Student Thesis

Systematic model reduction of signal transduction pathways using observability analysis reveals system-inherent structures

cand. kyb Dirk Fey,
University of Stuttgart

Supervisors: Dipl.–Ing. Holger Conzelmann (ISYS)
Prof. Dr.–Ing. Dr.h.c. mult. Ernst D. Gilles

University of Stuttgart
Institute for System Dynamics
Prof. Dr.-Ing. O. Sawodny

October 2006

Acknowledgements

My tanks are to my supervisor Holger Conzelmann, who made this thesis possible, for his kind and uncomplicated guidance and support in all matters.

I also would like to thank Prof. Michael Zeitz for initializing my engagement with the Institute for Systems Dynamics.

Last but not least, my thanks are to my family and friends, whose support was indispensable for the success of my studies and the preparation of this thesis.

Contents

List of Figures

List of Tables

Introduction

Systems Biology is a wholistic approach in understanding Biology as it aims in a systems-theoretical understanding [Chong and Ray 2002]. Rather than to characterize and classify isolated biological compartments, it tries to reveal the underlying principles of cellular and organismic function. Usually, most characteristics of life first emerge at a systems level, making the isolated consideration of the systems compartments insufficient to explain its overall behavior [Kitano 2002b]. To reach this goal it uses mathematical models, computation and methods of systems theory, trying to integrate the rapidly growing amount of biological data[Kitano 2002a].

One aspect of systems biology is quantitative temporal modeling of molecular cellular processes [Sauter and Bullinger 2004; Stelling and Gilles 2004]. In order to understand the dynamics of the considered process, a mathematical model is built. Analyzing the model can not only reveal important structural properties, as for instance the necessary compartments and interactions for the general function, but also predict the outcome of biological experiments.

In this thesis, I will focus on receptor mediated signal transduction and consider biological systems as a network of biochemical reactions, composed of biological functional units like receptors, scaffold proteins and signaling molecules. These biological functional units often can be modified on several binding domains leading to distinguishable 'operation modes' with different biochemical properties. A good example is a cell surface receptor, which propagates an external signal (cell stimulus) into the cells. After binding of an extracellular hormone, it catalyzes one or more biochemical reaction

in the cytoplasm. If the extracellular hormone separates from the receptor, this catalytic activity stops. So the receptor can be in two modes, i.e. 'on' or 'off', depending on the occupancy of its extracellular binding domain. Normally, receptors and other signaling molecules possess several binding domains, each able to assume several states, leading to an exponential increase of operation modes. In addition these molecules are highly interconnected (also through their binding domains). This variety of cellular signaling and its compartments easily yields to unsuitable large models [Csete and Doyle 2002; Hlavacek et al. 2003]. Even for isolated parts, for instance of the EGF-signaling pathway, the modeling of several million species is necessary to account for its complexity.

The first problem with complex models is, that they are difficult to parameterize, due to the lack of quantitative measurements. Mostly, only measurements of a few concentrations, and only qualitative knowledge some reactions is available, as for instance reaction A is considerable faster than reaction B. The quality of biological models can benefit a great deal by using this kind of qualitative knowledge. The second problem with complex models is, that they are difficult to analyze and to understand. A model that is as complex as reality is nearly as difficult to understand as reality itself. Therefore, mathematical methods are necessary, which reduce the models dimension, or modularize the models structure.

The modeling technique presented in this thesis, which we call 'domain oriented modeling', accounts for the biological complexity, and reduces the number of necessary parameters using quantitative knowledge about the domain interactions. In addition, it deals with the complexity of the resulting model, by reducing its dimension and revealing system inherent structures.

Chapter 1

Molecular biological and systems theoretical background

1.1 Signaling in cell biology

Many different kinds of molecules transmit information between the cells of multicellular organisms. Although all these molecules act as ligands that bind to receptors expressed by their target cells, there is considerable variation in the structure and function of the different types of molecules that serve as signal transmitters. Structurally, the signaling molecules used by plants and animals range in complexity from simple gases to proteins. Some of these molecules carry signals over long distances, whereas others act locally to convey information between neighboring cells. In addition, signaling molecules differ in their mode of action on their target cells. Some signaling molecules are able to cross the plasma membrane and bind to intracellular receptors in the cytoplasm or nucleus, whereas most bind to receptors expressed on the target cell surface. [Cooper 2000]

1.1.1 Modes of cell-cell Signaling

Cell signaling can result either from the direct interaction of a cell with its neighbor or from the action of secreted signaling molecules. Signaling by direct cell-cell (or cell-matrix) interactions plays a critical role in regulating the behavior of cells in animal tissues. For example, the integrins and

cadherins function not only as cell adhesion molecules but also as signaling molecules that regulate cell proliferation and survival in response to cell-cell and cell-matrix contacts. In addition, cells express a variety of cell surface receptors that interact with signaling molecules on the surface of neighboring cells. Signaling via such direct cell-cell interactions plays a critical role in regulating the many interactions between different types of cells that take place during embryonic development, as well as in the maintenance of adult tissues.

The multiple varieties of signaling by secreted molecules are frequently divided into three general categories based on the distance over which signals are transmitted.

- In **endocrine signaling**, the signaling molecules (hormones) are secreted by specialized endocrine cells and carried through the circulation to act on target cells at distant body sites. A classic example is provided by the steroid hormone estrogen, which is produced by the ovary and stimulates development and maintenance of the female reproductive system and secondary sex characteristics. In animals, more than 50 different hormones are produced by endocrine glands, including the pituitary, thyroid, parathyroid, pancreas, adrenal glands, and gonads.

In contrast to hormones, some signaling molecules act locally to affect the behavior of nearby cells.

- In **paracrine signaling**, a molecule released by one cell acts on neighboring target cells. An example is provided by the action of neurotransmitters in carrying signals between nerve cells at a synapse.

Finally, some cells respond to signaling molecules that they themselves produce.

- One important example of such **autocrine signaling** is the response of cells of the vertebrate immune system to foreign antigens. Certain types of T lymphocytes respond to antigenic stimulation by synthesizing a

> growth factor that drives their own proliferation, thereby increasing the number of responsive T lymphocytes and amplifying the immune response. It is also noteworthy that abnormal autocrine signaling frequently contributes to the uncontrolled growth of cancer cells. In this situation, a cancer cell produces a growth factor to which it also responds, thereby continuously driving its own unregulated proliferation.

[Alberts et al. 2002; Cooper 2000]

1.1.2 Functions of cell surface receptors

As already reviewed, most ligands responsible for cell-cell signaling (including neurotransmitters, peptide hormones, and growth factors) bind to receptors on the surface of their target cells. Consequently, a major challenge in understanding cell-cell signaling is clarifying the mechanisms by which cell surface receptors transmit the signals initiated by ligand binding. Some neurotransmitter receptors are ligand-gated ion channels that directly control ion flux across the plasma membrane. Other cell surface receptors, including the receptors for peptide hormones and growth factors, act instead by altering the activity of intracellular proteins. These proteins then transmit signals from the receptor to a series of additional intracellular targets, frequently including transcription factors. Ligand binding to a receptor on the surface of the cell initiates a chain of intracellular reactions, finally often reaching the target cell nucleus and resulting in programmed changes in gene expression.

1.1.3 Classification of cell-surface receptors

The different types of cell-surface receptors that interact with water-soluble ligands are schematically represented in Figure 1.1. Binding of ligand to some of these receptors induces second-messenger formation, whereas ligand binding to others does not. For convenience, we can sort these receptors into four classes:

- G protein coupled receptors (see Figure 1.1a): Ligand binding activates a G protein, which in turn activates or inhibits an enzyme that gener-

ates a specific second messenger or modulates an ion channel, causing a change in membrane potential. The receptors for epinephrine, serotonin, and glucagon are examples.

- Ion-channel receptors (see Figure 1.1b): Ligand binding changes the conformation of the receptor so that specific ions flow through it; the resultant ion movements alter the electric potential across the cell membrane. The acetylcholine receptor at the nerve-muscle junction is an example.

- Tyrosine kinase linked receptors (see Figure 1.1c): These receptors lack intrinsic catalytic activity, but ligand binding stimulates formation of a dimeric receptor, which then interacts with and activates one or more cytosolic protein-tyrosine kinases. The receptors for many cytokines, the interferons, and human growth factor are of this type. These tyrosine kinase linked receptors sometimes are referred to as the cytokine-receptor superfamily.

- Receptors with intrinsic enzymatic activity (see Figure 1.1d): Several types of receptors have intrinsic catalytic activity, which is activated by binding of ligand. For instance, some activated receptors catalyze conversion of GTP to cGMP; others act as protein phosphatases, removing phosphate groups from phosphotyrosine residues in substrate proteins, thereby modifying their activity. The receptors for insulin and many growth factors are ligand-triggered protein kinases; in most cases, the ligand binds as a dimer, leading to dimerization of the receptor and activation of its kinase activity. These receptors often referred to as receptor serine/threonine kinases or receptor tyrosine kinases autophosphorylate residues in their own cytosolic domain and also can phosphorylate various substrate proteins.

[Alberts et al. 2002; Lodish et al. 2000]

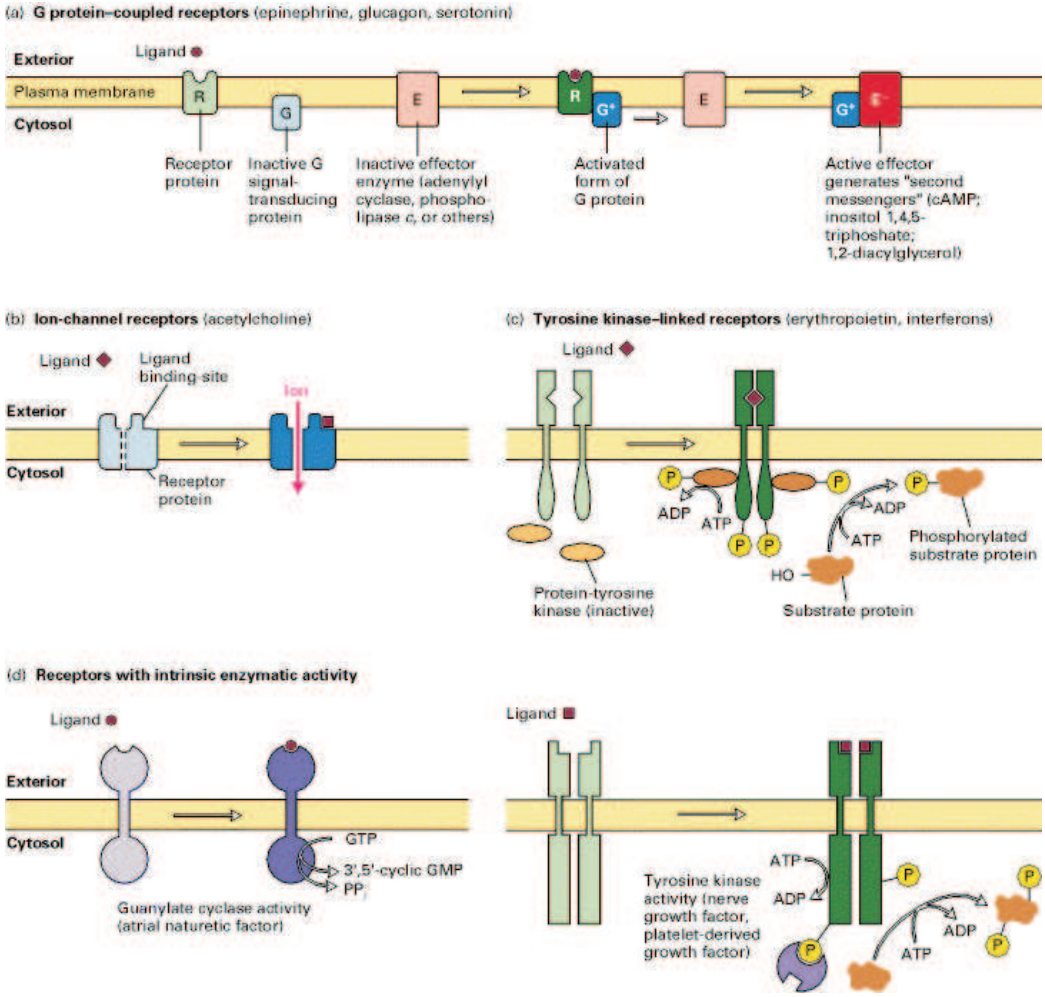

Figure 1.1: Classification of cell-surface (transmembrane) receptors in G protein coupled receptors, ion channels, tyrosine kinase linked receptors and receptors with intrinsic enzymatic activity. Figure taken from [Alberts et al. 2002]

1.1.4 Adapter and scaffold proteins

Many signal-transduction pathways contain large multiprotein signaling complexes, which often are held together by adapter proteins. Adapter proteins do not have catalytic activity, nor do they directly activate effector proteins. Rather, they contain different combinations of domains, which function as docking sites for other proteins. For instance, different domains bind to phosphotyrosine residues (SH2 and PTB domains), proline-rich sequences (SH3 and WW domains), phosphoinositides (PH domains), and unique C-terminal sequences with a C-terminal hydrophobic residue (PDZ domains) (see Figure 1.2). In some cases adapter proteins contain arrays of a single binding domain or different combinations of domains. In addition, these binding domains can be found alone or in various combinations in proteins containing catalytic domains. These combinations provide enormous potential for complex interplay and cross-talk between different signaling pathways. An example is the IRS1 (insulin receptor substrate 1) of the insulin signaling pathway. The ISR1 is able to bind at least three molecules: The insulin receptor, Grb2 (growth factor receptor-bound protein 2) and Shp2 (tyrosine protein phosphatase 2). Such proteins can be referred to as scaffolding proteins due to the assimilation of multiple molecules generating a scaffold. Providing sufficiently high concentrations at the destination, these scaffolds enable the submission of a signal along time and space. [Alberts et al. 2002; Lodish et al. 2000]

1.1.5 Cytoplasmic receptors

Cytoplasmic receptors are soluble proteins localized within the cytoplasm. The hormone has to pass through the plasma membrane, usually by passive diffusion, to reach the receptor and initiate the signal cascade. The cytoplasmic receptors are ligand-activated transcription activators. On binding with the ligand (the hormone), they will pass through the nuclear membrane into the nucleus and enable the transcription and translation of a certain gene and, thus, the production of a protein.

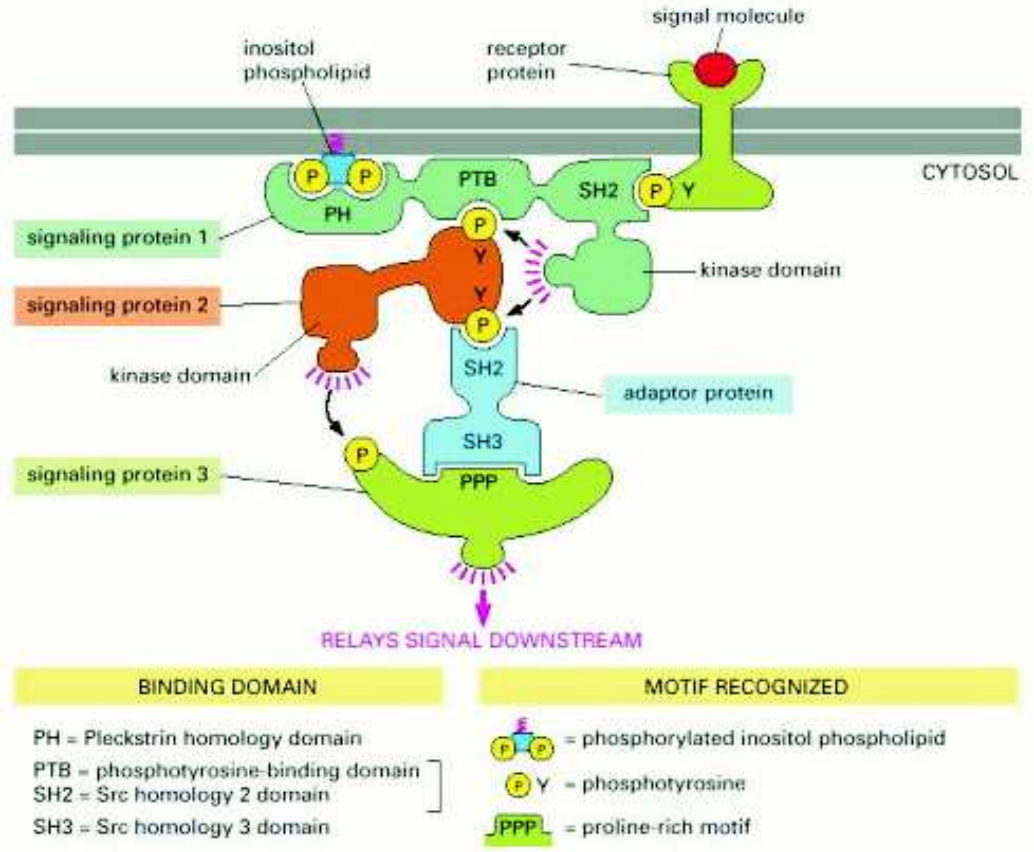

Figure 1.2: Complex formation in signal transduction. Receptors and scaffolds assemble different signaling and adaptor proteins, transmitting the signal through time and space. Figure taken from [Lodish et al. 2000]

1.1.6 Feedback control of hormone levels

The synthesis and/or release of many hormones are regulated by positive or negative feedback. This type of regulation is particularly important in coordinating the action of multiple hormones on various cell types during growth and differentiation. Often, the levels of several hormones are interconnected by feedback circuits, in which changes in the level of one hormone affect the levels of other hormones. One example is the regulation of estrogen and progesterone, steroid hormones that stimulate the growth and differentiation of cells in the endometrium, the tissue lining the interior of the uterus. Changes in the endometrium prepare the organ to receive and nourish an embryo. The levels of both hormones are regulated by complex feedback circuits involving several other hormones. [Lodish et al. 2000]

Furthermore there are often intracellular feedback loops within a signal transduction pathway for several tasks. Negative feedback for example often ensures robustness, whereas positive feedback loops can be used to build up bio-molecular switches or flip-flops. These feedback occurs on several levels and timescales. For example GAP junctions close a negative feedback loop over particular receptor appropriate kinase, in order to inhibit the receptors activity and such regulate the signal strength on reception level. Another possible mechanism occurs on gene expression level. Hereby the involved proteins are altered in concentrations through gene expression, this can be either the receptor or one or more of its adaptor proteins. Examples are the regulators of G-protein coupled receptors (RGS).

1.2 Modeling of signal transduction pathways

The interactions of the cells compartments (functional units like receptors or enzymes, signaling molecules, effectors) are based on several chemical forces and can be described in terms of biochemical reactions. Receptors and most of its adaptor molecules are macro-molecules, largely composed of long amino acid chains. They are fold into a particular three-dimensional structure (often referred to as conformation) which determines its function.

The conformation as well as the assembly of several macro molecules and ligands is stabilized by different types of chemical bounds:

- **Covalent bonds** are characterized by the sharing of one or more pairs of electrons between two elements, producing a mutual attraction that holds the resultant molecule together. Atoms tend to share electrons in such a way that their outer electron shells are filled. They most frequently occur between atoms with similar electronegativities.

- An **ionic bond** can be formed after two or more atoms lose or gain electrons to form an ion. Ionic bonds occur between metals, losing electrons, and non-metals, gaining electrons. Ions with opposite charges will attract one another creating an ionic bond. Such bonds are stronger than hydrogen bonds, but similar in strength to covalent bonds.

- **Van der Waals** interactions refer to intermolecular forces that deal with forces due to the polarization of molecules. Forces that deal with fixed or angle averaged dipoles (Keesom forces) and free or rotation dipoles (Debye forces) as well as shifts in electron cloud distribution (London Forces).

- **Hydrogen bonds** exists between two partial electric charges of opposite polarity. Although stronger than most other intermolecular forces, the typical hydrogen bond is much weaker than both the ionic bond and the covalent bond. Within macromolecules such as proteins and nucleic acids, it can exist between two parts of the same molecule, and figures as an important constraint on such molecules' overall shape.

- **Hydrophobic interactions** are performed by water, which is electrically polarized and able to form hydrogen bonds internally. Non-polarized matter is repelled by water and tends to cluster together since one larger area disturbing the internal structure of water is more energetic favorable than two smaller areas.

