms of macro-states obtained using Eqs. 18 and 21 by the stochastic Gillespie algorithm implemented in Jarnac (Sauro et al., 2003). The stochastic "noise" is seen in all the curves of these panels. Whereas the curves in panels A1 and A2 travel similar paths owing to the relatively large total number of molecules (6000) in the system, this is not at all true for curves in panels C1 and C2, where the stochastic fluctuations are comparable to the average value of the abundance of molecules in the corresponding states.

**DISCUSSION.**

A plethora of plasma membrane receptors and signaling proteins is involved in transfer and processing of extracellular information to various cytoplasmic targets, including the nucleus. Activated receptors and large adapter proteins often act as scaffolds that assemble multiprotein complexes on their docking sites. Following activation, each docking site on a scaffold can be covalently modified, e.g., phosphorylated (or dephosphorylated), and the phosphorylated site can be either free or occupied by a binding partner. A scaffolding adapter protein can either be associated with or dissociated from a receptor. Proteins bound to a scaffold can be phosphorylated and dephosphorylated, and may associate with other signaling proteins, assembling multicomponent complexes. All these different possibilities multiply leading to a combinatorial variety of feasible molecular species (micro-states) and extensively branched pathway graphs.

The present paper demonstrates that a mechanistic description of a highly combinatorial network generated by various phosphorylation and binding forms of receptors and scaffolds may be drastically reduced using a "domain-oriented" approach. This novel general framework for pathway macro-modeling is referred to as the (**a**,**b**,**h**)-formalism and exploits the hierarchy of the regulation of different domains on signaling proteins. A key prerequisite for the replacement of a micro-state description by this formalism is the existence of protein domains/sites (referred to as *a*-type sites) that do not influence other sites of that protein, allosterically or through their bound partners. In addition to *a*-type binding sites, other docking sites on a scaffold may interact allosterically. They are referred to as *b*-type docking sites. The states of each *b*-site may influence the chemical transformations occurring with any other *b*-type site, but do not influence the transformations of *a*-type docking sites. The hierarchy of control of docking sites on receptors and signaling scaffolds manifests itself through a group of specific sites, called *controlling* or *h*-type sites. They may influence the transitions between the states of all other sites on a signaling molecule. In models of receptor signaling, sites responsible for the ligand binding, receptor dimerization and phosphorylation of the activation loop are examples

of *controlling sites*. Importantly, the controlling hierarchy does not allow for the reverse interaction, in which *a*-sites and/or *b*-sites would influence the transformations of *h*-sites (for instance, ligand-receptor interactions are assumed to be independent of the states of receptor docking sites in our model of IR).

For many signaling proteins these conditions are fulfilled, and a signaling system is modeled in terms of the introduced "macro-formalism" that follows the states of different domains, including subsequent downstream signal transduction (Eqs. 6 and 7). In contrast to the combinatorial explosion of micro-states and equations in a mechanistic model, the number of the macro- and meso-states and equations is becoming tractable. Even in a simplified macro-model of a signaling protein with one controlling site, two independent and two allosterically interacting docking sites, each of which can display two states, the number of states and differential equations reduces from 32 to 16, and the number of transitions decreases from 80 to 20. Realistic models of signaling networks can provide even more impressive illustrations of how a macro-formalism can reduce the number of variables required for pathway description. For example, the EGFR network contains at least two important scaffold proteins, EGFR itself and the adapter protein GAB-1. Microscopic description of this network up to and including ERK activation would use roughly 163,000 variables (manuscript in preparation). However, only about 350 macro-variables are needed to monitor independent and allosterically interacting domains. Similarly, for the IR and IGF-1R networks, where scaffolds such as IRS-1, GAB-1 and Grb10 are involved, the assumption of independent interactions at docking sites provides a reduction in the number of equations from hundreds of thousands to a few hundreds. Yet the reduced description allows for accurate predictions of temporal signaling patterns.

We showed that, in some cases, a mechanistic description can be restored, in an exact or approximate manner, from the macro-formalism that operates with significantly reduced equation systems (see Eq. 16). Such a reconstruction of microscopic behavior is required when different micro-states within the same macro-state present biologically different activities. Fig. 3 illustrates the high quality of this reconstruction for a model of IR signaling.