- A **disulfide bond**, also called a disulfide bridge, is a strong covalent bond between two sulfhydryl groups (-SH). The presence of disulfide bonds help to maintain the tertiary structure of the protein.

In the above presented processes of ligand binding, conformational change and multi-molecule assembly bounds are broken down and formed new continuously. In consequence all these processes can be seen as chemical reactions, and described in terms of reaction equations. These approach is very common and widely used in modeling molecular biological processes and results in metabolic and signaling pathways completely described by a network of biochemical reactions. On a molecular level, meaning that all involved substances (in Biology often referred to as species) are considered, a mathematical model can derived by applying the law of mass action.

1.2.1 The law of mass action

The law of mass action is the origin of our quantitative treatment of chemical reactions and their kinetics. In a reaction system, the probability, that two molecules come close enough together to interact and perform a chemical bound does highly depend on the temperature, which is a measure of the nondirectional molecule movement, as well as the active masses (concentrations) of the substrates. For reactions occurring in cell-biological systems, the temperature can be assumed constant since the changes are relatively small according to the absolute value measured in Kelvin. Consequently the reactive activity of the molecules does not depend on the temperature[1]. The the law of mass action can be obtained from a kinetic point of view:

. *The rate of a chemical reaction is directly proportional to the active masses (concentrations) of each of the reactants.*

To illustrate this relationship, think on the randomly moving molecules a substrate. If the concentration is doubled, the probability that two molecules

[1]To include the dependency of the temperature, Arrhenius proposed the reaction rate proportional to the term $e^{\frac{E}{RT}}$, where E is the energy, R the gas constant and T the absolute temperature.

come close enough together to react will be double as high. Consequently the reaction rate will be double as high.

The reaction rates describe the flux of substance, which is a measure of how much substrate is converted to product per time. They lead finally to a description of the reaction system in terms of ordinary differential equations. For a general reversible chemical reaction of the form

$$aA + ... + bB \rightleftharpoons cC + ... + dD$$

the reaction rates are

$$\begin{aligned} r_{products} &= k_{products} c_A^a \cdot \ldots \cdot c_B^b \\ r_{substrates} &= k_{substrates} c_C^c \cdot \ldots \cdot c_D^d \end{aligned}$$

The total reaction rate is

$$r_{total} = r_{products} - r_{substrates}$$

To get the change of concentrations in time $\frac{dc_i}{dt}$ the stoichiometric parameters $a, ..., b$ and $c, ..., d$ have to be considered as pre-factors of the reaction rate, which hereby are related to an one molar change. For instance the above reaction needs a mole substrate A to form c mole product C, and thus the ordinary differential equations for the concentration-change of A and C are

$$\frac{d}{dt} c_A = -a\, r_{total}, \quad \frac{d}{dt} c_C = -c\, r_{total}$$

It is common sense, that the reactions are always assumed to direct from the left side (substrates) to the right side (product) of the reaction equation. Consequently the stoichiometric parameters are negative for substrates, and positive for product. If the reaction runs actually vice versa, the net reaction rate becomes negative.

1.2.2 System of reactions

Signal transduction pathways are biological reaction networks, generally possessing an abundant number of different substances and a variety of reactions connecting them. The changes of concentrations are now described by the

sum of all producing and consuming terms, i.e. the sum of all reaction rates multiplied by their stoichiometric coefficient. This finally leads to a system of ordinary differential equations of the form

$$\frac{d\mathbf{c}}{dt} = N\mathbf{r}(\mathbf{c}), \quad \mathbf{c} \in \mathbb{R}^{n,+},\ N \in \mathbb{R}^{n\times m},\ \mathbf{r} \in \mathbb{R}^m$$

where $\mathbf{c}$ denotes the vector of all n species concentrations, N denotes the stoichiometric matrix and $\mathbf{r}(\mathbf{c})$ denotes all m reaction rates depending in the species concentrations. As we will see in the next subsection, a system of reactions can include direct dependencies. As a consequence, the stoichiometric matrix N has no full row rank.

1.2.3 Conserved moieties

Conserved moieties are sums of concentrations which are constant over time. They state linear dependencies within an reaction system, and lead to an inherent model reduction of the ordinary differential equations system. A famous example is the enzyme driven reaction from a substrate S to a product P. The substrate binds to a Enzyme E (catalytic unit) reversibly building an enzyme-substrate complex C which triggers the transformation of bound S to the product P, which is irreversible. The reaction scheme is

$$S + E \rightleftharpoons C \rightharpoonup E + P$$

The reaction rates are

$$\begin{aligned} r_1 &= k_{+1}c_S c_E - k_{-1}c_C \\ r_2 &= k_2 c_C \end{aligned}$$

leading to the system of ordinary differential equations

$$\frac{d}{dt}\begin{pmatrix} c_S \\ c_E \\ c_C \\ c_P \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ -1 & 1 \\ 1 & -1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$$

The rows two and three of the stoichiometric matrix N are linear dependent, and consequently sum of the corresponding equations become zero.

$$\frac{d}{dt}c_E + \frac{d}{dt}c_C = -r_1 + r_2 + r_1 - r_2 = 0 \tag{1.1}$$

Thus the sum c_E+c_C is constant, and denoted as conserved moiety of enzyme. Integration of 1.1 from the starting point $t = 0$ to any time point t gives

$$c_E(t) + c_C(t) = c_E(0) + c_C(0)$$

For convenience we omit the dependence of time in notation and denote the states $c_E(t) = c_E$, $c_C(t) = c_C$ and the initial conditions $c_E(0) = c_{E0}$, $c_C(0) = c_{C0}$ in the following. Solving the equation for c_E and applying to the system results in elimination of c_E. Accordingly only three differential equations must be solved, and the concentration of enzyme is determined by the algebraic condition

$$c_E = c_{E0} + c_{C0} - c_C$$

Generally reaction systems are of the form $\frac{d}{dt}\mathbf{c} = N\mathbf{r}(\mathbf{c})$ the number of necessary differential equation determining the system is given by the row-rank of the stoichiometric matrix N.

$$N = \begin{pmatrix} N_0 \\ N_D \end{pmatrix}, \qquad \text{with } N_0 \text{ has full row-rank}$$

A link matrix L is set up describing the algebraic conditions between the dependent and independent states ($N_D = LN_0$).

$$\frac{d}{dt}\begin{pmatrix} \mathbf{c_{indep}} \\ \mathbf{c_{dep}} \end{pmatrix} = \frac{d}{dt}\begin{pmatrix} N_0 \\ LN_0 \end{pmatrix}\mathbf{r} = \begin{pmatrix} I \\ L \end{pmatrix} N_0\mathbf{r}$$

This separation of independent states ($\mathbf{c}_{indep}$)and linear dependent states ($\mathbf{c}_{dep}$) allows solving of a $dim(\mathbf{c}) - rowrank(N)$ smaller differential system. Remind that the reaction rates are functions of the concentrations $\mathbf{r} = \mathbf{r}(\mathbf{c}_{indep}, \mathbf{c}_{dep})$. The link matrix is used to reduce the dependencies to the independent states, and only the solving of the following differential equations is necessary.

$$\frac{d}{dt}\mathbf{c}_{indep} = N_0\mathbf{r}(\mathbf{c}_{indep}, L\mathbf{c}_{indep})$$

The dependent states are algebraically calculated from the independent ones through the link matrix:

$$\mathbf{c}_{dep} = L\mathbf{c}_{indep}$$

Note that the given initial conditions of a real system $\mathbf{c}(0)$ have to be consistent according to the dependencies expressed through the link matrix L. This means $\mathbf{c}_{dep}(0) = L\mathbf{c}_{indep}(0)$.

1.3 Common model reduction methods

As we have seen in the previous section, biological processes as occur in cell signaling networks are highly complex. In order to understand the underlying principles of cell regulatory systems models have to be set up, which in consequence, become highly complex as well possessing thousands or millions of state variables. In fact, these high dimensional models are as difficult to understand as reality. But neglecting state variables based on intuitively made preconditions does not take into account the inherent complexity of biological reaction systems, and thus often fails in emitting its underlying principles [Faeder et al. 2003; Hlavacek et al. 2003]. It turns out, that on that task, microscopic modeling of all possible states is necessary for a beneficial analysis [Blinov et al. 2004]. On the other hand, the analysis of high dimensional models results often in mathematical ratios or characteristics like sensitivities, not providing a understanding of basic regulatory mechanisms. To aim in the great goal of revealing underlying principles of cell regulation, model reduction is essential [Conzelmann et al. 2004]. Thereby a systematic approach is most desirable. Domain oriented modeling, which is the focus of this thesis, is a method to model signal transduction networks taking their combinatorial complexity into account, and reduce the resulting model systematically. In this section, the most common model reduction methods referring to biological systems will be presented briefly.

There is a variety of model reduction methods aiming in simpler descriptions. Most of them are based on a moderate understanding of the system enabling the modeler to presuppose strong assumptions or just neglect inessential reactions or states. They can be structured into three parts.

- Less species

  - Elimination of inessential species
  - Lumping species

- Less reactions
  - Elimination of inessential reactions
  - Assume equilibrium either for single reactions (**rapid equilibrium**) or for all reaction corresponding to a concentration (**quasi-steady-state**)

- Time-scale based decomposition.

Most of these techniques present approximative methods, where the modeled behavior shall represent the real behavior as good as possible, thereby achieving a certain accuracy necessary for the analysis or predictions. Mostly the reduction is related to the process of modeling itself, for instance elimination of inessential species or reactions is already conducted through defining the boundaries of the system. Further, whilst deriving the mathematical model, species appearing as sinks are normally not balanced (that means no differential equation is set up for this concentration), since the continuous increase of a concentration makes no sense in real biological systems, normally tending to any kind of homoeostasis (or limit circle). In some cases, species showing very similar properties, as for example two subtypes of a receptor, can be lumped together before even hooking up the model.

More formal approaches are rapid-equilibrium and quasi-steady-state assumptions, as they are very common in enzyme kinetics. Here either a reaction or a concentration is formally[2] assumed to be in equilibrium. If a concentration is quasi-steady-static, its derivation can formally be set to zero ($\frac{d}{dt}c_{steady-state} = 0$). This means that the change of this concentration is relatively small compared to the changes of the other concentrations (The time-course of $c_{steady-state}$ is much slower). If a reaction is in rapid equilibrium, its reaction rate can formally[2] be set to zero ($r_{rapid-equlibrium} = 0$). This means the driving force of this reaction in much greater than the driving forces of the surrounding reactions, and thus this reaction occurs much faster

than the other reactions. It is closely related to the time-scale based reduction.

Time-scale based reduction is based on the idea that many systems can be divided into a slow and a fast part, whereby the system's dynamics are approximated by the slow part.

$$\begin{aligned} \frac{d}{dt}\mathbf{x}_{fast} &= f_{fast}(\mathbf{x}_{fast}, \mathbf{x}_{slow}) \\ \frac{d}{dt}\mathbf{x}_{slow} &= f_{slow}(\mathbf{x}_{fast}, \mathbf{x}_{slow}) \end{aligned}$$

The principle is to determine the slow invariant manifold, and to project the systems equations onto it. If the difference of time characteristics is sufficient big, the systems trajectory approaches the slow manifold quickly. Then it can be described by the projection accurately, thereby neglecting the fast manifold. Since the slow manifold is an invariant subset of the state space, it can be expressed by a relation between the state variables $\mathbf{x}_{SM,fast}(\mathbf{x}_{slow})$, leading to the invariance

$$\frac{\partial}{\partial \mathbf{x}_{slow}}\mathbf{x}_{SM,fast} \cdot f_{slow} = f_{fast}$$

Several approaches exists in order to determine the slow manifold.The most common technique is an approximative method using Taylor-series. But there also exists an iterative method from [Roussel and Fraser 2001] and a computational technique based on linearization referred to as intrensic low dimensional manifold (ILDM) [Maas and Pope 1992].

There is another formal approach in lumping species from [Toth et al. 1997], which considers two systems of differential equations of different dimension and a transformation from the larger to the smaller system. The work of [Toth et al. 1997] concerns the change of the properties of the solution. Addressed questions are: Under which conditions equilibria, periodic solutions, invariant sets are lumped to equilibria, periodic solutions, invariant sets, and how does this lumping effect stability properties.

[2]This does not mean, that the concentration is not changing anymore. But it establishes direct relationship between some concentrations (often $c_{steady-state}(c_1, ..., c_n)$)

The reduction method which is focused in this thesis presents a lumping of species as in [Toth et al. 1997], but rather than investigating the theoretical effects of lumping in general, it is focussed on its application and setup of linear transformations that allow model reductions for biological systems. An output of the system is defined, and a smaller system describing the same output behavior is achieved through a transformation based on an observability analysis. In the next section the mathematical theory of nonlinear observability analysis is briefly presented, giving the background for the next chapters.

1.4 Observability analysis

The observability analysis is a common tool in systems theory and control engineering highly related to the task of determination of the inner state of a system through measuring particular outputs (function of a subset of the space variables). According to the special task of this thesis in model reduction, the observability analysis is used to determine the reduced state space. In the following some essential definitions and results of the observability analysis are briefly presented providing the system theoretical background of domain oriented modeling.

1.4.1 Nonlinear observability analysis

The nonlinear system is given as follows

$$\begin{aligned} \mathbf{x} &= \mathbf{f}(\mathbf{x}) \quad t > 0, \quad \mathbf{x}(0) = \mathbf{x_0} \in M \subseteq \mathbb{R}^n \\ y_i &= h_i(\mathbf{x}) \quad t \geq 0 \quad i = 1, \cdots, p \end{aligned} \tag{1.2}$$

Two initial conditions $\mathbf{x}_{10} \neq \mathbf{x}_{20}$ of the system are called indistinguishable ($\mathbf{x}_{10} I \mathbf{x}_{20}$) if the output function $\mathbf{y}(t, \mathbf{x}_{10})$ and $\mathbf{y}(t, \mathbf{x}_{20})$ for the initial conditions $\mathbf{x}_1(\mathbf{0}) = \mathbf{x}_{10}$ and $\mathbf{x}_2(\mathbf{0}) = \mathbf{x}_{20}$ are identical.

Definition 1. *The system* (1.2) *is observable with the output* y_i*, if and only if from* $\mathbf{x}_{10} I \mathbf{x}_{20}$ *follows* $\mathbf{x}_{10} = \mathbf{x}_{20}$*. [Nijmeijer and Van der Schaft 1990]*

The observability analysis is based on a linear function space referred to as observability space.

$$\mathbb{O} = \left\{ h_1(\mathbf{x}), L_f h_1(\mathbf{x}), L_f^2 h_1(\mathbf{x}), \cdots, h_p(\mathbf{x}), L_f h_p(\mathbf{x}), \cdots \right\}$$

It is the space spanned from the output functions $y_i = h_i(x)$ and their derivations in time $\frac{d^j y_i}{dt^j}$. Hereby the time derivations are expressed in terms of Lee derivations [3] $L_f^j h_i(\mathbf{x}) = \frac{d^j y_i}{dt^j}$ of the function h_i along a vector field $\mathbf{f}(\mathbf{x})$.

In an analytical system the outputs y_i can be expressed in terms of Taylor-expansions as linear combinations of the Lee derivations $L_f^j h_i(\mathbf{x}) \in \mathbb{O}$. Calculating the Taylor-series for two different output functions $y_i(t, \mathbf{x}_{01}), y_i(t, \mathbf{x}_{02})$ (outputs with different initial conditions), and plugging into the Definition 1 of observability gives:

$$\begin{aligned} y_i(t, \mathbf{x}_{01}) - y_i(t, \mathbf{x}_{02}) &= \sum_{j=0}^{\infty} \left[\frac{t^j}{j!} (L_f^j h_i(\mathbf{x}_{01}) - L_f^j h_i(\mathbf{x}_{02})) \right] \\ &= \begin{cases} = 0, & \mathbf{x}_{01} = \mathbf{x}_{02} \\ \neq 0, & \mathbf{x}_{01} \neq \mathbf{x}_{02} \end{cases} \end{aligned}$$

Consequently the systems is observable if and only if for each two different states $\mathbf{x}_{01}, \mathbf{x}_{02} \in M,\ \mathbf{x}_{01} \neq \mathbf{x}_{02}$ there exists a function

$$\Phi(\mathbf{x}) = \sum_{j=0}^{\infty} a_j(t) L_f^j h_i(\mathbf{x}), \quad \Phi(\mathbf{x}) \in \mathbb{O}$$

in the observability space which distinguishes the two states, i.e.

$$\Phi(\mathbf{x}_{01}) \neq \Phi(\mathbf{x}_{02})$$

On that account, all components of the state vector $\mathbf{x}$ can be determined by use of the linear independent functions $\Phi(\mathbf{x}) \in \mathbb{O}$. Here the observability space contains an infinite number of functions, a linear independent set of n functions have to be chosen. The easiest way to begin with is to take directly

[3] For a given System (1.2) the Lee derivation is defined recursively as follows

$$L_f^{j+1} h_i(\mathbf{x}) = L_f^j h_i(\mathbf{x}) \,, \quad L_f^1 h_i(\mathbf{x}) = \frac{dh_i(\mathbf{x})}{d\mathbf{x}} \mathbf{f}(\mathbf{x})$$

the output functions and to continue with the Lee derivations. If an linear dependent function $L_f^j h_i(\mathbf{x})$ is found, the subsequent Lee derivations may not be considered, since they are linear dependent as well.

However there is no upper limit of taking higher order Lee derivations (instead of the above mentioned) and thus, it is possible that the state $\mathbf{x}$ of the system is not well-determined by the chosen set of functions in $\mathbb{O}$ (i.e. the equations can not be solved for $\mathbf{x}$ in a unique manner). In that case it must be searched for higher order Lee derivations, which solves for $\mathbf{x}$. If such functions are found, the system is still observable, but not by an uninterrupted chain of Lee derivations.

The observability map $\mathbf{q}(\mathbf{x})$ is defined to calculate the systems state from the observed outputs, therefore it contains n equations (a selection of the outputs and its derivations).

Definition 2. *The observability map is*

$$\mathbf{q}(\mathbf{x}) = \begin{pmatrix} \mathbf{q}_1(\mathbf{x}) \\ \mathbf{q}_2(\mathbf{x}) \\ \vdots \\ \mathbf{q}_p(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_p \end{pmatrix} = \mathbf{y}$$

with

$$\mathbf{q}_i(\mathbf{x}) = \begin{pmatrix} L_f^0 h(\mathbf{x}) \\ L_f^1 h(\mathbf{x}) \\ \vdots \\ L_f^{n_i-1} h(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} y_i \\ \frac{d}{dt} y_i \\ \vdots \\ \frac{d_i^n - 1}{dt_i^n - 1} y_i \end{pmatrix} = \mathbf{y}_i, \quad \sum_{i=1}^{p} n_i = n$$

The maximal order of the derivations $n_i - 1$ for each output y_i is not specified, and accordingly there may exist several maps which are distinct in the order of the particular output maps n_i. However the total number of all equations $\sum_{i=1}^{p} n_i$ must equal the dimension of the system n.

To furnish proof of the invertibility of the observability map is very difficult in the most cases. On that account it is often performed a local observability analysis where a small neighborhood $\mathbb{U} = \mathbf{x} \in M : |\mathbf{x} - \mathbf{x}_p| \leq \varepsilon$ of a point of interest $\mathbf{x}_p \in M$ is considered.

Theorem 1. *A System is locally observable at a point* $\mathbf{x}_p \in M$*, if a neighborhood* $\mathbb{U}(\mathbf{x}_p) = \mathbf{x} \in M : |\mathbf{x} - \mathbf{x}_p| \leq \varepsilon$ *of the state* $\mathbf{x}_p$ *exists, so that the observability map* $\mathbf{q}(\mathbf{x})$ *is invertible in* $\mathbb{U}(\mathbf{x}_p)$*.*

The local observability can be checked through the observability rank condition, which can be derived directly from the inverse function theorem. The inverse function theorem gives sufficient conditions for a vector-valued function to be invertible on an open region containing a point in its domain. The theorem states, that if a function $\mathbf{f} : \mathbb{R}^n \mapsto \mathbb{R}^n$ is continuously differentiable near a point $\mathbf{x}_p$ and has at $\mathbf{x}_p$ a Jacobian matrix with a determinant that is nonzero, then f is an invertible function near $\mathbf{x}_p$. That is, an inverse function to $\mathbf{f}$ exists in some neighborhood of $\mathbf{f}(\mathbf{x}_p)$. The Jacobian matrix of the inverse function $\mathbf{f}^{-1}$ at $\mathbf{f}(\mathbf{x}_p)$ is then the inverse of the Jacobian matrix $\frac{\partial}{\partial \mathbf{x}}\mathbf{f}$, evaluated at $\mathbf{x}_p$.

The inverse function theorem can be generalized to differentiable maps between differentiable manifolds. In this context, the theorem states that for a differentiable map $\mathbf{f} : M \mapsto N$, if the derivative of $\mathbf{f}$, $(Df)_p : T_pM \mapsto T_{\mathbf{f}(p)}N$ is a linear isomorphism at a point p in M, then there exists an open neighborhood U of p such that $\mathbf{f}|U : U \mapsto \mathbf{f}(U)$ is a diffeomorphism. Note that this implies that M and N must have the same dimension. If the derivative of $\mathbf{f}$ is an isomorphism at all points p in M then the map $\mathbf{f}$ is a local diffeomorphism.

For the sense of this thesis it is adequate to use the definitions of the inverse function theorem in terms of linear algebra. Then a system is locally observable at a point $\mathbf{x}_p$, if the Jacobian matrix of the observability map, referred to as observability matrix $Q(\mathbf{x})$, has full rank at $\mathbf{x}_p$. This means that the space spanned by the observability matrix (which is a co-distribution of the observability space) has full dimension (the same dimension as the observability space).

Definition 3. *The observability matrix is defined as the Jacobian matrix of the observability map.*

$$Q(\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}}\mathbf{q}(\mathbf{x})$$

Theorem 2. *A system is locally observable at x_p if*

$$rank\{Q(\mathbf{x}_p)\} = n, \quad \mathbf{x}_p \in M \subseteq \mathbb{R}^n$$

As already mentioned this conditions of checking the observability matrix is only sufficient. That means if the observability matrix Q has full rank, the system is locally observable, but the reverse does not hold. The system can still be observable, even though Q has not full rank.

1.4.2 Observability canonical form

For locally observable systems, there may exist local canonical coordinates in which the system takes a specific form [Isidori 1995]. The observability canonical form is then

$$\begin{aligned}
\dot{x}_{i,1} &= x_{i,2} \\
&\vdots \\
\dot{x}_{i,n_i-1} &= x_{i,n_i} \\
\dot{x}_{i,n_i} &= \varphi_i(x_{i,1}, \cdots, x_{i,n_i}) \\
y_i &= x_{i,1} \qquad\qquad i = 1, \cdots, p.
\end{aligned}$$

All nonlinearities are condensed in the functions φ_i. The rest of the system consists of a linear chain of integrators. In fact, the above form exists, if Theorem 1 holds. This means, that the system is locally observable with a subsequent, uninterrupted chain of Lee derivations. The state space transformation bringing such a system to the observability canonical form is given by the observability map of Definition 2. The condition of Theorem 2 guarantees that the observability map $\mathbf{q}(\mathbf{x})$ is a local diffeomorphism, and thus can be used as a local coordinate transformation (Liapunov transformation).

1.4.3 Linear observability analysis

Linear systems can be treated as a special form of nonlinear systems. All above definitions and statements still hold. However, in the case of linear

system the observability analysis can be performed much easier. Consider a linear system of the form

$$\begin{aligned} \dot{\mathbf{x}} &= A \cdot \mathbf{x}, \quad t > 0,\ \mathbf{x} \in R^n \\ y &= h \cdot \mathbf{x} \end{aligned} \tag{1.3}$$

with system matrix is $A \in R^{n \times n}$ and the output vector $h \in R^{1 \times n}$. The observability map is linear as well, and is defined by the observability matrix.

$$q(\mathbf{x}) = \begin{pmatrix} h \cdot \mathbf{x} \\ h \cdot A \cdot \mathbf{x} \\ \vdots \\ h \cdot A^{n-1} \cdot \mathbf{x} \end{pmatrix} = Q \cdot \mathbf{x} \quad with\ Q = \begin{pmatrix} h \\ h \cdot A \\ \vdots \\ h \cdot A^{n-1} \end{pmatrix}$$

Context

In Chapter 3, the observability analysis is used to determine the necessary number of differential equations that describe the output behavior of the system. If the dimension of the observability space is less than the number of states, a linear state space transformation is searched, which lumps necessary states onto the observability space. In such transformed coordinates, the output behavior can be described completely with a lower dimensional differential equations system within the observability space. The states, which do not influence the systems output behavior present the internal dynamics and can be neglected for the sake of model reduction (if the internal dynamics are stable, which is true for all systems considered here), since by definition the observability space is decoupled from the 'internal space'.

Chapter 2

Review: Domain-oriented approach

Proteins involved in cell signaling possess several functional subunits, referred to as domains, performing binding interactions. It is a general concept of classification and structure analysis in molecular biology, to dissect proteins into several subunits of preserved (i.e. in several proteins similar) regions. These so called domains achieve the tasks of the protein, and display a variety of different functions. An example is the transmembrane domain of a cell surface receptor. The domains can be seen as functional subunits fulfilling their function autonomously. However, there may be dependencies between different domains of a protein, as for example in an allosteric regulated enzyme, where activator or inhibitor binding to the regulatory domain changes the affinity of the substrate domain, and such alters the catalytic activity. Since all the domains considered in this thesis correspond to binding processes, they can be referred to as binding or docking sites. Signal transduction is performed by signal and effector binding to the receptor or scaffold, and its functionality is determined by the dependencies between the different binding sites. For instance, a transmembrane receptor can bind a signal molecule on its extracellular domain, therewith causing conformational changes of the receptor's three dimensional structure altering the affinity of one or more intracellular domains. Since such a binding site alters the affinities of other binding domains, it can be referred to it as controlling domain. In this chap-

ter the available literature on the domain oriented approach, which solely refers to the modeling of signal transduction, and its complexity is briefly reviewed to set the work of this thesis (especially the new defined occurrence levels and the results in Chapter 4) into context.

2.1 Domain interactions

In cellular organization the molecular-biological functionality relies on protein-protein interactions, providing an underlying framework, through which signaling pathways are assembled and controlled [Pawson and Nash 2003; Pawson 2003, 2004]. These protein-protein interaction occur on specific functional subunits, the domains, providing a modularized structure on signaling events on the bio-molecular level. Domains present a specific amino-acid sequence within a protein, which relates to a specific function, as for instance membrane crossing domains (hydrophobic) or Shc2 ant PTB domains (phosphorylation sites). Signaling events, which (on a molecular level) present protein assemblies or modifications, rely on specific interacting domains which are for instance able to phosphorylate (activate) or to bind other proteins or signal-molecules. These domains can be referred to as binding domains or docking sites.

In proteomics (study of protein structure and function) a protein is regarded as the together-chaining of its domains possessing a certain three dimensional folding, and functional analysis considers the interactions between these domains. Domains with similar functions are often highly conserved (very similar in amino-acid sequence and genetic code), a fact that has been used in bio-informatics to accelerate the identification of binding partners. Modifications (mutations) of a protein remote a binding domain does not influence its performance, as long as the folded three dimensional structure of the domain is not disturbed. Further, the domains can be seen as separate interacting modules, which can recognize numerous protein modifications (including phosphorylation, methylation, acetylation hydroxylation and ubiquitination) and, in this sense, control the dynamic state of the cell.

Surely the interacting domains within a protein must not be fully independent from each other. A transmembrane receptor involved in tyrosine signaling for instance, recognizes its extracellular signal-ligand through binding it to the extracellular domain. This leads to changes in the receptors three dimensional conformation allowing the phosphorylation of several intracellular Shc2 domains. Such a domain triggering subsequent processes within a protein are referred to as controlling domain.

There is a variety of binding domains involved in cell signaling processes. In tyrosine signaling Shc2 and PTB domains play an important role, since their tyrosine residues can be phosphorylated and so enable the subsequent binding of other signaling molecules.

2.2 Signal transduction and combinatorial complexity

As we have seen in Section 1.1.4, receptors and scaffold proteins often display multiple binding domains which engage several downstream signal proteins. Each of these docking sites can bind various combinations of adapter proteins. This results in a large number of multi-protein-complexes, or species, within the cell. Well studied examples of such complex formations are EGFR [Jorissen et al. 2003; Schlessinger 2000], FceRI [Kinet 1999; Turner and Kinet 1999], Ste5p [Elion 2001] and FcgRIIB [March and Ravichandran 2002]. The number of possible species increases exponentially with the number of binding domains, since each binding domain can be in several different states. An entire collection of potential molecular species corresponding to different forms of an receptor and/or scaffold, is referred to as a set of micro-states. A systematic, mechanistic description has to account for all possible micro-states and all transitions between them.

The combinatorial complexity has been largely ignored by both experimentalists and modelers [Blinov et al. 2004], and is a major barrier to predictive understanding of signal transduction [Faeder et al. 2003; Hlavacek et al. 2003]. Experimental resolution of protein states and complexes is usually

limited to a small number of sites and interactions, but rapidly advancing proteomic technologies are likely to provide a wealth of more detailed information about signaling complexes in the near future [Blagoev et al. 2004; Aebersold and Mann 2003; Meyer and Teruel 2003; Salomon et al. 2003; Mann and Jensen 2003]. A number of studies already confirm that a diverse range of molecular complexes arise during signal transduction [Blagoev et al. 2004; Pacini et al. 2000; Bunnell et al. 2002]. Because the full spectrum of protein states and complexes is difficult to enumerate, computational modeling will play an important role in interpreting such data and assessing the functional significance of specific interactions and complexes [Hlavacek et al. 2003]. In reference to [Faeder et al. 2003] the key questions to be addressed include, whether networks favor the formation of specific complexes from the multitude of potential complexes, and, if so, how these favored complexes affect signaling outcomes. But if regarded from a systems theoretical point of view, it might be more interesting to know which information is necessary to process signaling and how it is used to perform regulatory action. To address these questions is the purpose of this thesis. As will be shown in Chapter 4.1 and 4.2, the domain oriented approach leads to information flow diagrams, as widely used in control engineering, and uncovers underlying principles of control and regulation in the cell.

Few biochemical network models of signaling developed so far cover the breadth of states and complexes required to address these questions of the cell's control strategy. Instead, most models given a particular set of proteins and interactions, make additional (usually implicit) assumptions excluding the vast majority of possible species from consideration. As will be seen in the next example, the EGF-receptor model published in [Kholodenko et al. 1999] and its extensions in [Schoeberl et al. 2002] do not account the full combinatorial complexity.

Example: EGF-receptor

Growth factors and related molecules control cell growth and trigger cell proliferation and other cellular responses through interaction with cell-surface

receptors. A well-studied receptor of this type is EGFR [Jorissen et al. 2003; Schlessinger 2000; Wiley et al. 2003], the receptor for the epidermal growth factor (EGF).

In Figure 2.1 the EGF signaling pathway is illustrated. It shows the EGFR associated with several signaling and adaptor proteins. The processes leading to this receptor-complex can roughly be explained as follows: Ligand binding of EGF stabilizes interaction between receptors [Ferguson et al. 2003; Garrett et al. 2002; Ogiso et al. 2002] leading to a dimerization of two receptors. This co-location enables the cytoplasmic kinase domain of one EGFR to transphosphorylate various cytoplasmic receptor tyrosine residues of the other EGFR [Jorissen et al. 2003; Schlessinger 2000], which means an activation of these domains to a higher energy level. Phosphotyrosine-containing sites can be occupied by the cytosolic adapter proteins Grb2 and Shc [Batzer et al. 1994; Okabayashi et al. 1994] which are attracted by phosphorylated receptors. If Shc is bound to a receptor, it can be phosphorylated by EGFR [Pelicci and Lanfrancone L 1992]. The phosphorylated form of Shc interacts with Grb2 [Rozakis-Adcock et al. 1993] which, in turn, interacts constitutively with the guanine nucleotide exchange factor Sos [Egan et al. 1993; Li et al. 1993; Rozakis-Adcock et al. 1993]. The possibilities of effector aggregation and complex formation is given by the interacting domains [Pawson 2003, 2004]. For example, Grb2 binds EGFR and Shc via its Src homology 2 (SH2) domain [Lowenstein et al. 1992; Rozakis-Adcock et al. 1993], and Grb2 binds Sos via its two SH3 domains [Egan et al. 1993; Li et al. 1993; Rozakis-Adcock et al. 1993]. Summarizing, all this processes lead, in terms of possible molecule complexes, to a fully occupied receptor as sketched in Figure 2.2.

A mathematical model of early events in EGF-EGFR mediated intracellular signaling is published in [Kholodenko et al. 1999] and was extended including downstreaming signaling in [Schoeberl et al. 2002]. However, several implicit assumptions have been made: In the model of [Kholodenko et al. 1999], the only monomers of EGFR considered are those lacking cytoplasmic modifications. The only dimers of EGFR considered are those in which

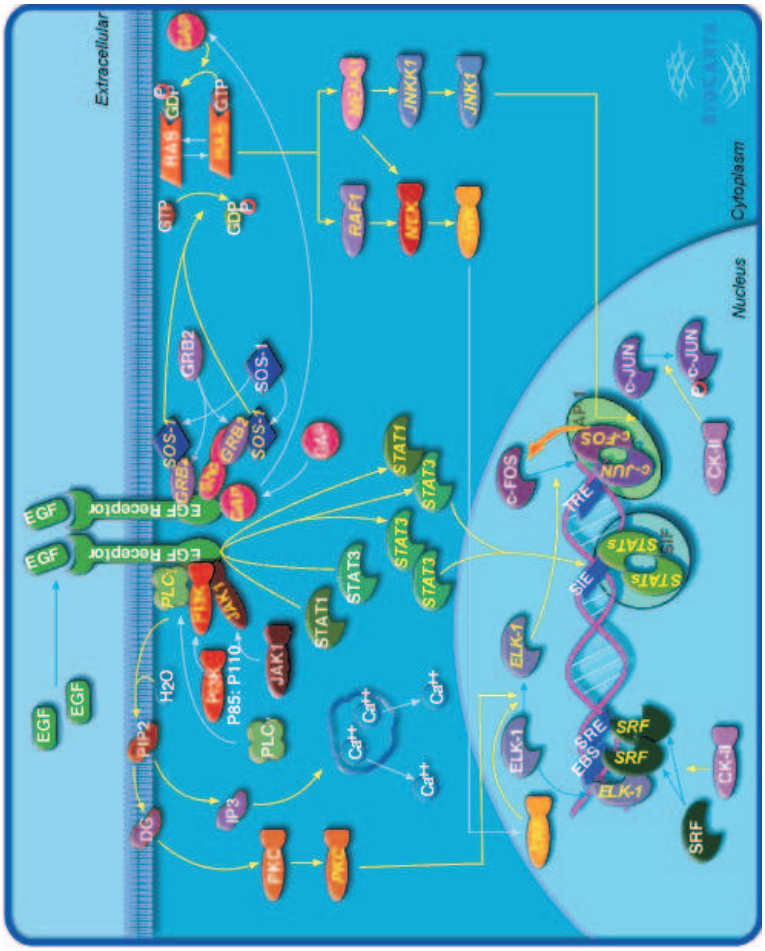

Figure 2.1: Schematically representation of the EGF receptor signal transduction pathway. It is able to dimerize and possesses distinct binding domains which can engage several intracellular signaling molecules [www.biocarta.com].

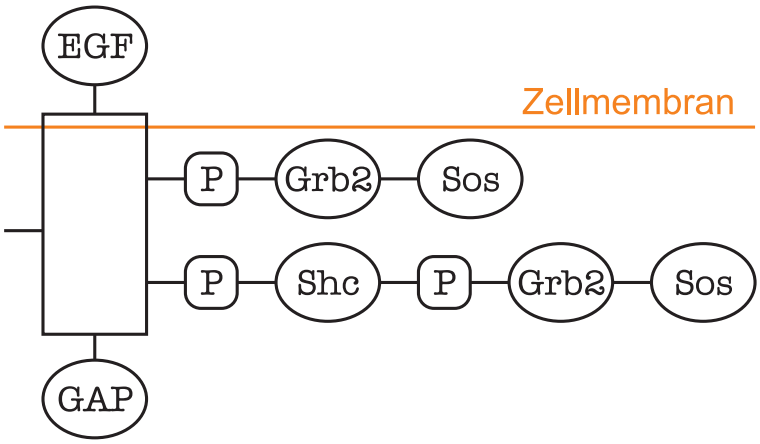

Figure 2.2: Schematic representation of the EGF receptor monomer occupied by its adaptor proteins. The dimer-receptor shows a symmetric structure, where the second receptor monomer is bound to the empty domain on the left.

both receptor-parts have bound EGF, and only one receptor half is in direct contact with, at most, a single adapter protein, either Grb2 or Shc, but not both. Unaggregated receptor monomers with modified/bound cytoplasmic domains, dimers of EGFR involving one or two EGF-free receptors [Jorissen et al. 2003], and dimers of EGFR in direct contact with more than a single adapter protein [Jiang and Sorkin 2002], for example, are assumed not to form. If we wish to account for all possible species, the number of necessary states can be determined by (1) considering all possible states of each domain and (2) taking all combinatorial combinations of these.

For example, we can identify 1232 potential chemical species based on the following assumptions about the possible states of the relevant protein domains. The extracellular domain of a receptor can be either free or bound to EGF. The Grb2 binding site on EGFR can be (1) unphosphorylated, (2) phosphorylated, (3) bound to Grb2, or (4) bound to Grb2 associated with Sos. The Shc binding site on EGFR can be (1) unphosphorylated, (2) phosphorylated, (3) bound to Shc, (4) bound to phosphorylated Shc, (5) bound to Shc associated with Grb2, or (6) bound to Shc associated with Grb2 and Sos in complex. Thus, from combinatorics, there are $2 \cdot 4 \cdot 6 = 48$

species containing a single receptor, an equal number of species containing a symmetric dimer of EGFR, and $\binom{48}{2} = 1128$ species containing an asymmetric dimer of EGFR. In addition to these receptor-containing species, there are seven cytosolic, chemical species and the free extracellular EGF concentration.

2.3 Domain-oriented modeling

In this section the state of the art in domain oriented model reduction, which considers receptor and scaffold related signal transduction using macroscopic states, is presented briefly by reviewing the related articles [Borisov et al. 2005] and [Conzelmann et al. 2006].

[Borisov et al. 2005] demonstrated that in some cases signaling events corresponding to distinct docking sites of a receptor or scaffold molecule can be considered separately. Therefore the assumption of independent binding domains and the use of so called macroscopic states, that correspond to experimentally verifiable variables, is necessary. Then, the macro-states follow the states of independent docking domain separately, including subsequent downstream signal transduction, whereby the other docking domains are unimportant. Compared to the combinatorial explosion of micro-states and equations in a mechanistic model, for a macro-description, the number of macro-states increases linearly, as the sum of distinct domains and binding partners. Provided there is a set of docking sites where molecular events are independent (i.e, allosteric interactions are absent), the signaling system can be modeled macroscopically, and the temporal dynamics of any given micro-state can be expressed explicitly or approximated using the product of the relative concentrations of individual docking sites (see Table 2.3). The basic system considered was a receptor or scaffold possessing one controlling docking site h, influencing the binding properties of the other binding sites dependent on its occupation by a signal-ligand (see Figure 2.3). The original mathematical description of the considered systems as well as the derivation of the statements presented in [Borisov et al. 2005] are not presented in this

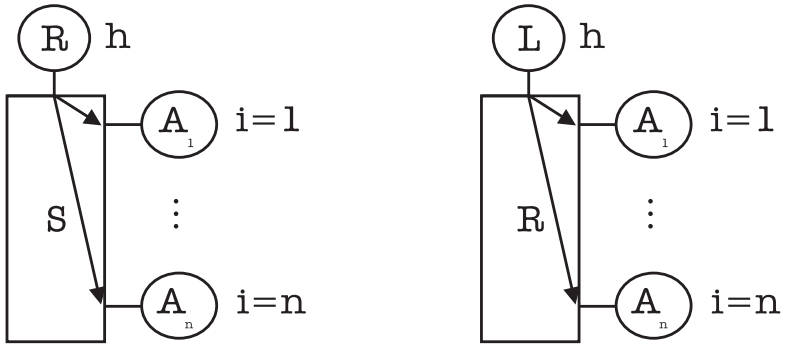

Figure 2.3: Schematic representation of a scaffold (left) and a receptor (right) as considered in [Borisov et al. 2005] with one controlling site h which can either be occupied by the receptor R or the ligand L respectively and several distinct and independent downstreaming sites i which can be unphosphorylated, phosphorylated or occupied by one of the downstreaming signaling molecules A_i, B_i, ... Binding of R and L respectively on the controlling domain h changes the affinities of the the other docking sites (indicated by the arrows).

thesis, since the here used approach in terms of linear algebra contains the mathematical description of these systems. Further this would make notations necessary differing from these used in Chapter 3. However the methods and the results of this work are related to former work and discussed in relevance whenever possible. To facilitate these reference, the systems for which the macroscopic variables were originally defined [Borisov et al. 2005] are summarized in Table 2.3.