For rarely populated signaling systems, the stochastic master equation describes the physical reality more closely than deterministic differential equations. Exact stochastic algorithms developed by Gillespie are related to the master equation and provide a tool for simulation of signaling systems with low population numbers (Gillespie, 1977). Owing to the combinatorial complexity of multiprotein receptor-mediated signaling systems, the computational cost of stochastic calculations appears to be very high. Eqs. 17-22 prove that our macro-state approach can be exploited in stochastic algorithms,

drastically reducing the number of variables and reactions to be considered and, hence, the calculation cost. Moreover, similar to the deterministic case, the microscopic variables can be approximated in terms of macro- and mesoscopic states, as illustrated in Fig. 4 for activation of a scaffold adaptor protein by a receptor in the confines of small subcellular volumes with low molecular numbers. We conclude that our domain-oriented approach can be applied to develop reduced, tractable models of "noisy" signaling systems.

**ACKNOWLEDGEMENTS**

This work was supported by the Grants GM59570 and AA08714 from the National Institute of Health.

**References**

Ablooglu, A.J., Kohanski, R.A., 2001. Activation of the insulin receptor's kinase domain changes the rate-determining step of substrate phosphorylation. Biochemistry, 40, 504-513.

Arnold, V., Afrajmovich, V., Ilyashenko, Y., Shil'nikov, L., 1994. Dynamical Systems V: Bifurcation Theory and Catastrophe Theory. In Encyclopaedia Math Sci. Springer-Verlag, Berlin.

Asthagiri, A.R., Lauffenburger, D.A., 2001. A computational study of feedback effects on signal dynamics in a mitogen-activated protein kinase (mapk) pathway model. Biotechnol Prog, 17, 227-239.

Blinov, M.L., Faeder, J.R., Goldstein, B., Hlavacek, W.S., 2004. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. Bioinformatics.

Borisov, N.M., Markevich, N.I., Hoek, J.B., Kholodenko, B.N., 2005. Signaling through receptors and scaffolds: independent interactions reduce combinatorial complexity. Biophys. J., (submitted for publication).

Bray, D., 1998. Signaling complexes: biophysical constraints on intracellular communication. Annu Rev Biophys Biomol Struct, 27, 59-75.

Cai, D., Dhe-Paganon, S., Melendez, P.A., Lee, J., Shoelson, S.E., 2003. Two new substrates in insulin signaling, IRS5/DOK4 and IRS6/DOK5. J Biol Chem, 278, 25323-25330.

Conzelmann, H., Saez-Rodriguez, J., Sauter, T., Kholodenko, B.N., Gilles, E.D., 2005. A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. 1st FEBS Advanced Lecture Courses "Systems Biology: From Molecules & Models to Cells".

Faeder, J.R., Hlavacek, W.S., Reischl, I., Blinov, M.L., Metzger, H., Redondo, A., Wofsy, C., Goldstein, B., 2003. Investigation of early events in Fc epsilon RI-mediated signaling using a detailed mathematical model. J Immunol, 170, 3769-3781.

Fenichel, N., 1971. Persistance and smoothness of invariant manifolds for flows. Indiana Univ Math J, 21, 193-255.

Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem., 81, 2340-2361.

Hatakeyama, M., Kimura, S., Naka, T., Kawasaki, T., Yumoto, N., Ichikawa, M., Kim, J.H., Saito, K., Saeki, M., Shirouzu, M., Yokoyama, S., Konagaya, A., 2003. A computational model on the modulation of mitogen-activated protein kinase (MAPK) and Akt pathways in heregulin-induced ErbB signalling. Biochem J, 373, 451-463.

Haugh, J.M., Schooler, K., Wells, A., Wiley, H.S., Lauffenburger, D.A., 1999. Effect of epidermal growth factor receptor internalization on regulation of the phospholipase C-gamma1 signaling pathway. J Biol Chem, 274, 8958-8965.
Haugh, J.M., Wells, A., Lauffenburger, D.A., 2000. Mathematical modeling of epidermal growth factor receptor signaling through the phospholipase C pathway: mechanistic insights and predictions for molecular interventions. Biotechnol Bioeng, 70, 225-238.
Heinrich, R., Rapoport, S.M., Rapoport, T.A., 1977. Metabolic regulation and mathematical models. Prog Biophys Mol Biol, 32, 1-82.
Hlavacek, W.S., Faeder, J.R., Blinov, M.L., Perelson, A.S., Goldstein, B., 2003. The complexity of complexes in signal transduction. Biotechnol Bioeng, 84, 783-794.
Hunter, T., 2000. Signaling--2000 and beyond. Cell, 100, 113-127.
Ingham, R.J., Holgado-Madruga, M., Siu, C., Wong, A.J., Gold, M.R. 1998. The Gab1 protein is a docking site for multiple proteins involved in signaling by the B cell antigen receptor. J Biol Chem, 273, 30630-30637.
Kholodenko, B.N., Demin, O.V., Moehren, G., Hoek, J.B., 1999. Quantification of short term signaling by the epidermal growth factor receptor. J Biol Chem, 274, 30169-30181.
Kholodenko, B.N., Schuster, S., Garcia, J., Westerhoff, H.V., Cascante, M., 1998. Control analysis of metabolic systems involving quasi-equilibrium reactions. Biochim Biophys Acta, 1379, 337-352.
Le Novere, N., Shimizu, T.S., 2001. STOCHSIM: modelling of stochastic biomolecular processes. Bioinformatics, 17, 575-576.
Lehr, S., Kotzka, J., Herkner, A., Sikmann, A., Meyer, H.E., Krone, W., Muller-Wieland, D., 2000. Identification of major tyrosine phosphorylation sites in the human insulin receptor substrate Gab-1 by insulin receptor kinase in vitro. Biochemistry, 39, 10898-10907.
Levchenko, A., Bruck, J., Sternberg, P.W., 2000. Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. Proc Natl Acad Sci U S A, 97, 5818-5823.
Luo, H.R., Hattori, H., Hossain, M.A., Hester, L., Huang, Y., Lee-Kwon, W., Donowitz, M., Nagata, E., Snyder, S.H., 2003. Akt as a mediator of cell death. Proc Natl Acad Sci U S A, 100, 11712-11717.
Mischenko, E., Rozov, N., 1980. Differential Equations with Small Parameters and Relaxation Oscillations. Plenum Press, New York.
Moehren, G., Markevich, N., Demin, O., Kiyatkin, A., Goryanin, I., Hoek, J.B., Kholodenko, B.N., 2002. Temperature dependence of the epidermal growth factor receptor signaling network can be accounted for by a kinetic model. Biochemistry, 41, 306-320.
Morton-Firth, C.J., Bray, D., 1998. Predicting temporal fluctuations in an intracellular signalling pathway. J Theor Biol, 192, 117-128.
Munshi, S., Hall, D.L., Kornienko, M., Darke, P.L., Kuo, L.C., 2003. Structure of apo, unactivated insulin-like growth factor-1 receptor kinase at 1.5 A resolution. Acta Crystallogr D Biol Crystallogr, 59, 1725-1730.
Myers, M.G., Jr., Zhang, Y., Aldaz, G.A., Grammer, T., Glasheen, E.M., Yenush, L., Wang, L.M., Sun, X.J., Blenis, J., Pierce, J.H., White, M.F., 1996. YMXM motifs and signaling by an insulin receptor substrate 1 molecule without tyrosine phosphorylation sites. Mol Cell Biol, 16, 4147-4155.
Nishida, K., Hirano, T., 2003. The role of Gab family scaffolding adapter proteins in the signal transduction of cytokine and growth factor receptors. Cancer Sci, 94, 1029-1033.
Ottensmeyer, F.P., Beniac, D.R., Luo, R.Z., Yip, C.C., 2000. Mechanism of transmembrane signaling: insulin binding and the insulin receptor. Biochemistry, 39, 12103-12112.
Pawson, T., Nash, P., 2003. Assembly of cell regulatory systems through protein interaction domains. Science, 300, 445-452.