System properties	Definition of macro-states	Is the macro-state description exact?	Estimation of micro-states in terms of macro-states	
			During the time-evolution of the system	At the steady-states
A scaffold with kinetically independent docking sites. No controlling hierarchy, no allosteric interactions or interactions through bound partners.	A macro-state is the sum of all micro-states having a particular state of a docking site	Yes	Exact	Exact
A scaffold-receptor module. Docking sites on the scaffold do not interact allosterically or through bound partners. The state of a controlling site (h) on the scaffold influences the chemical transformations of docking sites.	A macro-state is the sum of all micro-states having a particular state of a docking site and a certain state h. (called mesoscopic states in [Conzelmann et al. 2006])	Yes	Approximate. The accuracy of the approximation is higher if the changes in h-states occur much faster than the transformations of docking sites.	Approximate. The accuracy of the approx. is higher if the changes in h-states occur much faster or much slower than the transf. of docking sites
A receptor acting as a scaffold with independent docking sites. Ligand-receptor interactions are independent of the states of receptor docking sites. The state of a controlling site (h) is determined by ligand binding and affects the docking sites.	A macro-state is the sum of all micro-states having a particular state of a docking site and a certain state h. (called mesoscopic states in [Conzelmann et al. 2006])	Yes	Approximate. The accuracy of the approximation is higher if the changes in h-states occur much faster than the transformations of docking sites.	Approximate. The accuracy of the approximation is higher if the changes in h-states occur much faster or much slower than the transformations of docking sites
A receptor that acts as a scaffold with independent sites. No controlling hierarchy (no h-sites), but ligand dissociation occurs only if all docking sites on the receptor are unphosphorylated.	A macro-state is the sum of all micro-states of the receptor with bound ligand and a particular state of a docking site.	No. Only an approximate macrodescription.	Approximate, usually with high accuracy.	Exact
Each site influences any other site, either directly or indirectly (via a bound partner).	Macro-description is not applicable.			

Table 2.1: Macro-description of receptors and scaffolds

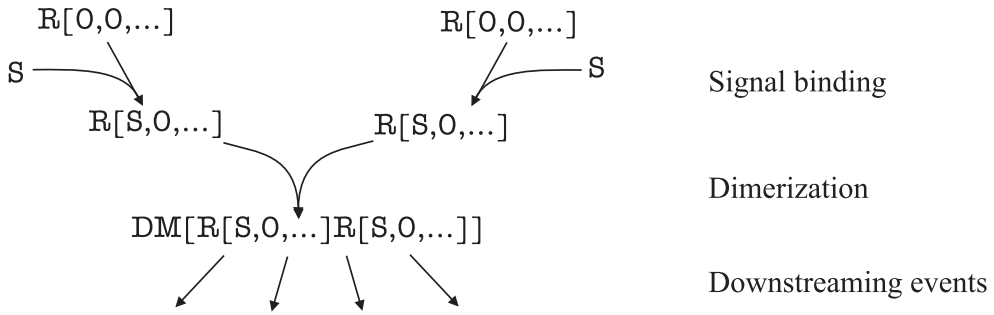

Figure 2.4: Dimerization as considered in the supplementary material of [Borisov et al. 2005]. Both monomeric dimerization partners R[...] have to bind their signal ligand first, before they can form a dimeric receptor-complex DM[...]. All downstreaming signaling events, namely phosphorylation and effector assembly on intracellular domains, occur after the dimerization on the fully signal-occupied receptor-dimer DM[R[S,...],R[S,...]], which in terms of macroscopic modeling can be treated as a single receptor with doubled binding sites.

In their work [Borisov et al. 2005] demonstrated that for scaffold proteins with independent binding sites and scaffolds with one controlling domain, the dynamics of these macroscopic states can be accurately described by reduced models. However, a methodology to derive the reduced model equations for any scaffold with a more complex pattern of domain interactions is missing in this work.

[Conzelmann et al. 2006] introduced a new systematic approach formalizing and extending the above presented model reduction, and structured it in three essential steps. First we start generating a complete mechanistic description of the considered receptor or scaffold protein as in [Faeder et al. 2003; Hlavacek et al. 2003; Blinov et al. 2004]. The resulting microscopic system is of the form:

$$\frac{d}{dt}\mathbf{x} = f(\mathbf{x}), \quad y = h_1(\mathbf{x}) \tag{2.1}$$

All possible binding events (including phosphorylation) have to be accounted and expressed in terms of biochemical reactions by mass action kinetics. Postulated qualitative dependencies between different binding processes can be

taken into account through assumptions about the association and dissociation constants (see Chapter 2.1 on page 41 or Example 8 on page 64).

The second step is the introduction of macroscopic [Borisov et al. 2005] and mesoscopic [Conzelmann et al. 2006] states (levels of occurrence of molecule complexes consisting of one or more species) following a hierarchical pattern and providing a smooth and bijective state space transformation:

$$\mathbf{z} = T\mathbf{x} \quad \Rightarrow \quad \frac{d}{dt}\mathbf{z} = Tf(\mathbf{x})\Big|_{\mathbf{x}=T^{-1}\mathbf{z}}, \quad y = h_1(\mathbf{x})\Big|_{\mathbf{x}=T^{-1}\mathbf{z}} \tag{2.2}$$

The transformed system possesses a modular and partwise decoupled structure, dissecting the ODEs into two subsystems g_1 and g_2.

$$\frac{d}{dt}\begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} = \begin{bmatrix} g_1(\mathbf{z}_1) \\ g_2(\mathbf{z}_1, \mathbf{z}_2) \end{bmatrix}, \quad y = h_2(\mathbf{z}_1) \tag{2.3}$$

The model reduction is in the third step achieved by neglecting the variables $\mathbf{z}_2$ (if stable) and solving $\mathbf{z}_1 = g_1(\mathbf{z}_1)$ for itself. Therefore the dynamics of $\mathbf{z}_2$ must be stable, which is however always the case for biological systems as they are considered here. The number of necessary equations g_1 depends on a given set of concentrations or sums of concentrations of interest y. To achieve a dissection, the output function h_2 as well the equations g_1 must be describable as function of $\mathbf{z}_1$.

This 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. It links the microscopic approach in [Blinov et al. 2004] and the macroscopic approach in [Borisov et al. 2005] by applying a linear state space transformation (Equation 2.2).

Each possible pattern of domain interactions can be realized in the modeling step by adjusting the parameters. The state space transformation which has to be performed is completely independent of this interaction pattern. The 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. The method is able to reveal system inherent structures, leading to a modularization of the system of ordinary

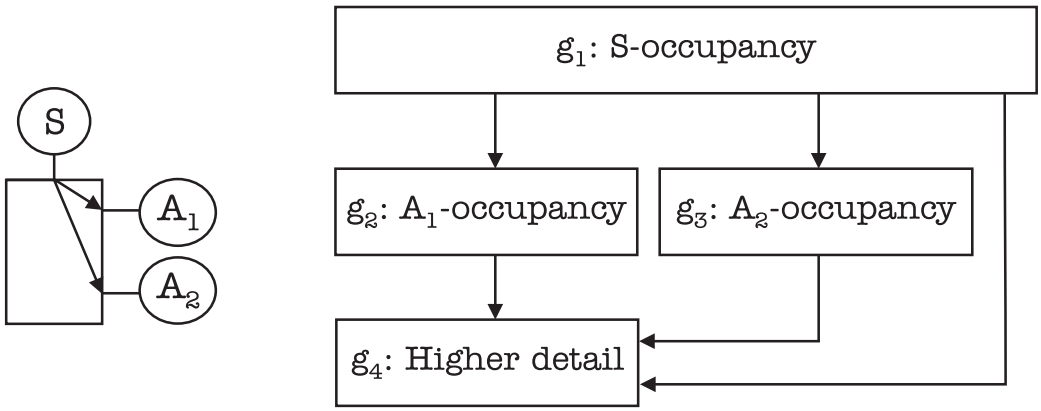

Figure 2.5: Modularization of a receptor or scaffold system with one controlling domain and two effector domains. Description in the main text.

differential equations in terms of partwise decoupled subsystems, each describing a distinct biological process.

Example 1. *The transformed system, of a receptor or scaffold possessing one controlling domain as shown in Figure 2.5, can be written in the form:*

$$\frac{d}{dt}\begin{bmatrix} z_0 \\ z_1 \end{bmatrix} = g_1(z_0, z_1) \tag{2.4}$$

$$\frac{d}{dt}\begin{bmatrix} z_2 \\ z_4 \end{bmatrix} = g_2(z_0, z_1, z_2, z_4) \tag{2.5}$$

$$\frac{d}{dt}\begin{bmatrix} z_3 \\ z_5 \end{bmatrix} = g_3(z_0, z_1, z_3, z_5) \tag{2.6}$$

$$\frac{d}{dt}\begin{bmatrix} z_7 \\ z_8 \end{bmatrix} = g_4(z_0, z_1, z_3, z_4, z_5, z_6, z_7, z_8) \tag{2.7}$$

The transformed system is structured in four modules with a hierarchical pattern of dependency (see Figure 2.5). The first module g_1 (2.4) *,containing two states, describes the signal occupancy* S*. The second and third module* g_2 (2.5) *and* g_3 (2.6) *contain two additional equations each, describing the effector binding of* A_1 *and* A_2 *respectively. Note that they depend on the states of the first module. The last and fourth module* g_1 (2.7) *contains another two states, and is necessary for a more detailed (no information loss) description of the system, needing all states.*

In [Conzelmann et al. 2006] different dependency patterns have been analyzed. However, if several signaling molecules are subsequently engaged on one binding site forming a chain-like multiprotein-complex, the macroscopic and mesoscopic states as defined in [Borisov et al. 2005] and [Conzelmann et al. 2006] do not result in any modularization or model reduction. It was a aim of this thesis, to use the above described method, to search for specific macroscopic and mesoscopic states resulting in some sort of modularization of this process of chain-formation (Chapter 4.1).

Although [Borisov et al. 2005] discussed the possibility of receptor dimerization, a detailed model of the receptor dimerization process in accordance to the domain oriented microscopic approach as in [Blinov et al. 2004], was not yet analyzed in macroscopic terms. The very simplified model of [Borisov et al. 2005], whereby both receptor monomers have to bind their signal-ligand before they can undergo dimerization (see Figure 2.4), allowed the application of the macroscopic modeling approach for the already dimerized receptor analogous to the monomeric case. The second task of this thesis however, was to include and analyze the dimerization process itself. Therefore a detailed model, respecting the full combinatorial complexity of dimer formation is necessary (Chapter 4.2).

Chapter 3

Methods

In reference to [Conzelmann et al. 2006] the reduction method presented in this thesis can be structured in the already mentioned three essential steps.

1. generating a complete mechanistic description of the considered receptor or scaffold protein as in [Faeder et al. 2003; Hlavacek et al. 2003; Blinov et al. 2004]

2. introduction of macroscopic [Borisov et al. 2005] and mesoscopic [Conzelmann et al. 2006] states providing a smooth and bijective state-space transformation

3. neglecting the equations not influencing the dynamics of the macro states

The problem for a given system, with a certain pattern of interactions is to introduce the macroscopic and mesoscopic states following a hierarchical pattern and providing a smooth and bijective state-space transformation [Conzelmann et al. 2006]. A task of this thesis was to search those macroscopic and mesoscopic variables for the chain-formation problem as well as the receptor-dimerization process, that result in some sort of modularization and/or model reduction and still possess biological relevance. Therefore it was necessary to modify the macroscopic and mesoscopic variables as defined in [Borisov et al. 2005] and [Conzelmann et al. 2006].

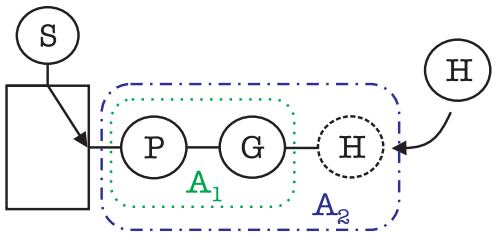

Figure 3.1: Exemplary schematic representation of the chain formation problem. Subsequent binding of P, G and H on one binding site, leading to the chain-like signal molecule complex PGH. In reference to [Borisov et al. 2005] and [Conzelmann et al. 2006] the PG complex and the PGH complex site can be considered as the effectors A_1 and A_2 respectively. Then, the docking of H is described by the biochemical reaction of A_1 into A_2 on the corresponding docking site: $A_1 \rightarrow A_2$.

Chain-formation occurs, if several signaling molecules are engaged subsequently on one docking site (see Figure 3.1). Each new binding of a signaling molecule leads to a new state of the docking site. The subsequent engagement of a signaling molecule on the domain i can be interpreted as the transformation of the effector E_i^k into the effector E_i^l. To each state of the docking site corresponds one-to-onely a macro-state as defined in [Borisov et al. 2005]. The definitions of the macro- and mesoscopic states in [Borisov et al. 2005] and [Conzelmann et al. 2006] do not take into account, that the effectors E_i^l and E_i^k are mainly composed of the same signaling molecules, and differ only in the last bound signaling molecule. Therefore new mesoscopic and macroscopic variables . Therefore, new macroscopic and mesoscopic variables (the effector occupancy levels and the levels of occurence) are defined, which take the composition of the effectors into account. Further, in contrast to [Borisov et al. 2005] and [Conzelmann et al. 2006], the here defined levels of occurence do not need any kind of reference molecule like a receptor or a scaffold.

It is possible to predict the maximal level of reduction, i.e. how many equations can be omitted, for a given output vector $\mathbf{y}$, which is usually one or more macroscopic variables. Therefore an observability analysis of the linearized model is performed, and the equations of the internal dynam-

ics are neglected[1]. Note that the output is a linear function of the states, but the system's dynamics are nonlinear, generally resulting in constricted statements of the linear observability analysis. This means that the original system still might be observable, even if the linearized model is not (see also Section 1.4 on page 25). In that case no reduction is possible, although it has been predicted by the linear observability analysis. However in this thesis, all considered cases were predicted correctly, meaning that the rank of the observability matrix was equal to the minimal number of equations necessary to describe the output behavior.

In order to describe the considered processes accurately and unambiguous, clear notations and definitions are necessary, especially for the setup of quantitative mathematical models. For this reason the notations used in this thesis are presented below, and the macroscopic and mesoscopic states are defined in the following.

3.1 Notations: Complexes and receptors

3.1.1 Notation of [Conzelmann et al. 2006]

In the following we will consider a receptor or scaffold protein R with b different binding sites each able to bind e_i different effectors or adapter proteins, and hence be in $e_i + 1$ different states: either unoccupied or occupied by one of the e_i effectors. Therefore we get the total number of different species as

$$n = \prod_{i=1}^{b} (e_i + 1)$$

A chemical reaction on the domain i can be seen as the biochemical transformation of the effector E_i^k to the effector E_i^l (see also the chain-formation problem on page 46).

$$R[*, \cdots, E_i^k, \cdots, *] \quad \rightarrow \quad R[*, \cdots, E_i^l, \cdots, *]$$

[1]Note that the internal dynamics have to be stable, which is however true for all considered models of signal transduction pathways.

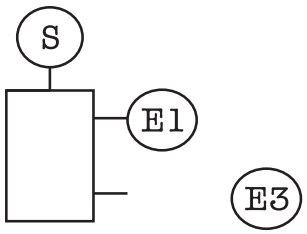

Figure 3.2: Exemplary receptor with three binding domains, the first is occupied with S, the second with $E1$ and the last unoccupied. As you can see there exists a third effector $E3$ which is not bound. The notation for this receptor species is $R[S, E1, 0]$.

Each effector E_i^k on a binding site i can react into another effector E_i^l ($k \neq l \leq e_i$), or separate from the receptor reversibly, resulting in e_i changes at this site i. Since each of the n species can change the state of each of its b binding sites, the total number of theoretical reactions is

$$\sum_{j=1}^{b} e_j \prod_{i=1, i\neq j}^{b} (e_i + 1)$$

In order to get a clear notation of the receptor species, they are labelled as follows. A specie is $R[E_1^{j_1}, \cdots, E_b^{j_b}]$, where $E_i^{j_i}$ denotes the j_i-th effector occupying the i-th binding domain. So the i-th position in $R[\ldots]$ describes the i-th binding domain of the receptor (see Figure 3.2). If one domain is unoccupied, we write a zero at the corresponding position. The advance of this notation is, that it is clear which molecules are forming the multi-molecule-complex.

3.1.2 Extensions

Complex

The notations in [Conzelmann et al. 2006] were sufficient to consider and decouple signaling events occurring on distinct docking sites. To each docking site a formal set of effectors and all possible reactions between these effectors were specified. However, for the analysis of special processes on one binding site, a closer look at these effectors, their composition and biological

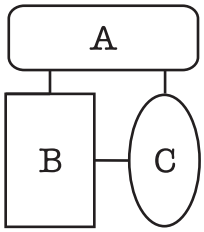

Figure 3.3: Three molecules A, B, C able to bind each other on distinct domains, leading to the possible molecule complexes AB, AC, BC and ABC.

function is necessary. A very common process is the subsequent assimilation of signaling molecules in a sequence of reactions, resulting in a chain-like multi-molecule complex. In terms of effectors as in [Conzelmann et al. 2006] chains with different sizes are formally equivalent to different effectors. The assimilation of another signaling molecule prolonging the chain is a chemical transformation from one effector to another. However, these two effectors share common properties. They consist mainly of the same signaling molecules, only differing in the latest attached molecule. To benefit from this similarities, the notation of the complex is introduced as the assimilation of multiple signaling-molecules. A complex can be denoted as a string of its compartments, similar to the chemical notation of a substance.

Example 2. *A complex consisting of the three molecules A, B, C as illustrated in Figure 3.3, is denoted as ABC and contains the complexes AB, AC, BC, A, B, C since any binding combination of the single molecules is possible.*

Dimer

Another necessary extension concerns dimeric receptors. The notation of the single receptors as above is kept, and DM is used to denote the fact that the receptors have formed a complex of two receptors: $DM[R[\ldots], R[\ldots]]$.

3.2 Domain interactions

Distinct binding domains can influence each other by means of conformational changes, allosteric regulation and blocking of binding sites through

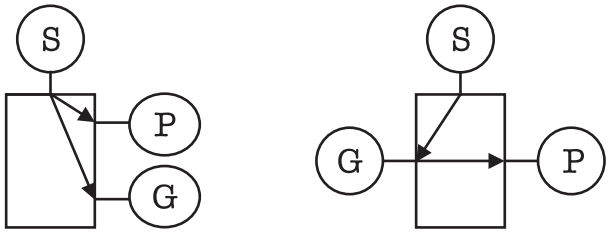

Figure 3.4: Two receptors with different domain dependencies. On the left hand side a receptor is shown with one controlling domain S, influencing the the two effector binding sites P and G directly. There are no crossover effects between the effector binding sites P and G and no influence of these on the controlling binding site S. On the right hand side a sequential structure of dependency is illustrated. The controlling domain S influences the intermediate controlling domain G which in turn effects the binding site P. Again there exists no other dependencies.

effector binding. Through different patterns of these interactions or dependencies certain biological functionalities are implemented. For example, ligand binding to one controlling domain can have great influence on several other domains, namely the increase or decrease of association and dissociation rates, by changing the molecules three dimensional conformation, easing downstreaming signaling molecules the access to the docking site. To illustrate this fact an arrow from the triggering or controlling domain is drawn to the dependent one, as is shown exemplarily in Figure 3.4. For mutual dependencies an arrow with two heads ($\leftrightarrow$) is used.

To implement the domain interactions mathematically, consider the binding of a signal molecule on a docking site i. The chemical reaction on this domain i transforms the effector E_i^k to the effector E_i^l.

$$R[*, \cdots, E_i^k, \cdots, *] \quad \rightarrow \quad R[*, \cdots, E_i^l, \cdots, *]$$

However, there are several reactions which have to be accounted for, since these reaction can occur for all species having the effector E_k bound to domain i, and the states of the other domains $j \neq i$ (above denoted by $*$) are undefined. If the domain i is completely independent from all other domains, all these reactions share the same association and dissociation constants. If however the domain i depends on another domain j, these reactions must be

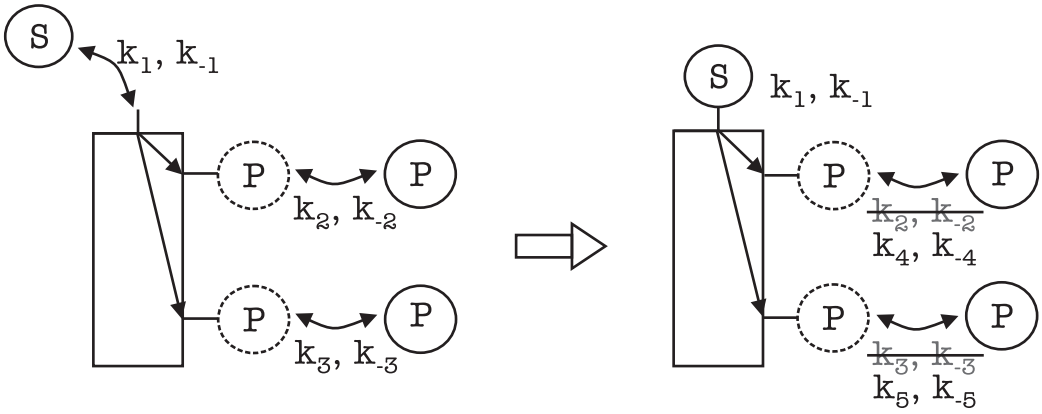

Figure 3.5: A receptor R or scaffold S with one controlling domain as considered in [Borisov et al. 2005] and [Conzelmann et al. 2006]. The S-binding domain influences the P- and G-binding domain, as indicated by the arrows. When S binds, the association and dissociation constants of the P- and G-binding domain change:

	without S	with S
S-domain	k_1, k_{-1}	k_1, k_{-1}
P-domain	k_2, k_{-2}	k_4, k_{-4}
G-domain	k_3, k_{-3}	k_5, k_{-5}

grouped according to the state of the domain j. Then, a different association and dissociation constant is assigned to each group of reactions (see Figure 3.5).

3.3 Macroscopic and mesoscopic states

Although already mentioned in Chapter 2.3, a brief explanation of macroscopic and mesoscopic variables is given here. Macroscopic states describe a system at lower information level than microscopic states, but still meet the given requirements of accuracy. For instance in thermodynamics science, the consideration of macroscopic states like temperature, pressure, enthalpy or entropy is sufficient to describe the behavior of the system and knowledge of microscopic states like position and velocity of each of the involved molecules is not necessary for most technical applications. The macroscopic states introduced in [Borisov et al. 2005] and [Conzelmann et al. 2006] transfer these

concept of lower detailed description to biochemical reaction networks of receptors and scaffolds as they occur in signal transduction. As we will see on page 56, mesoscopic states are states of intermediate detail and provide a one-to-one and onto (bijective) and smooth state-space transformation, transforming species onto mesoscopic space, and since the transformation is one-to-one also back from mesoscopic to species space [Conzelmann et al. 2006]. In addition to the levels of occupancy as first defined in [Conzelmann et al. 2006], another set of macroscopic and mesoscopic variables, the more generally applicable levels of occurrence are defined below.

The transformation of the microscopic state-space (Equation (2.1) on page 41) to the macro- and mesoscopic space (Equation (2.3) on page 42) allows the systematic investigation of signal transduction networks under the aspect of modularization and model reduction by integrating different interaction patterns. Depending on the pattern of dependency, certain sets of reactions possess equal association and dissociation rates, which can result in a partwise decoupled system of ordinary differential equations of macro- and mesoscopic states (see Chapter 2.3). As we will see in Chapter 4, these subsystems possess a hierarchical pattern of interactions and describe different biological processes. Depending on the set of output variables (interesting concentrations which are usually macroscopic variables), some of these subsystems are sufficient to describe the output and can be solved independently. In that case (and if the neglected subsystems are stable) a model reduction is possible.

3.3.1 Occupancy levels

As we have seen in Section 1.1.4 receptors and scaffolds are able to bind multiple adaptor and signaling proteins and present fundamental participants in signal transduction networks. Occupancy levels present a measure for the amount of particular effectors docked on their specific binding domains and can be interpreted as concentration of this receptor-effector-complex corresponding to a given domain independently from all docking sites [Borisov et al. 2005].

Let R be a receptor or scaffold protein with b different docking sites $i \in \{1, \ldots, b\}$, each able to assemble one of the several effectors E_j, which may consist of multiple signaling proteins. Thus for each docking site a related set of effectors can be defined as:

Definition 4. *Domain specific effector set*

$$\mathbb{E}_i \quad = \quad \textit{set of all effector concentrations able to be assembled on the } i\textit{-th domain of the receptor}$$

For each docking site there are $m_i = \|\mathbb{E}_i\|$ different effectors as binding partners for the receptor or scaffold, and consequently each site can be in one of $m_i + 1$ states, namely either occupied by one of the m_i effectors or unoccupied. By the b different effector sets, each of size m_i, $\sum_i^b m_i$ different levels of occupancy are given for the whole receptor or scaffold (see Example 3). The following defined occupancy levels can be interpreted as the concentration of this different occupancies in the system.

The $\sum_i^b m_i$ occupancy levels corresponding to the b different domains of the receptor or scaffold are defined as the sum over all receptor species having bound the particular effectors to the appropriate domains.

Definition 5. *Occupancy level*

$$\begin{aligned} \mathbb{R}_i[E_j] \quad &= \quad \textit{set of all receptor species concentrations with bound effector } E_j \textit{ on the } i\textit{-th docking site} \\ occR_i[E_j] \quad &= \quad \sum_i x_i \,, \quad x_i \in \mathbb{R}_i[E_j] \end{aligned}$$

As already mentioned the occupancy levels can be seen as the concentrations of the receptor or scaffolding proteins having bound a given effector to a specified domain, independently from the states of all other docking sites. To make this relation clearer consider the following example:

Example 3. *Consider a receptor with two binding domains $i \in \{1, 2\}$. The first domain is able to bind the effectors S and P, the second can be occupied*

by P only. Hence the domain specific effector sets are

$$\begin{aligned} \mathbb{E}_1 &= \{S, P\}, \quad m_1 = 2 \\ \mathbb{E}_2 &= \{P\}, \quad m_2 = 1 \end{aligned}$$

The receptor species are R[0,0], R[S,0], R[P,0], R[0,P], R[S,P], R[P,P]. The first docking site can be occupied by $m_1 = 2$ *different effectors, the second by* $m_2 = 1$ *different effectors, leading to* $\sum_i^2 m_i$ *different occupancy levels.*

$$\begin{aligned} occR_1[S] &= R[S,0] + R[S,P] \\ occR_1[P] &= R[P,0] + R[P,P] \\ occR_2[P] &= R[0,P] + R[S,P] \end{aligned}$$

The occupancy levels of Definition 5 are pure macroscopic states as they were first defined in [Borisov et al. 2005], and have high biological relevance as domains are identified as functional units of proteins and we talk about domain interactions in biological terms. To make this biological relevance clearer, consider the following example.

Example 4. *The GAB scaffolding protein have a central role in cellular growth, transformation and apoptosis. Binding of Grb2 and Sos enables the activation of the small GTPase Ras and leads to the activation of the MAPK-cascade, which promotes mitogenesis and differentiation. Independently GAB is able to bind p85, which initiates the P13K/AKT pathway involved in glucose and lipid metabolism. This two processes can occur successively or simultaneously and incidences corresponding to one binding domain (for example the binding of Sos) would generate a different microscopic state or species whether this incidences have no effect on the processes of the other domains (as for example the P13K/AKT activation).*

The second feature of the occupancy levels is measurability, i.e. they can be quantified in experimental studies by western blot analysis using site specific antibodies, whereas species concentrations cannot be determined experimentally at the state-of-the-art.

3.3.2 Higher order occupancy levels

The number of species in a system describing a receptor or scaffolding protein is much higher than the number of occupancy levels, therefore additional mesoscopic states are introduced providing a state space transformation so that the receptor or scaffolding protein system can be described in the macro- and mesoscopic space. With the following defined higher order occupancy levels a macro- and mesoscopic space of equal dimension to the species space is given, and a smooth and bijective transformation between the two spaces can be conducted.

Continuative to the occupancy levels mesoscopic states called higher order occupancy levels can be defined as follows:

Definition 6. *p-th order occupancy level*

$$\begin{aligned} \mathbb{R}_{i\cdots j}[E_k, \cdots, E_l] &= \textit{set of all receptor species concentrations} \\ &\quad \textit{having bound p effectors } E_k, \cdots, E_l \\ &\quad \textit{such that } E_k \textit{ bound to the i-th, ...,} \\ &\quad E_l \textit{ bound to the j-th docking site.} \\ occR_{i\cdots j}[E_k, \cdots, E_l] &= \sum_i x_i \,, \quad x_i \in \mathbb{R}_{i\cdots j}[E_k, \cdots, E_l] \end{aligned}$$

These higher order occupancies are states of intermediate detail, i.e. the higher the order of the occupancy level, the more specific is it. The b-th order occupancy denotes a receptor, where all docking sites have bound an effector, and thus the b-th order occupancy level is equivalent to the species of the full occupied receptor or scaffold. The number of all occupancy levels is equivalent to the number of species.

Example 5. *The higher order occupancy levels in the example above can only be of second order since the receptor only has two docking sites. Thus they are the most specific mesoscopic states possible; they are equal to species concentrations (microscopic) and no sums have to be made.*

$$\begin{aligned} occR_{12}[S, P] &= R[S, P] \\ occR_{12}[P, P] &= R[P, P] \end{aligned}$$

The higher order occupancy levels were first defined in [Conzelmann et al. 2006] as mesoscopic states and present a full set of states necessary to describe the system without information loss, allowing the exact analysis of different dependency patterns. This can be illustrated with the following scheme of the state-space transformation

$$\begin{array}{ccc} \text{original coordinates:} & & \text{new coordinates:} \\ R[\cdots, E_i, \cdots], & \overset{\mathbf{z}=T\mathbf{x}}{\Longleftrightarrow} & R[\cdots, E_i, \cdots], \\ E_i = \{\{\}, E_i^1, \cdots, E_i^{e_i}\} & & E_i = \{*, E_i^1, \cdots, E_i^{e_i}\} \end{array}$$

where $\{\}$ denotes the unoccupied docking site and $*$ means the docking site is not specified. As can be seen, each docking site in the original coordinates can either be unoccupied or occupied by one of its effectors E_i^k, $k \in \{1, \cdots, e_i\}$. In the new coordinates, the docking site can either be unspecified $(*)$ or be occupied by one of its effectors. Consequently the dimension of the original space equals the dimension of the new space.

3.3.3 Effector occupancy levels

The above defined occupancy levels are sums of concentration that relate to specific states of the specified docking sites. The docking sites appropriated to the occupancy level are denoted by the indexes $i \ldots j$ of $occR_{i\ldots j}[\cdot]$. There are several cases, where a slightly different approach is appropriate, and non-binding-site specific variables are more useful. An example is a dimeric receptor possessing pairs of docking sites, possessing the same properties and thus are able to bind the same effectors, and triggering equal subsequent events. In this case, it is advantageous to integrate the state of both similar docking sites in one macro- or mesoscopic variable (see results in Chapter 4.2). On that account effector occupancy levels are defined, which are independent from the receptor's or scaffold's binding sites and denote the relative concentration of a specific effector bound to the whole receptor (not a specific binding site).

Definition 7. *Effector occupancy level*

$$occR[E_i] \;=\; \sum_j R_j[E_i]$$

An effector occupancy level is the weighted sum over all species having bound the given effector, and can be interpreted as the concentration of the effector assimilated with the receptor or scaffold. Note that for instance species occupied twice by the effector on two different domains are counted twice. The effector occupancy level $occR[E_i]$ can be calculated by summating the (domain specific) occupancy levels of the specified effector $\sum_j occR_j[E_i]$.

Example 6. *In the example above the effector P can be assimilated by two binding domains, the first and the second. Hence the effector occupancy level of P is different from any of the "normal" or domain specific occupancy levels.*

$$occR[P] \;=\; R[0,P] + R[P,0] + 2R[P,P] \;=\; occR_1[P] + occR_2[P]$$

The advantage of the effector occupancy levels is the integration on system inherent symmetries, resulting in a further model reduction than the ordinary occupancy levels. A good example is a dimerized receptor possessing each binding domain of the monomer twice, each triggering equal subsequent events (see Chapter 4.2).

The common advantage of all occupancy level definitions is the dissection of complex signaling processes of receptors and scaffolds according to distinct docking sites. However there are other processes for which no modularization or reduction is achieved. For instance, the subsequent binding of signaling molecules on one domain, resulting in a chain-like multi protein complex (see Figure 3.1), is a process which cannot be modularized by the levels of occupancy (see Chapter 4.1).

3.3.4 Occurrence levels

For more complex systems with several distinct receptors and scaffold proteins a more general approach leading to macroscopic and mesoscopic states is suitable. The major disadvantage of the occupancy levels is, that they always refer to some kind of carrier molecule like a receptor or a scaffold. Taking a different approach, the occupancy of a receptor by an effector can

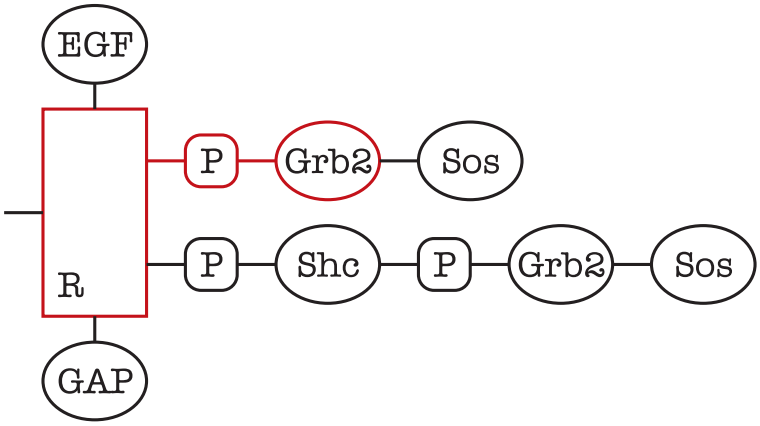

Figure 3.6: Schematic representation of one dimer half of the EGF receptor with several of its adaptor proteins assembled on different binding domains. The occurrence level of the R-P-Grb2-complex O[RPGrb2], for example, is the sum of all species having R-P-Grb2 assembled together.

be seen as the occurrence of a receptor-effector complex. Consequently a generalized concept of macroscopic states is presented by the consideration of multi-molecule-complexes and their occurrence in the system.

With occurrence levels we want to describe the frequency of a molecule or molecule complex occurring in the system. From this point of view it does not matter whether it occurs as pure complex, or it is bound or assembled to other complexes. If we take for example the commonly known Michaelis Menten mechanism, the sum of enzyme and enzyme-substrate-complex would be the total number of enzyme molecules occurring in the system and such be equivalent to the occurrence level of the enzyme.

Definition 8. *Occurrence level*

$$\begin{aligned} \mathbb{O}[C] &= \textit{set of all species having the multi-protein complex } C \textit{ as compartment, including } C \textit{ itself} \\ Occ[C] &= \sum_i x_i\,, \quad x_i \in \mathbb{O}[C] \end{aligned}$$

The compared to the occupancy levels more general concept of occurrence levels evades the reference of the macroscopic states to a carrier molecule. The consideration of occurrences of complexes enables us to describe

all kinds of biological reaction networks, in terms of macroscopic states. Several receptors or scaffolds and general complex formation independent of any "carrier-molecule" can be described in an formalized manner.

Example 7. *The EGF-receptor as shown in Figure 3.6 possesses several docking sites able to bind multiple signal molecules in a chain-like manner. The Grb2 molecule itself can be seen as a signal molecule with several binding domains able to bind the phosphorylated receptor, the phosphorylated signaling molecule Shc and the signaling molecule Sos. A processes that is not necessarily associated with the receptor, and thus not describable completely with the occupancy levels, is for instance the binding of Sos to Grb2. In terms of occupancy P-Grb2 would be treated as effector ($E_2^2 = PGrb$) on the receptors docking site, as well as P-Grb2-Sos ($E_2^3 = PGrb2Sos$), and the binding of Sos would rather be described as the reaction of the receptor's domain from E_2^2 to E_2^3, than the attachment from Sos onto Grb2. However, it is possible, that free Grb2 engages Sos during not connected to the receptor. To model this event macroscopically, the occurrence level description is necessary.*

3.3.5 Summary

The occupancy levels (Definition 5) are pure macroscopic states as they were first defined in [Borisov et al. 2005], and have high biological relevance as domains are identified as functional units of proteins and we talk about domain interactions in biological terms. The second feature of the occupancy levels is measurability, i.e. they can be quantified in experimental studies by western blot analysis using site specific antibodies, whereas species concentrations cannot be determined experimentally at the state-of-the-art.

The higher order occupancy levels (Definition 6) were first defined in [Conzelmann et al. 2006] as mesoscopic states and present a full set of states necessary to describe the system without loss of information. They provide a bijective state-space transformation, which allows the exact analysis of different dependency patterns.

The novel defined effector occupancy levels (Definition 7) integrate bio-

logical and structural symmetries of the considered systems for the sake of modularization and model reduction. Symmetric processes as for instance receptor dimerization are structured more efficient and describable with less equations than with the ordinary occupancy levels.

The more general concept of the novel defined occurrence levels (Definition 8) takes into account that effectors composed of several signaling molecules may possess equal compartments. Further it evades reference to a carrier molecule, as for instance a receptor or a scaffold. The consideration of occurrences of complexes enables us to describe all kinds of biological reaction networks, in terms of macroscopic states. Several receptors or scaffolds and general complex formation independent of any carrier-molecule can be described in an formalized manner. In this way occurrence levels provide an accurate way to model receptors, scaffolds and adapter proteins all with several binding domains, able to assimilate multi-protein-complexes with each other, as well as with other effector and signaling molecules.

3.4 Transformation, model reduction and observability analysis

One aim of the macroscopic approach is the systematically derivation of a model reduction through the application of a state-space transformation and the neglect of decoupled dynamics of minor relevance. Since macro- and mesoscopic states as defined above are natural linear combinations of species, the corresponding state-space transformation between the microscopic space and the macro- and mesoscopic space is linear. Observability analysis of biological relevant outputs is used to predict a complete basis for the transformed space and construct suitable mesoscopic states. In the following section the context between macroscopic states, linear state-space transformations, observability analysis and model reduction is discussed.