Pawson, T., Scott, J.D., 1997. Signaling through scaffold, anchoring, and adaptor proteins. Science, 278, 2075-2080.
Paz, K., Voliovitch, H., Hadari, Y.R., Roberts, C.T., Jr., LeRoith, D., Zick, Y., 1996. Interaction between the insulin receptor and its downstream effectors. Use of individually expressed receptor domains for structure/function analysis. J Biol Chem, 271, 6998-7003.
Rao, C.V., Wolf, D.M., Arkin, A.P., 2002. Control, exploitation and tolerance of intracellular noise. Nature, 420, 231-237.
Rodrigues, G.A., Falasca, M., Zhang, Z., Ong, S.H., Schlessinger, J., 2000. A novel positive feedback loop mediated by the docking protein Gab1 and phosphatidylinositol 3-kinase in epidermal growth factor receptor signaling. Mol Cell Biol, 20, 1448-1459.
Saltiel, A.R., Pessin, J.E., 2002. Insulin signaling pathways in time and space. Trends Cell Biol, 12, 65-71.
Sauro, H.M., Hucka, M., Finney, A., Wellock, C., Bolouri, H., Doyle, J., Kitano, H., 2003. Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. Omics, 7, 355-372.
Schlessinger, J., 2000. Cell signaling by receptor tyrosine kinases. Cell, 103, 211-225.
Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D., Muller, G., 2002. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nat Biotechnol, 20, 370-375.
Shepherd, P.R., Withers, D.J., Siddle, K., 1998. Phosphoinositide 3-kinase: the key switch mechanism in insulin signalling. Biochem J, 333 ( Pt 3), 471-490.
Shimizu, T.S., Le Novere, N., Levin, M.D., Beavil, A.J., Sutton, B.J., Bray, D., 2000. Molecular model of a lattice of signalling proteins involved in bacterial chemotaxis. Nat Cell Biol, 2, 792-796.
White, M.F., 1998. The IRS-signalling system: a network of docking proteins that mediate insulin action. Mol Cell Biochem, 182, 3-11.
White, M.F., 2002. IRS proteins and the common path to diabetes. Am J Physiol Endocrinol Metab, 283, E413-422.
Yamasaki, S., Nishida, K., Yoshida, Y., Itoh, M., Hibi, M., Hirano, T., 2003. Gab1 is required for EGF receptor signaling and the transformation by activated ErbB2. Oncogene, 22, 1546-1556.

**FIGURE LEGENDS**

**Fig. 1. Digital flags denoting domain states of the insulin receptor.** The receptor-dimer $R$ binds the ligand $L$ via domain $h_1$. For unbound $L$, $h_1 = 0$, and when $R$ is bound to $L$, $h_1 = 1$. Binding the ligand causes autophosphorylation of the activation loop, which switches on the tyrosine kinase activity of the receptor. Variable $h_2$ represents the activity of the kinase domain: $h_2 = 0$ the unphosphorylated loop and inactive kinase domain, $h_2 = 1$ phosphorylated loop and active kinase domain. The active kinase domain phosphorylates two tyrosine docking sites in the cytoplasmic tail, whose states, $a_1$ and $a_2$, respectively, are described as follows: 0 indicates free unphosphorylated site (Y); 1 stands for free phosphorylated site (pY). Arrows indicate hierarchical control relationships, i.e. transitions between different values of $h_2$ are affected by $h_1$, and transitions between different values of $a_1$ or $a_2$ are affected by $h_2$.

**Fig. 2. Digital flags denoting domain states of the scaffolding adapter protein.** Adapter protein $S$ binds via domain $h$ to the phosphorylated residue (pY) of the receptor tyrosine kinase $R$. For unbound $R$, $h = 0$, for bound $h = 1$. Binding to the receptor enables phosphorylation of two docking sites, whose states, $a_1$ and $a_2$, are described as follows: 0, free unphosphorylated site (Y); 1, free phosphorylated site (pY). Arrows indicate hierarchical control relationships, i.e. transitions between different values of $a_1$ or $a_2$ are affected by $h$.