3.4.1 Linear state space transformation of a nonlinear system

The models we use here are set up by applying the law of mass action to the reaction networks. The resulting model describes the change of the species concentration in time, and is a set of nonlinear ordinary differential equations on the manifold M. We can write it in the form:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \quad \mathbf{x} \in M \subseteq \mathbb{R}^n$$

We are interested in the macroscopic states, especially the occupancy levels of bound effectors, which is taken into account by a linear output function, which we can write as a matrix multiplication.

$$y_i = \mathbf{h}_i^T \cdot \mathbf{x}, \quad \mathbf{h} \in \mathbb{N}^n$$

Where $\mathbb{N}^n$ is the n dimensional space of natural numbers, whereby the output is a weighted sum of species concentrations. Now we perform a linear state space transformation $\mathbf{z} = T \cdot \mathbf{x}$.

$$\begin{aligned} \dot{\mathbf{z}} &= T \cdot \dot{\mathbf{x}}, \quad T \in \mathbb{R}^{n \times n} \\ &= T \cdot \mathbf{f}(\mathbf{x}) = \underbrace{T \cdot \mathbf{f}(T^{-1} \cdot \mathbf{z})}_{\mathbf{g}(\mathbf{z})} \end{aligned}$$

If this transformation has decoupled some states $z_1, \cdots, z_r$ the equations are of the form:

$$\mathbf{g}(\mathbf{z}) = \begin{bmatrix} g_1(z_1, \cdots, z_r) \\ \vdots \\ g_r(z_1, \cdots, z_r) \\ g_{r+1}(z_1, \cdots, z_r, \cdots, z_n) \\ \vdots \\ g_n(z_1, \cdots, z_r, \cdots, z_n) \end{bmatrix}$$

We will now consider the Jacobian matrix of the transformed system to analyze if we can reach the decoupling through such a linear transformation and how it has to look like. For all $\mathbf{z} \in \mathbb{R}^+$ the jacobian matrix of the

transformed decoupled system has the form:

$$\tilde{A} = \frac{\partial \mathbf{g}(\mathbf{z})}{\partial \mathbf{z}} = \begin{bmatrix} a_{1,1}(\mathbf{z}) & \cdots & a_{1,1}(\mathbf{z}) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{r,1}(\mathbf{z}) & \cdots & a_{r,r}(\mathbf{z}) & 0 & \cdots & 0 \\ a_{r+1,1}(\mathbf{z}) & \cdots & a_{r+1,r}(\mathbf{z}) & a_{r+1,r+1}(\mathbf{z}) & \cdots & a_{n,n}(\mathbf{z}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n,1}(\mathbf{z}) & \cdots & a_{n,r}(\mathbf{z}) & a_{n,r+1}(\mathbf{z}) & \cdots & a_{n,n}(\mathbf{z}) \end{bmatrix}$$

We can express $\frac{\partial \mathbf{g}(\mathbf{z})}{\partial \mathbf{z}}$ through the Jacobian matrix of the original system $\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}}$:

$$\begin{aligned} & \mathbf{g}(\mathbf{z}) && = T \cdot \mathbf{f}(T^{-1} \cdot \mathbf{x}) \\ \Rightarrow \quad & \frac{\partial \mathbf{g}(\mathbf{z})}{\partial \mathbf{z}} && = T \cdot \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \cdot \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \\ & && = T \cdot \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \cdot T^{-1} \end{aligned}$$

As we can see it holds that the Jacobian matrix of the transformed system is equivalent to the transformed Jacobian matrix of the original system, as far as we perform a linear transformation. Thus the transformed system is decoupled if and only if the transformation T decouples the linearized original system $\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}}, \quad \forall\, \mathbf{x} \in M$. It is clear that such transformation can be found only under certain conditions for $\mathbf{f}(\mathbf{x})$. We will briefly discuss these conditions in the following.

3.4.2 The observability analysis helps to construct a suitable transformation

In the previous section we have seen, that a decoupling of a nonlinear dynamical system of the form

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}) \\ y_i &= h_i^T \cdot \mathbf{x}, \quad \mathbf{h} \in \mathbb{N}^n \end{aligned} \tag{3.1}$$

can be achieved, if the jacobian matrix at any point is decoupled. In order to check if this is possible with a linear state space transformation, the observability analysis is used.

At the beginning of this section we have defined linear outputs of the form $y_i = h_i^T \cdot \mathbf{x}$. Performing an observability analysis as in Chapter 1.4 gives the observability matrix Q. By reducing Q through building linear combinations of the rows until ranks-form is reached, we get a constant matrix Q_B with rank $r = \max_{\mathbf{x}\in M}\{rank(Q(\mathbf{x}))\}$. Thereby $\mathbf{x}$ is treated as an arbitrary parameter. The simplest basis for the by Q spanned observability space is given by the rows of Q_B.

$$Q_B \quad = \quad RowReduce(Q), \quad Z_r = span(Q_B)$$

This basis matrix is constant, i.e. all entries are independent from $\mathbf{x}$, and defines the maximal[2] observability space for all $\mathbf{x} \in M$. We use Q_B as the first r rows of the transformation matrix T and fill up the rest such that T is smooth and nonsingular. This analysis predicts the decoupling of the first r transformed states $z_1, \cdots, z_r$ since $\mathbf{y}$ has rank r and the transformation lies within the from Q_B spanned observability space. In order to get biological relevant states the first r rows of the performed transformation was constructed of linear combinations of Q_B. The reduction is achieved just in omitting the resulting equations of the internal dynamics.

3.4.3 Linearization

Even for the simple and small systems in this work, analytic nonlinear observability analysis was very expensive. For this reason, the system was linearized at an arbitrary point, and the kinetic parameters were chosen randomly as integers. Although it was not clear if the row reduced observability matrix Q_B is independent from the linearization point[2], this method predicted the reduction, which was finally achieved through the state space transformation, in all cases correctly.

Performing a Taylor series linearization of the system (3.1) in an arbitrary point $\mathbf{x}_0$ gives

$$\frac{d}{dt}(\mathbf{x}_0 + \tilde{\mathbf{x}}) \quad = \quad \mathbf{f}(\mathbf{x}_0) + \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\mathbf{x}_0) \cdot \tilde{\mathbf{x}} + O(\tilde{\mathbf{x}}^2)$$

[2]There might be singular points $\mathbf{x}_0$ for which the observability of some additional states get lost, i.e. $rank(Q(x_0)) < r = \max_{\mathbf{x}\in M}\{rank(Q(x))\}$

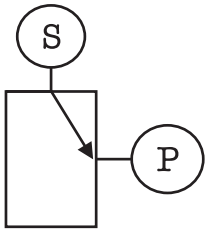

Figure 3.7: Monomer receptor with one extracellular and one intracellular effector binding domain. The extracellular docking site is able to bind the signal molecule S, which changes the affinity of the intracellular or downstreaming docking site. There do not exist any backtracking effects from the downstreaming domain to the signal domain.

If we neglect the terms of higher order $O(\tilde{\mathbf{x}})$ and if the Jacobian matrix of $\mathbf{f}$ is non-singular ($rank(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}) = n$), then the change of the system in the neighborhood of $\mathbf{x}_0$ can be approximated by the equations

$$\begin{aligned} \dot{\tilde{\mathbf{x}}} &= \underbrace{-\dot{\mathbf{x}}_0 + \mathbf{f}(\mathbf{x}_0)}_{=0} + \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\mathbf{x}_0) \cdot \tilde{\mathbf{x}} \\ y_i &= h_i^T \cdot (\mathbf{x}_0 + \tilde{\mathbf{x}}) \end{aligned} \tag{3.2}$$

The Equations (3.2) possess a linear system[3] for $\tilde{\mathbf{x}}$. If the linearization point $\mathbf{x}_0$ is not a steady state of the System (3.1) (i.e. $\mathbf{f}(\mathbf{x}_{steady-state}) = 0$), then the system runs away from $\mathbf{x}_0$ and the approximation above is only good for a short period of time. In order to observe the system during the entire run of an trajectory, $\mathbf{x}_0$ has to be updated continuously leading to a time-variant problem: $\mathbf{x}_0 = \mathbf{x}_0(t)$ with $\mathbf{x}_0(t)$ is a solution of Equation (3.1).

Example 8. *Let us consider a receptor possessing two binding domains. The binding of S triggers a conformational change of the receptors structure accelerating effector binding of P as shown in Figure 3.7, and we can talk from the effector binding site S as a signal binding site or controlling domain. In order to get a complete mathematical model, all possible reactions must be*

[3]The System (3.2) is a special form of a linear systems class in control theory referred to as multiple input, multiple output systems (MIMO), if the linearization point $\mathbf{x}_0$ is considered as input: $\mathbf{x} = A\mathbf{x} + B\mathbf{u}$, $y = C\mathbf{x} + D\mathbf{u}$ with $A = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\mathbf{x_0})$, $B = 0$, $C = h^T$, $D = h^T$,$\mathbf{u} = \mathbf{x}_0$

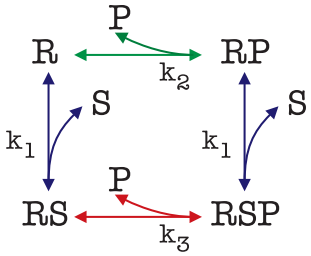

Figure 3.8: Example reaction scheme. We have a molecule R, which can react with two other molecules S and P and so form the molecule complexes RS, RP, RSP. The reaction constants of the two reactions with S are equal

considered. This includes the binding of S to the pure receptor as well as to the receptor bound by P and the binding of P to the pure receptor as well as to the receptor bound by S respectively. The complete reaction scheme is shown in Figure 3.8. Taking into account that the signal binding site is independent from from the P-binding domain, we can assume the signal binding to be independent from the receptors state and set the reaction constants for both S assimilating reactions equal k_1*. The other reactions engaging P have distinct constants* k_2, k_3*. Applying the law of mass action gives the reaction rates to*

$$\begin{aligned} r_1 &= k_1 R[0,0]S - k_{-1} R[S,0] \\ r_2 &= k_1 R[0,P]S - k_{-1} R[S,P] \\ r_3 &= k_2 R[0,0]P - k_{-2} R[0,P] \\ r_4 &= k_3 R[S,0]P - k_{-3} R[S,P] \end{aligned}$$

The system is described through the set of ordinary differential equations of the receptor concentrations:

$$\frac{d}{dt}\begin{bmatrix} R[0,0] \\ R[S,0] \\ R[0,P] \\ R[S,P] \end{bmatrix} = \begin{bmatrix} -r_1 - r_3 \\ +r_1 - r_4 \\ r_2 + r_3 \\ +r_2 + r_4 \end{bmatrix}$$

A biologically interesting variable is the concentration of the receptor having bound the signal S, thus let us consider the occupancy level occR[S] as output.

$$y = occR[S] = R[S,0] + R[S,P]$$

With $\mathbf{x} = \begin{bmatrix} R[0,0] & R[S,0] & R[0,P] & R[S,P] \end{bmatrix}^T$ *and* $h = [\ 0\ 1\ 0\ 1\ ]$, *the system can be written as*

$$\dot{\mathbf{x}} = f(\mathbf{x}), \qquad y = h^T \cdot \mathbf{x}$$

The linear observability analysis is performed as follows,

$$A = \frac{\partial f}{\partial \mathbf{x}}, \qquad h^T = \frac{\partial y}{\partial \mathbf{x}}, \qquad Q = \begin{bmatrix} h^T \\ \vdots \\ h^T \cdot A^n \end{bmatrix}$$

giving the observability matrix as

$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ k_1 S & k_{-1} & k_1 S & k_{-1} \\ k_1 S(k_1 S + k_{-1}) & k_{-1}(k_1 S + k_{-1}) & k_1 S(k_1 S + k_{-1}) & k_{-1}(k_1 S + k_{-1}) \\ k_1 S(k_1 S + k_{-1})^2 & k_{-1}(k_1 S + k_{-1})^2 & k_1 S(k_1 S + k_{-1})^2 & k_{-1}(k_1 S + k_{-1})^2 \end{bmatrix}$$

As we can easily see there are dependent rows and we can reduce this matrix by adding multiples of one row to another until rank-form is reached, and take only the first r independent rows (omitting all zero rows in rank-form). Note that this does not change the spanned space for $S \neq 0$ *(*$span[Q] = span[Q_B]$*).*

$$Q_B = RowReduce(Q) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

This would give the first two states for the transformation, in fact the fist two rows present the easiest coordinates for the predicted reduction. In general, one can construct states with biological relevance, by taking linear combinations of the rows. Here, instead of taking row one (no S bound to R)

$$\begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R \\ RS \\ RP \\ RSP \end{bmatrix} = R + RP$$

we get a more interesting variable by adding row two on row one.

$$\tilde{Q}_B = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

The corresponding transformed states are the total concentration of all receptor molecules or the occurrence level of the receptor (z_1) and the occupancy level of S (z_2):

$$\begin{aligned} z_1 &= R[0,0] + R[S,0] + R[0,P] + R[S,P] \\ z_2 &= R[S,0] + R[S,P] \end{aligned}$$

The remaining $n - r$ states have to be chosen such that the resulting transformation is smooth and nonsingular, which is not shown in detail here. However the transformed system becomes:

$$\frac{d}{dt}\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} 0 \\ -k_1 \underbrace{(R[0,0] + R[0,P])}_{z_1 - z_2} S + k_{-1} \underbrace{(R[S,0] + R[S,P])}_{z_2} \\ g_3(z_1, z_2, z_3, z_4) \\ g_4(z_1, z_2, z_3, z_4) \end{bmatrix}$$

Only two states z_1 and z_2 are necessary, to describe the dynamic output behavior exactly.

$$\frac{d}{dt}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 0 \\ g_2(z_1, z_2) \end{bmatrix}, \qquad y = z_2$$

Further one of them is a conserved moiety itself ($\dot{z}_1 = 0$) and only one differential equation remains to be solved.

3.5 Mathematical programming tools

Because of the complexity of the considered models and the resulting calculatory expense, it was made use of the programming tool Mathematica. Nonlinear and linear observability analysis could be performed efficiently and fast in order to search the observability space for biologically relevant basis vectors. In order to double-check these vectors for the reduction performance and for the sake of analysis of the macroscopic model, the state space transformation could be done equally efficient. The result of this search are the levels of occurrence as defined in Chapter 3.3. The relevance of this definitions and its benefits for model analysis, namely the unveiling of system

inherent structures, were analyzed exemplarily on a variety of different models, and these which are sufficient to clarify these relevance are presented in Chapter 4.

Additionally a notebook script has been written, which allows the automatized setup of models of dimeric receptors as described in Chapter 4.2. A rule-based modeling algorithm was implemented, which allows the declaration of an arbitrary number of docking sites and effectors. A set of rules specifies the reactions and the dependency pattern of the docking sites through setting the reaction constants for the interacting proteins. On the basis of this information (the binding sites, their corresponding effectors and the rules) the reaction rates and the ordinary differential equations for the species concentrations are derived automatically.

Chapter 4

Results

4.1 Monomers

In this section we will apply the macro-state description using the occupancy levels of a monomeric receptor. The cases we will consider are based by the precondition, that the exterior signal binding influences (normally accelerates) the binding of the interior effectors (see Chapter 1). The signal concentration is treated as input signal (time varying). For simplifying the derivation process the effector concentrations are not balanced (as in [Borisov et al. 2005]), meaning that no ode's for them are set up. Hence this concentrations are treated as known time-varying values or constants (i.e. effector-pools). At the end of this section a brief interpretation of this simplification is given and it is shown how the equations can be extended considering balanced effectors without loosing the results. However the in chapter 3.3 newly presented concept of occurrence levels allows the consideration of effectors as balanced states, and leads to the same results directly.

A task of this thesis was to analyze the chain-formation problem as already mentioned in Chapter 3 . Therefore, the two simplest cases, the subsequent binding of two effectors and the binding of one effector subsequent to the phosphorylation of the docking site were considered. Using macroscopic and mesoscopic states as defined earlier [Borisov et al. 2005; Conzelmann et al. 2006], the here called occupancy levels, neither a model reduction nor a modularization was achieved. However we will see in this chapter, that

the here newly defined effector occupancy levels (Definition 7) and levels of occurrence (Definition 8) result in a modularization. A model reduction however could only be achieved under strong assumptions.

4.1.1 Review: Decoupling distinct binding sites

There has been previous work on model reduction through macroscopic description of monomeric receptors and scaffolds [Borisov et al. 2005; Conzelmann et al. 2006]. For the sake of completeness the results are briefly presented in the following.

Fully independent binding domains

If we apply the method of occupancy levels to a receptor with fully independent binding domains i each able to bind m_i domain specific effectors, we only need $\sum_{i=1}^{d} m_i$ equations - namely the levels of occupancy - to describe the system. We have one conserved moiety, the total concentration of the receptor, and m_i ordinary differential equation for each domain describing its level of occupancy, which can be solved independently from each other. A model reduction from $\prod m_i + 1$ down to $\sum_{i=1}^{d} m_i$ equations is achieved.

The easiest way to show this is to consider a receptor with two independent domains, one able to bind S the other able to bind P (Figure 4.1). The necessary macroscopic states are:

$$\begin{aligned} occR[] &= R[0,0] + R[S,0] + R[0,P] + R[S,P] \\ occR_1[S] &= R[S,0] + R[S,P] \\ occR_2[P] &= R[0,P] + R[S,P] \end{aligned}$$

Not necessary is the only second order mesoscopic state in the system $occR_{1,2}[S,P] = R[S,P]$.

One controlling domain

In the presence of a controlling domain able to bind a signal S, which influences downstreaming domains, the levels of occupancy are also suitable.

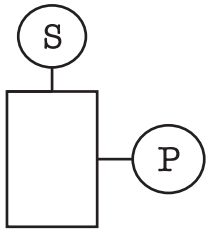

Figure 4.1: Monomer receptor with two totaly independent domains. Events occurring on one binding site have no effects on events of the other binding site.

The level of the signal bound to the controlling domain can be solved independently from all others, only the total concentration of the receptor is needed, and signaling events corresponding to distinct downstreaming domains can be decoupled from each other. To each downstreaming effector domain belongs a set of equations of the occupancy levels of the effectors and the second order occupancy levels of the signal and effectors that can be solved independently.

To make these modularization clear, the simplest system to consider is a receptor, possessing one controlling domain and two additional downstreaming domains as illustrated in Figure 4.2. Each docking site can only bind one effector, the controlling domain binds S, the first effector domain P and the second G, leading to 2^3 receptor species. The first module describes the signal binding and contains two states, the total concentration of the receptor and the first-order occupancy level of the signal:

$$\begin{aligned} z_1 &= occR[] = R[0,0,0] + R[S,0,0] + R[0,P,0] + R[0,0,G] + \\ &\qquad R[S,P,0] + R[S,0,G] + R[0,P,G] + R[S,P,G] \\ z_2 &= occR[S] = R[S,0,0] + R[S,P,0] + R[S,0,G] + R[S,P,G] \end{aligned}$$

Two additional states for each effector binding site are necessary:

$$\begin{aligned} z_2 &= occR[P] = R[0,P,0] + R[S,P,0] + R[0,P,G] + R[S,P,G] \\ z_3 &= occR[S,P] = R[S,P,0] + R[S,P,G] \end{aligned}$$

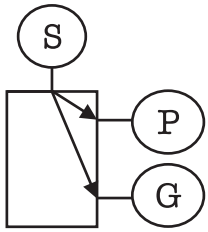

Figure 4.2: Monomer receptor with one extracellular and two intracellular effector binding domains. The extracellular docking site is able to bind the signal molecule S changing the affinity of the intracellular or downstreaming docking sites. These downstreaming sites are independent from each other. One is able to bind P, the other able to bind G. There do not exist any backtracking effects from the downstreaming domains to the signal domain.

Respectively

$$z_4 = occR[G]$$
$$z_5 = occR[S, G]$$

The remaining occupancy levels $occR[P, G]$ and $occR[S, P, G]$ can be neglected.

4.1.2 Modularization of multiple signaling events at one site

In the section above previously analyzed cases were presented, which deal with the modularization of distinct binding sites of monomers by means of occupancy levels. Since these modularization dissects the model into binding site specific subsystems, the analysis of the signaling events corresponding to these independent binding sites was suitable. The aim of these analysis was to reveal the limitations of the occupancy levels, unfold system inherent structures and, if possible, expand it to modularize more biological relevant cases. Therefore, it is suitable to distinguish between two cases that differ greatly in their biological meaning as well as their modeled reaction schemes: Phosphorylation on the one hand and binding of a signaling protein on the other.

Binding of two ordinary effectors without phosphorylation building a chain

Let us consider a receptor R as can be seen in Figure 4.3 with two distinct binding domains, one able to bind the extracellular signal molecule S the other the intracellular effector molecule P. P can form a complex with another effector G before or after docking to the receptor, thus the PG complex can bind to R as well, leading to a ordered chain of effectors on the receptors intracellular docking site. The binding of S does not depend on the occupancy of the other domains, and so the association and dissociation rates of the corresponding reactions rates are equal (k_1, k_{-1}). Binding of P to the receptor is assumed to be influenced by the signal binding domain, meaning that the affinity of the receptor R to the effector P changes upon ligand binding of S. Thus $k_2 \neq k_3$ and consequently $k_{-2} \neq k_{-3}$. The effector G reacts with the corresponding domain of P and thus we assume that it does not depend on the receptors conformation as well. Now we can parameterize the corresponding reaction rates with k_4 and k_{-4} for all G-related reactions. To analyze this event in detail, we assume the binding of PG complex to the receptor to be different from the binding of solitary P, which allows a feedback effect within the adaptor protein P from the G-binding domain to

the R-binding domain (illustrated in Figure 4.4). All possible reactions are

$$
\begin{aligned}
R[0,0] + S &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} R[S,0] \\
R[0,P] + S &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} R[S,P] \\
R[0,PG] + S &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} R[S,PG] \\
R[0,0] + P &\underset{k_{-1}}{\overset{k_2}{\rightleftharpoons}} R[0,P] \\
R[S,0] + P &\underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} R[S,P] \\
R[0,P] + G &\underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} R[0,PG] \\
R[S,P] + G &\underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} R[S,PG] \\
R[0,0] + PG &\underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} R[0,PG] \\
R[S,0] + PG &\underset{k_{-6}}{\overset{k_6}{\rightleftharpoons}} R[S,PG]
\end{aligned}
$$

A transformation to macroscopic states will be performed using occurrence levels. The use of occupancy levels is not suitable, because they neither lead to any kind of model reduction, nor provide any suitable modularization (equations not shown). Furthermore occupancy levels do not take into account basic properties of receptor-unrelated complex assimilation or chain-formation. For instance receptor species having bound P are summed up forming the occupancy level $occR_2[P]$, and receptors having bound the molecule complex PG generate the occupancy level $occR_2[PG]$. Although a receptor having bound the PG-complex implies that it has bound P as well, the corresponding concentrations $\{R[0,P], R[S,P]\}$ are not considered in the occupancy level $occR_2[P]$. This example shows that for each different state of the docking site distinct occupancy levels are generated, although there may be close relation between this different states. This circumstance can be avoided using in chapter 3.3 defined occurrence levels.

The occurrence levels used to transform the system, relate to multi-molecule-complexes of the receptor to keep close meaning to the occupancy

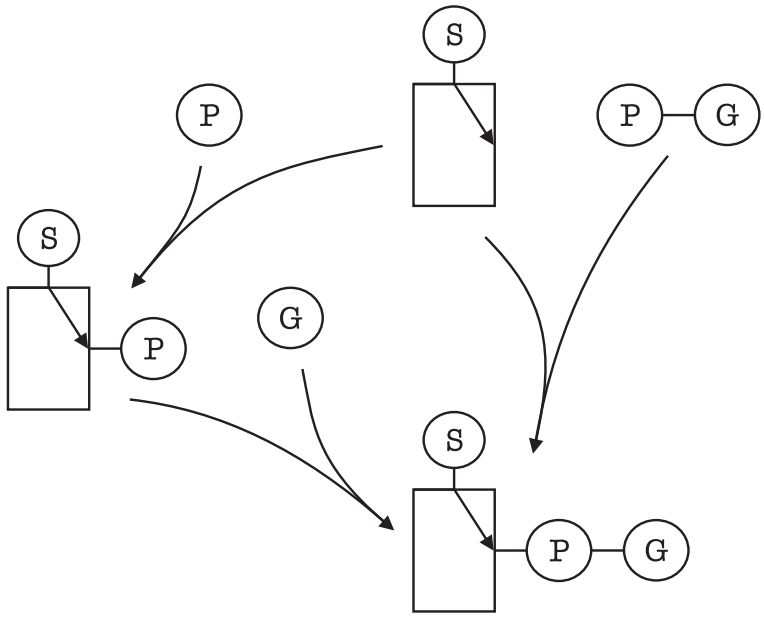

Figure 4.3: Monomer receptor with one extracellular and one intracellular effector binding domain. The extracellular docking site is able to bind the signal molecule S, which changes affinity of the intracellular docking site. Taking this dependence into account we can talk from the binding domain for S from an controlling or signal docking site. The intracellular binding domain is referred as effector binding site. This intracellular effector docking site is only able to bind P, but with the addition that P can be associated with G. That means P has two binding domains one linking to the receptor and one binding the sequenced effector G. Consequently the intracellular effectors can occur as single molecules (P, G) and as two-molecule-complex PG. Since G has no binding domain corresponding to R, no interaction occurs between the effector G and the receptor.

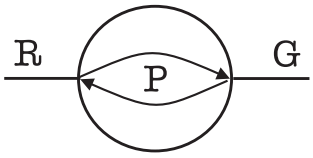

Figure 4.4: Domain interactions of effectors. The effector P possesses two domains, one for the receptor and one for the subsequent signaling molecule G. The arrows indicates that the binding of G changes the affinity of the receptor binding site.

levels. For instance the occurrence of the receptor-signal complex $Occ[RS]$ is equivalent to the occupancy of the receptor R by the signal S if no other effectors can bind to the receptor-bound S. All possible occurrence levels are (if effectors are not taken into account)

$$\begin{aligned}
z_0 &= Occ[R] \\
&= R[0,0] + R[S,0] + R[0,P] + R[S,P] + R[0,PG] + R[S,PG] \\
z_1 &= Occ[RS] \\
&= R[S,0] + R[S,P] + R[S,PG] \\
z_2 &= Occ[RP] \\
&= R[0,P] + R[S,P] + R[0,PG] + R[S,PG], \\
z_3 &= Occ[RSP] \\
&= R[S,P] + R[S,PG] \\
z_4 &= Occ[RPG] \\
&= R[0,PG] + R[S,PG] \\
z_5 &= Occ[RSPG] \\
&= R[S,PG]
\end{aligned}$$

The first state z_0 is the total sum of all complexes where the receptor molecule is involved. There is neither consumption nor production of receptor molecules. Thus we expect a vanished derivation and consequently this concentration to be constant. The state z_1 is the occurrence level of the receptor-signal complex, where the signal has bound to the first docking site of the receptor. Since the signal cannot bind to any other molecule, z_1 is equivalent

to the occupancy level $occR_1[S]$. The state z_3 is the occurrence level of the complex where the effector P has bound to the second docking site of the receptor. It is different from the occupancy $occR_2[P]$ since the specie $R[S, PG]$ is taken into account. The states z_3 and z_4 are occurrence levels referring to complexes of three molecules: z_3 has close meaning to the second order occupancy $occR_{12}[S, P]$ and sums species where S and P has bound to the receptor; z_4 is equivalent to the first order occupancy level $occR_2[PG]$. The last state z_5 is equal to the species $R[S, PG]$ and is the concentration of the fully occupied receptor having bound all subsequent effectors.

After setting up the ordinary differential equation for the above reaction scheme using mass action kinetics as described in Chapter 3, we can transform the system to macroscopic description using the above states ($T = \frac{\partial z}{\partial x}$). With $z = T \cdot c$ follows $\dot{z} = T \cdot f(c)|_{T^{-1} \cdot z}$, and the transformed differential equations are:

$$
\begin{aligned}
\dot{z_0} &= 0 \\
\dot{z_1} &= k_1 S(z_0 - z_1) - k_{-1} z_1 \\
\dot{z_2} &= k_3 P(z_1 - z_3) + k_6 PG(z_1 - z_3) + k_2 P(z_0 - z_1 - z_2 + z_3) \\
&\quad + k_5 PG(z_0 - z_1 - z_2 + z_3) - k_{-6} z_5 + k_{-3}(-z_3 + z_5) \\
&\quad + k_{-5}(-z_4 + z_5) - k_{-2}(z_2 - z_3 - z_4 + z_5) \\
\dot{z_3} &= k_1 S z_2 + k_3 P(z_1 - z_3) + k_6 PG(z_1 - z_3) - k_{-1} z_3 - k_{-3} z_3 \\
&\quad - k_1 S z_3 + k_{-3} z_5 - k_{-6} z_5 \\
\dot{z_4} &= G k_4 z_2 + k_6 PG(z_1 - z_3) + k_5 PG(z_0 - z_1 - z_2 + z_3) \\
&\quad - G k_4 z_4 - k_{-4} z_4 - k_{-5} z_4 + k_{-5} z_5 - k_{-6} z_5 \\
\dot{z_5} &= k_6 PG(z_1 - z_3) + G k_4 (z_3 - z_5) + k_1 S(z_4 - z_5) \\
&\quad - k_{-1} z_5 - k_{-4} z_5 - k_{-6} z_5
\end{aligned}
$$

The transformation has led to a decoupling of the first two states from the others. This does make sense, since we have postulated no inner interaction from the receptor or its domains to the signal binding domain. In the following we will discuss which assumptions have to be taken to reach a model reduction, due to a decoupling of all macroscopic states (z_0, z_1, z_2,

z_4) from any mesoscopic states (higher order occurrence levels z_3, z_5), and discuss the biological relevance.

If we assume that the PG complex docks to R like pure P, i.e. the association and dissociation of the effector or effector-complex is fully determined by the interacting domains. This can be interpreted as the existence of an binding domain of P responsible for the linkage to the receptor. This is a justifiable assumption, since the effector binding is rather determined by local interactions between the corresponding domains than by a global interaction of the entire receptor and the entire effector [Pawson 2004]. So the effector P can be seen as a molecule with two binding domains; one responsible for the binding to the receptor, and one for the binding of G. If the G-binding-domain should not influence the receptor-binding-domain, we have to set[1]:

$$k_5 = k_2, \quad k_{-5} = k_{-2}, \quad k_6 = k_3, \quad k_{-6} = k_{-3}$$

And the equations become:

$$\begin{aligned}
\dot{z_0} &= 0 \\
\dot{z_1} &= k_1 S(z_0 - z_1) - k_{-1} z_1 \\
\dot{z_2} &= -(k_{-2} z_2) + k_3(P + PG)(z_1 - z_3) + k_{-2} z_3 - k_{-3} z_3 \\
&\quad + k_2(P + PG)(z_0 - z_1 - z_2 + z_3) \\
\dot{z_3} &= k_3(P + PG)(z_1 - z_3) + k_1 S(z_2 - z_3) - (k_{-1} + k_{-3}) z_3 \\
\dot{z_4} &= G k_4 z_2 + k_3 PG(z_1 - z_3) + k_2 PG(z_0 - z_1 - z_2 + z_3) \\
&\quad - G k_4 z_4 - k_{-2} z_4 - k_{-4} z_4 + \mathbf{k_{-2} z_5} - \mathbf{k_{-3} z_5} \\
\dot{z_5} &= k_3 PG(z_1 - z_3) + G k_4(z_3 - z_5) + k_1 S(z_4 - z_5) \\
&\quad - k_{-1} z_5 - k_{-3} z_5 - k_{-4} z_5
\end{aligned}$$

The equation for occupancy-level-G $\dot{z_4}$ is the only one still coupled with $\dot{z_5}$, but also the one of most interest (the fully occupied domain often triggers

[1] To decouple the equations of z_2 and z_3 from z_5, it is mathematically not necessary to set the forward reaction constants for the effector binding equal ($k_5 = k_2, \quad k_6 = k_3$), since the all couplings if the system to z_5 concern backward reaction rates k_{-i} ($i = \{2, 3, 4, 5\}$).

further subsequent processes. See also Chapter 1). We can decouple it if we assume that the dissociation of the effector is independent from the signal binding site, i.e.

$$k_{-2} = k_{-3}$$

This proposal can be interpreted as follows. The binding of S leads to a conformational change of the receptors structure from the informally called "tensed conformation" to the "easy conformation", allowing the effector molecule P to bind much easier. In the above case, bound P fixes the "easy conformation", and arbitrary binding of P to the "tensed" domain forces it to the "easy" conformation. In such a process the dissociation of P would be the same by S-occupied receptor as well as by S-unoccupied receptor. Keep in mind that binding of S still has a great effect, namely the increase of effector association. This ensures that the effector binding occurs at different rates depending on the S-occupancy of the receptor.

$$\begin{aligned}
\dot{z_0} &= 0 \\
\dot{z_1} &= k_1 S(z_0 - z_1) - k_{-1} z_1 \\
\dot{z_2} &= -(k_{-2} z_2) + k_3(P + PG)(z_1 - z_3) + k_2(P + PG)(z_0 - z_1 - z_2 + z_3) \\
\dot{z_3} &= k_3(P + PG)(z_1 - z_3) + k_1 S(z_2 - z_3) - (k_{-1} + k_{-2}) z_3 \\
\dot{z_4} &= G k_4 z_2 + k_3 PG(z_1 - z_3) + k_2 PG(z_0 - z_1 - z_2 + z_3) \\
&\quad -(G k_4 + k_{-2} + k_{-4}) z_4 \\
\dot{z_5} &= k_3 PG(z_1 - z_3) + G k_4(z_3 - z_5) + k_1 S(z_4 - z_5) - (k_{-1} + k_{-2} + k_{-4}) z_5
\end{aligned}$$

Now the first five equations do not depend on z_5. And we can omit the last equation, if we are only interested in the first order occupancy levels.

Extension

In the above section we did not consider the formation of the PG complex:

$$P + G \underset{k_{-7}}{\overset{k_7}{\rightleftharpoons}} PG$$

If the effector P is considered as adaptor protein with two distinct binding domains, one responsible for linkage to the receptor, the other one for binding

of the effector G, consequently the binding from G to P is generated by domain-domain interactions, and we can assume this process independently from the receptor domain of P. Thus the corresponding reaction constants are equal, $k_5 = k_2$, $k_{-5} = k_{-2}$,$k_6 = k_3$,$k_{-6} = k_{-3}$, and the macro-state description is:

$$\begin{aligned}
\dot{z_0} &= 0 \\
\dot{z_1} &= k_1 S(z_0 - z_1) - k_{-1} z_1 \\
\dot{z_2} &= -(k_{-2} z_2) + k_3(P + z_6)(z_1 - z_3) + k_2(P + z_6)(z_0 - z_1 - z_2 + z_3) \\
\dot{z_3} &= k_3(P + z_6)(z_1 - z_3) + k_1 S(z_2 - z_3) - (k_{-1} + k_{-2}) z_3 \\
\dot{z_4} &= G k_4 z_2 + k_3 z_6(z_1 - z_3) + k_2 z_6(z_0 - z_1 - z_2 + z_3) \\
&\quad -(G k_4 + k_{-2} + k_{-4}) z_4 \\
\dot{z_5} &= k_3 z_6(z_1 - z_3) + G k_4(z_3 - z_5) + k_1 S(z_4 - z_5) - (k_{-1} + k_{-2} + k_{-4}) z_5 \\
\dot{z_6} &= G k_7 P + k_{-2} z4 - (k_{-7} + k_3(z_1 - z_3) + k_2(z_0 - z_1 - z_2 + z_3))
\end{aligned}$$

where z_6 means the concentration of the free PG complex. As you can see the first four equations do not depend on z_5 but coupled indirectly through the equation of the PG complex z_6. We can remove this coupling, if we assume the dissociation of P and PG from the receptor to be independent from the occupancy of S analogously to the previous section: $k_{-2} = k_{-3}$.

In difference to the previous model, where we did not balance the PG complex, the occupancy level of G (z_4) is not directly dependent on the fully occupied receptor (z_5). However it is now indirectly coupled through the free PG complex and to get it decoupled, the same assumptions as before have to be taken. Thus it seems to be a structural property of those effector chains, that they only can be described in a reduced way, if the subsequent binding has no effect on the dissociation of the already bound molecules. This of course makes no sense, if the dissociation of the intermediate effector is conducted by any enzyme. An example is the dephosphorylation though the phosphatase which we will discuss in the next section.

Binding of one effector due to previous phosphorylation

In this section a receptor of similar structure as above is considered. Again it has an extracellular signal binding site, and an intracellular effector binding site, however the intracellular binding domain has to be phosphorylated before effector binding G can occur. This processes can easily be described with the above equations, just setting

$$k_5 = 0,\ k_{-5} = 0,\ k_6 = 0,\ k_{-6} = 0$$

This means docking of the effector molecule G is only possible, if the corresponding binding site is occupied by the effector P (matches phosphorylation) through prior binding. It is not possible that the single effector G is phosphorylated and able to bind at the unphosphorylated receptor.

Through observability analysis it was found that no transformation is possible that would result in any exact model reduction. However, to get a better understanding of the phosphorylation process let us have a closer look at the equations, and assume that dephosphorylation (dissociation of P from the receptor) might occur also if G is bound, i.e.

$$k_5 = 0,\ k_{-5} > 0,\ k_6 = 0,\ k_{-6} > 0$$

In that case the equations are:

$$\begin{aligned}
\dot{z_0} &= 0 \\
\dot{z_1} &= k_1 S(z_0 - z_1) - k_{-1} z_1 \\
\dot{z_2} &= k_3 P(z_1 - z_3) + k_2 P(z_0 - z_1 - z_2 + z_3) - k_{-6} z_5 \\
&\quad + k_{-3}(-z_3 + z_5) - k_{-2}(z_2 - z_3 - z_4 + z_5) + k_{-5}(-z_4 + z_5) \\
\dot{z_3} &= k_3 P(z_1 - z_3) + k_1 S(z_2 - z_3) - k_{-1} z_3 - k_{-3} z_3 + k_{-3} z_5 - k_{-6} z_5 \\
\dot{z_4} &= G k_4 (z_2 - z_4) - k_{-4} z_4 - k_{-5} z_4 + k_{-5} z_5 - k_{-6} z_5 \\
\dot{z_5} &= G k_4 (z_3 - z_5) + k_1 S(z_4 - z_5) - (k_{-1} + k_{-4} + k_{-6}) z_5
\end{aligned}$$

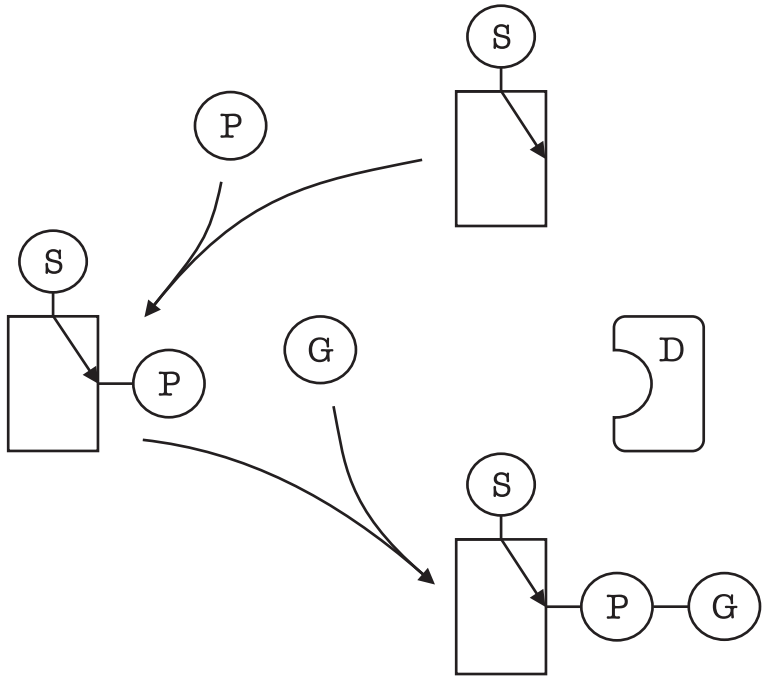

Figure 4.5: Monomer with one phosphorylation binding site and the effector G. The effector can bind to the receptor only if phosphorylated previously. Dephosphorylation (removing of P) is performed through the phosphatase D. Consequently occupancy of P by G has great effect on dephosphorylation, since it hampers the phosphatase's access.