**Fig. 3. Time courses of the receptor forms with both docking sites phosphorylated.** Micro-states $r(a_1, a_2, h_1, h_2)$ of a predimerized receptor activated by ligand binding ($h_1$) and autophosphorylation of the activation loop ($h_2$) are calculated using a mechanistic model (Eq. 2) and compared with their approximations obtained using a macro-description (Eqs. 8 and 16). Panels A and B show the micro-states, $r(1,1,0,0)$ and $r(1,1,1,0)$ (A), and $r(1,1,0,1)$ and $r(1,1,1,1)$, (B) and their estimates ($r_{est}$) in terms of macro-states, when the activation loop is unphosphorylated (A, $h_2 = 0$) and phosphorylated (B, $h_2 = 1$), respectively. The initial free ligand concentration and total receptor concentration are assumed to be 100 nM. The rate constants for the model are the following: $k_0 = 0.05$ nM$^{-1}\cdot$s$^{-1}$, $k_{-0} = 0.5$ s$^{-1}$, $K_d = 10$ nM; $k_{act} = 0.5$ s$^{-1}$, $k_{deact} = 0.5$ s$^{-1}$; $k_{p1} = 0.2$ s$^{-1}$, $k_{-p1} = 0.8$ s$^{-1}$; $k_{p2} = 0.8$ s$^{-1}$, $k_{-p2} = 0.2$ s$^{-1}$. The initial conditions for Eq. 2 and 8 were set as follows: $r(0,0,0,0) = 100$ nM, $r(0,0,1,0) = r(0,0,0,1) = r(0,0,1,1) = 1\cdot10^{-10}$ nM, whereas all other $r(a_1,a_2,h_1,h_2) = 0$; $R_1(0,0,0) = R_2(0,0,0) = 100$ nM, $R_1(0,1,0) = R_1(0,0,1) =$

$R_1(0,1,1) = R_2(0,1,0) = R_2(0,0,1) = R_2(0,1,1) = 1 \cdot 10^{-10}$ nM, and all other $R_i(a_i,h_1,h_2) = 0$. The freely available Jarnac software package was used for simulations (Sauro et al., 2003).

**Fig. 4. Time course of receptor-bound and unbound scaffold forms with both docking sites phosphorylated.** The number of molecules $x(1,1,1)$ and $x(1,1,0)$ correspond to micro-states of the phosphorylated scaffold ($a_1 = 1$, $a_2 = 1$) bound to ($h = 1$) or dissociated from ($h = 0$) the receptor. Their estimates ($x_{est}$) are made in terms of macro-states, similar to Eq. 16. Panels A, B, and C illustrate three cases, where the subcellular volume is $10^{-13}$, $10^{-14}$ and $10^{-15}$ l, respectively. The left and right panels (marked by numbers 1 and 2) present the results obtained via deterministic and stochastic algorithms, respectively. For stochastic simulation, every random event (molecular transformation) produces a new point in the time course, and every $1000^{th}$ (A), $100^{th}$ (B) and $10^{th}$ (C) time course point is plotted. Rate constants are: $k_0 = 5 \cdot 10^{-3}$ $nM^{-1} \cdot s^{-1}$, $k_{-0} = 0.5$ $s^{-1}$ ($K_d = 100$ nM); $k_{p1} = 0.2$ $s^{-1}$, $k_{-p1} = 0.8$ $s^{-1}$; $k_{p2} = 0.8$ $s^{-1}$, $k_{-p2} = 0.2$ $s^{-1}$. The initial molecular abundances are calculated separately for panels A, B, and C by multiplication of the initial concentrations by the corresponding volume. The initial concentrations for Eq. 3 and 9 were set as follows: $R = 100$ nM, $s(0,0,0) = 100$ nM, $s(0,0,1) = 1 \cdot 10^{-10}$ nM, whereas all other $s(a_1,a_2,h) = 0$; $S_1(0,0) = S_2(0,0) = 100$ nM, $S_1(0,1) = S_2(0,1) = 1 \cdot 10^{-10}$ nM, and all other $S_i(a_i,h) = 0$.

**Table 1. Types of sites and imposed conditions that allow for a macro-model reduction**

| Sites | Influence: | Depend on: | Transformation | Rate constants may depend on: |
|---|---|---|---|---|
| Controlling (*h*-sites) | All types of sites | Other controlling sites | $h_j \to \widetilde{h}_j$ | The states (***h***) of controlling sites |
| Independent docking sites (*a*-sites) | No effect on other sites | Controlling sites | $a_j \to \widetilde{a}_j$ | The states $a_j$ and $\widetilde{a}_j$ and the states (***h***) of controlling sites |
| Interacting allosterically or via bound partners (*b*-sites) | Other allosterically interacting sites only | Controlling and other allosterically interacting sites | $b_j \to \widetilde{b}_j$ | All states (***b***) of allosterically interacting sites and the states (***h***) of controlling sites |