We again consider the case $k_{-2} = k_{-3}$ to aim in decoupling of $\dot{z_2}$

$$\begin{aligned}
\dot{z_0} &= 0 \\
\dot{z_1} &= k_1 S(z_0 - z_1) - k_{-1} z_1 \\
\dot{z_2} &= k_3 P(z_1 - z_3) + k_2 P(z_0 - z_1 - z_2 + z_3) \\
&\quad -k_{-2}(z_2 - z_4) - k_{-5}(z_4 - z_5) - k_{-6} z_5 \\
\dot{z_3} &= k_3 P(z_1 - z_3) + k_1 S(z_2 - z_3) - k_{-1} z_3 - k_{-2}(z_3 - z_5) - k_{-6} z_5 \\
\dot{z_4} &= G k_4 (z_2 - z_4) - k_{-4} z_4 - k_{-5}(z_4 - z_5) - k_{-6} z_5 \\
\dot{z_5} &= G k_4 (z_3 - z_5) + k_1 S(z_4 - z_5) - (k_{-1} + k_{-4} + k_{-6}) z_5
\end{aligned}$$

We can decouple $\dot{z_2}$ and $\dot{z_3}$ from z_5 if and only if

$$k_{-5} = k_{-6}$$

and $\dot{z_3}$ if and only if

$$k_{-2} = k_{-6}$$

The first constraint is not such a strong one and means that dephosphorylation at the PG complex is independent from the signal binding site (consider the phosphatase as nearly independent from the receptors conformation). The second constraint is hardly to fulfill since the dephosphorylation of the unoccupied site should be much faster than the occupied one, thinking that in this case the phosphatase than can reach the phosphorylated domain much easier.

Approximations

Since an exact reduction is only possible under strong assumptions about association and dissociation constants, resulting in a biologically unreasonable model, we will discuss two possible approximations in the following.

(1) If we use the first constraint from above, namely the phosphatase of the occupied effector site is independent from the signal S ($k_{-2} = k_{-3}$ and $k_{-5} = k_{-6}$), and decouple our system just by neglecting the coupling term as implied in the work of [Borisov et al. 2005] (see also chapter 2.3). All

equations stay the same and the only error occurs in $\dot{z_3}$ with a deviation of $(k_{-2} - k_{-2})z_5$.

$$\dot{z_3} = k_3P(z_1 - z_3) + k_1S(z_2 - z_3) - k_{-1}z_3 - k_{-2}z_3 + (\mathbf{k_{-2}} - \mathbf{k_{-5}})\mathbf{z_5}$$

If we go back to the definitions of the occupancy levels z_i we can find some relations. First $z_5 = R[S, PG]$ is smaller than $z_3 = R[S, P]+R[S, PG]$, which miners the influence of all z_5 terms. Now we can compare $(k_{-2} - k_{-5})z_5$ with $(k_{-1} + k_{-2})z_3$ if we set $z_5 = z_3$ as upper limit. We can neglect the coupling term z_5 with making a little fault if $\mid k_{-2} - k_{-5} \mid \ll \mid k_{-1} + k_{-2} \mid$.

(2) If we neglect the process of effector docking in the first step, the ode's for the receptor and its phosphorylation in species coordinates are:

$$\begin{aligned}
\frac{d}{dt}R[0,0] &= -(k_2PR[0,0]) + k_{-2}R[0,P] + k_{-1}R[S,0] - k_1RS \\
\frac{d}{dt}R[S,0] &= -(k_{-1}R[S,0]) - k_3PR[S,0] + k_{-3}R[S,P] + k_1RS \\
\frac{d}{dt}R[0,P] &= k_2PR - k_{-2}R[0,P] + k_{-1}R[S,P] - k_1R[0,P]S \\
\frac{d}{dt}R[S,P] &= k_3PR[S,0] - k_{-1}R[S,P] - k_{-3}R[S,P] + k_1R[0,P]S
\end{aligned}$$

Transformed to occupancy space this gives:

$$\begin{aligned}
z\dot{e}_0 &= 0 \\
z\dot{e}_1 &= k_1S(ze_0 - ze_1) - k_{-1}ze_1 \\
z\dot{e}_2 &= -(k_{-2}ze_2) + k_3P(ze_1 - ze_3) + k_{-2}ze_3 - k_{-3}ze_3 \\
&\quad + k_2P(ze_0 - ze_1 - ze_2 + ze_3) \\
z\dot{e}_3 &= k_3P(ze_1 - ze_3) + k_1S(ze_2 - ze_3) - (k_{-1} + k_{-3})ze_3
\end{aligned}$$

In the second step, we consider the G binding separately as a pseudo module as illustrated in Figure 4.6. It is described by the differential equation:

$$\frac{d}{dt}P_{pseudo}[G] = k_4GP_{pseudo}[0] - k_{-4}P_{pseudo}[G]$$

As a constraint, the total concentration of the pseudo-module must equal $R_2[P]$, it follows $P_{pseudo}[0] = R_2[P] - P_{pseudo}[G]$. This links the two modules

and the occupancy of G in the pseudo-module $P_{pseudo,1}[G]$ becomes the fourth state of the total system ze_4:

$$\frac{d}{dt}ze_4 = Gk_4ze_2 - (Gk_4 + k_{-4})ze_4$$

Setting the dephosphorylation independent from the signal binding again ($k_{-2} = k_{-3}$ and $k_{-5} = k_{-6}$), a comparison to the correct model by building the difference gives:

$$\begin{aligned}
z\dot{e}_0 - \dot{z_0} &= 0 \\
z\dot{e}_1 - \dot{z_1} &= 0 \\
z\dot{e}_2 - \dot{z_2} &= (k_{-2} - k_{-5})z_4 \\
z\dot{e}_3 - \dot{z_3} &= (k_{-2} - k_{-5})z_5 \\
z\dot{e}_4 - \dot{z_4} &= -k_{-5}z_4
\end{aligned}$$

In contrast to the first method, the dephosphorylation is described approximately (separation P and PG is summarized in one reaction). Therefore we set $k_{-5} = 0$ to compare it accurately to the first approximate method. The error deviates become $k_{-2}z_4$ for $z\dot{e}_2 - \dot{z_2}$ and $k_{-2}z_5$ for $z\dot{e}_3 - \dot{z_3}$. The error according to z_5 in the Equation $z\dot{e}_3 - \dot{z_3}$ is obvious, since in this method the state z_5 does not even exist. The error in z_4 is due to the method inherent decoupling of the first four equations to the last equation (module R considered independent from pseudo module P).

equation	method (1)	method (2)
$z\dot{e}_2$	$-k_{-2}z_5$	$k_{-2}z_4$
$z\dot{e}_3$	0	$k_{-3}z_5$

Table 4.1: Comparison of the error in the differential equations for the approximative methods (1) and (2). The dephosphorylation is assumed to be independent from the signal binding domain ($k_{-2} = k_{-3}$) and not occurring on occupied effector sites ($k_{-5} = k_{-6} = 0$)

4.1.3 Notes on the effector description

In all above models the pure effector concentrations are not modeled and so have been considered as known time-varying values. In this case we can

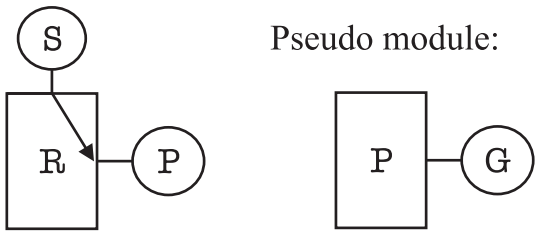

Figure 4.6: Pseudo-modules of chain formation. For the approximative method (2) the system is participated in two modules, which are modeled separately. They are linked together through the occupancy $R[P]$ which has to equal the total concentration of P_{pseudo}, since G can only bind to P if it is associated with the receptor. The pseudo module can be interpreted as the phosphorylated receptor.

regard them as inputs for the modules. Often however, the effector concentrations remain constant, either because of the high concentration of the effector in relation to the amounts engaged by the receptor or because of regulation mechanism of the cell. A good example are the second messengers ATP/ADP/AMP, which cause phosphorylation of specific Ser/Ter residues (phosphorylation sites), and whose concentrations are highly regulated. Consequently these concentrations are often considered constant for all processes that do not concern their regulation.

it is however possible, to extend the above equations and to consider the effectors as balanced variables without loosing the presented results on reduction. The above equations still hold, however we get additional differential equations for the free effector concentrations (see appendix). Note that the effector concentrations are not integrated in the concept of occupancy levels, but can be considered as additional microscopic equations.

If we assume no external production or consumption, and the effector is domain specific, which means that each effector only participates reactions downstream one single domain, we have the conserved moieties:

$$E_i + occR[E_i] = E_{i,total} = \text{const.}$$

Solving this for E_i and set in the equations of macro state description, it does not change the discussed modular structure.

The concept of occurrence levels however does not exclude non-receptor related species, and no divergency from the formal modeling process has to be made, to achieve a suitable mathematical description of the system. Further it reveals the above mentioned conserved moieties after its application, even in more complex cases. This is shown exemplarily in the appendix, where both here discussed modeling methods are compared in more detail.

4.1.4 Summary and conclusions

In this section subsequent binding events occurring at one binding site building a chain of effectors have been analyzed. Observability analysis showed, that with the use of occupancy levels no model reduction is possible. However a the system's inherent modular structure was expected. Further analysis showed, that with the definition of a more general concept of mesoscopic states, the levels of occurrence, a model reduction could be achieved under two preconditions:

1. **no backtracking effects:** All dependencies within the system are directed downstream. These assumption is for instance violated, if the release or undocking of an effector is performed by an enzyme, as for instance dephosphorylation by phosphatases. (Figure 4.5)

2. **dependencies effect only forward reactions** (docking) but not backwards reactions (release) of effector binding. This becomes clear if one dissects the reversible reaction of effector binding in two irreversible reactions. The effector docking runs with two different reaction constants to take into account that the signal binding influences the effector binding:

$$\begin{aligned} R + P &\xrightarrow{k_1} RP \\ RS + P &\xrightarrow{k_2} RSP \end{aligned}$$

However the release of the effector runs with the same reaction constant

for both cases (signal and no signal):

$$\begin{aligned} RP &\xrightarrow{k-} R + P \\ RSP &\xrightarrow{k-} RS + P \end{aligned}$$

As a further result of the performed analysis several conclusions can be drawn. First modeling of chain or cascade reactions as for instance modeled in the above sections should follow some suggestions, to ease the system's analysis. Whenever reasonable, binding site dependencies should only effect the forward reaction (the engaging process) but not the backward reaction (the releasing process), because this could lead to a less complex model under certain circumstances as exemplified in Section 4.1.2. Even though this restriction seems to be a strong one, it allows to constitute the equilibrium ($\frac{k_{forward}}{k_{backward}}$) as well as the time-frame ($k_{forward}$ or $k_{backward}$) to reach this equilibrium, since only the proportion of the reaction constants is specified, not the absolute values.

Second, from a systems theoretical point of view the levels of occurrence should be favored over the levels of occupancy, not only as more general concept, but also because of further modularization and model-reduction feasibilities. They provide a general framework of macroscopic modeling that is not restricted to any kind of carrier molecule as is the concept of occupancy levels. The levels of occupancy may be the more biological relevant in some cases, but they can be defined as linear output of an occurrence level model and can be determined easily by substraction. For example in the above models: $occR_2[P] = Occ[RP] - Occ[RPG]$. Summarizing, the levels of occurrence provide a more suitable framework for modeling domain specific signaling processes than the former defined levels of occupancy.

4.2 Dimers

In this section, we will set up a detailed micro-description model of receptors that are able to form a dimeric complexes. We start with a model possessing

all combinatorial possible species, symmetric and non-symmetric. It is shown how the levels of occupancy are used to modularize and reduce the model of the dimerization process, and how the model reduction can be extended by lumping the symmetric species and applying the effector occupancy levels. Further it is shown, that the levels of occurrence results in the same modularization and model reduction directly.

For receptors, that are able to dimerize, the extracellular signal binding brings the monomers together or stabilizes such allocalization. A good example is the EGF receptor (Example 2.2 on page 34). This dimerization proceeds the crossover phosphorylation of the intracellular binding domains, and such performs the acceleration of intracellular effector bindings [Ferguson et al. 2003; Garrett et al. 2002; Ogiso et al. 2002; Jorissen et al. 2003; Schlessinger 2000].

The presence of receptor dimers have already been considered ([Borisov et al. 2005] supplementary material), however under very strong assumptions rather allowing the application of the occupancy levels on a dimerized receptor and modeling the dimerization process separately. This includes the two following simplifications: The monomeric receptor have to bind its signal ligand first, before it can bind to a second monomeric receptor (also having engaged a signal molecule) to form a dimeric complex. Only receptor dimers are able to engage downstream effector molecules (see Figure 2.4).

However these are strong simplifications, since following reasonable scenarios are not possible:

- Dimerization occurs also if only one monomer has engaged its signal molecule
- Occurrence of spontaneous dimerization, which is then stabilized by signal binding
- Engagement of downstream effectors of monomeric receptors also, either after signal binding or spontaneously.

One task of this thesis was to take these possible scenarios into account and to start with a reaction scheme that includes all possible signaling events.

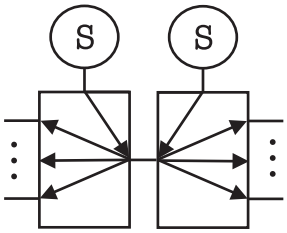

Figure 4.7: Receptor dimer. Signal binding accelerates the dimerization process, which subsequently effects the binding of downstreaming effectors on the intracellular domains. No other dependencies are assumed to exist.

Consequently effector binding on all domains is possible for dimeric and monomeric receptors, even if the signal is not bound, and dimerization is possible between all monomer-species [Jorissen et al. 2003; Jiang and Sorkin 2002]. To take the effects of dimerization as described at the beginning of this chapter into account, signal binding of S on the receptor is assumed to be independent from the state of the receptors domains, dimerization is assumed to be dependent on the occupancy of the extracellular signal docking site, and intracellular effector binding or phosphorylation is assumed to be dependent on the dimeric/monomeric status of the receptors (see Figure 4.7). With these preconditions two general processes are discussed in the following two sections. First the process of dimerization itself, second the binding of intracellular effectors on distinct independent domains.

Signal binding

The signal binding is assumed to be independent from the state of the receptor and its domains. Thus all reactions concerning the signal binding are modeled with the same reaction constant k_1.

$$
\begin{aligned}
R[0,\ldots]S &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} R[S,\ldots] \\
DM[R[0,\ldots],R[\ldots]]S &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} DM[R[S,\ldots],R[\ldots]] \\
DM[R[\ldots],R[0,\ldots]]S &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} DM[R[\ldots],R[S,\ldots]]
\end{aligned}
$$

Symmetric species, as for instance $DM[R[S,\ldots],R[0,\ldots]]$ and $DM[R[0,\ldots],R[S,\ldots]]$ are considered separately, which simplifies the derivation of the equations. However, since these symmetric species are biologically not distinguishable, one could lump two symmetric species. Then only one variable would represent both (theoretical) cases. Consequently, the reaction constant describing a binding process must be multiplied by the number of binding sites, where this reaction might occur. Note that the reverse reaction rate remains unchanged, since the molecule cleavage is only possible at the binding site where it is attached. However, reversed reaction rates of lumped symmetric species having bound multitudes of the same molecule must be treated analogous.

Example 9. *Consider for instance the lumped specie, that represents the dimer receptors having bound the signal S on either one of its monomer parts:*

$$\widetilde{DM}[R[0,\ldots],R[S,\ldots]] = DM[R[0,\ldots],R[S,\ldots]] + DM[R[S,\ldots],R[0,\ldots]]$$

It is the product product of the signal binding to $DM[R[0,\ldots],R[0,\ldots]]$. To take into account, that the signal molecule S can bind to two receptors signal docking sites, the reaction rate must be multiplied by 2.

$$\widetilde{DM}[R[0,\ldots],R[0,\ldots]] + S \underset{k_{-1}}{\overset{2k_1}{\rightleftharpoons}} DM[R[\ldots],R[S,\ldots]]$$

This can be shown mathematically. The reaction scheme for the above mention signal binding is

$$DM[R[0,\ldots],R[0,\ldots]]S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} DM[R[S,\ldots],R[0,\ldots]]$$

$$DM[R[0,\ldots],R[0,\ldots]]S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} DM[R[0,\ldots],R[S,\ldots]]$$

Using mass action kinetics, the corresponding reaction rates are

$$\begin{aligned}
\frac{d}{dt}DM[R[S,\ldots],R[0,\ldots]] &= k_1 DM[R[0,\ldots],R[0,\ldots]]S \\
&\quad -k_{-1}DM[R[S,\ldots],R[0,\ldots]] \\
\frac{d}{dt}DM[R[0,\ldots],R[S,\ldots]] &= k_1 DM[R[0,\ldots],R[0,\ldots]]S \\
&\quad -k_{-1}DM[R[0,\ldots],R[S,\ldots]]
\end{aligned}$$

Adding both equations gives

$$\begin{aligned}\frac{d}{dt}\widetilde{DM}[R[0,\ldots],R[S,\ldots]] &= 2k_1 DM[R[0,\ldots],R[0,\ldots]]S \\ &\quad -k_{-1}\widetilde{DM}[R[0,\ldots],R[S,\ldots]]\end{aligned}$$

The model of dimerization

Since the dimerization is dependent on the controlling domain, its reactions can be structured in three groups with different reaction constants, corresponding to the signal-occupancy of the binding partners.

1. Neither of the two receptor species has bound a signal molecule on its controlling domain. The dimerization is very unlikely. Consequently the corresponding reaction rate is small, which is modeled with the reaction constant k_2.

$$R[0,...]R[0,...] \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} DM[R[0,...],R[0,...]]$$

2. One of the binding partners has bound a signal, the other has not. The reaction rate is significantly altered compared to Case 1 ($k_3 \neq k_2$).

$$R[S,...]R[0,...] \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} DM[R[S,...],R[0,...]]$$

3. Both dimerization partners have bound the signal, the reaction rate is high ($k_4 \neq k_3 \neq k_2$).

$$R[S,...]R[S,...] \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} DM[R[S,...],R[S,...]]$$

This implements the dimerization process in a signal dependent manner, the more signal has been received by the receptor, the faster is its dimerization $k_2 < k_3 < k_4$. It is to note that for the results achieved in this thesis it is not necessary to make any assumptions concerning the actual value of the reaction constants, however it is presented here that way to give a reasonable model of dimerization.

There are reactions between two different monomers, that can form two formal different species, which are not distinguishable biologically, leading to a lumping problem similar to that of the signal binding. To be thorough, the consequences are presented here briefly. Consider reactions of the form:

$$R[\ldots,E_i,\ldots]R[\ldots,E_j,\ldots] \underset{k_-}{\overset{k_+}{\rightleftharpoons}} DM[R[\ldots,E_i,\ldots],R[\ldots,E_j,\ldots]]$$

$$R[\ldots,E_j,\ldots]R[\ldots,E_i,\ldots] \underset{k_-}{\overset{k_+}{\rightleftharpoons}} DM[R[\ldots,E_j,\ldots],R[\ldots,E_i,\ldots]]$$

If these species are lumped together,

$$\begin{aligned}\widetilde{DM}[R[\ldots,E_i,\ldots],R[\ldots,E_j,\ldots]] \;&=\; DM[R[\ldots,E_i,\ldots],R[\ldots,E_j,\ldots]] + \\ &\quad\; DM[R[\ldots,E_j,\ldots],R[\ldots,E_i,\ldots]]\end{aligned}$$

the probability of a reaction between two different monomers will be double as high as the probability of a reaction between two equal monomers.

$$R[\ldots,E_i,\ldots]R[\ldots,E_j,\ldots] \underset{k_-}{\overset{2k_+}{\rightleftharpoons}} \widetilde{DM}[R[\ldots,E_i,\ldots],R[\ldots,E_j,\ldots]]$$

This can be shown by adding both reaction rates (analogous to Example 9).

4.2.1 Modularization of the dimerization process

To understand the process of dimer-formation itself, it is sufficient to consider the simplest case of a receptor that is able to dimerize. Let us consider a dimeric receptor as you see in Figure 4.8 with two distinct binding domains: one extracellular for the signal S and one intracellular for an effector P. Binding of the signal molecule S accelerates dimer formation whereas dimer configuration causes high phosphorylation rates (binding of P). No other domain dependencies are preconditioned to exist. We consider all asymmetric dimer species like $DM[R[S,0],R[0,0]]$ and $DM[R[0,0],R[S,0]]$ separately. In that case, we need 20 species to describe the system, 2^4 dimers and 4 monomers but have only 6 different reaction constants. The similarities of the reactions lead to modularization of the system of ordinary differential equations in macroscopic coordinates as they are chosen here, and enables us to perform a model reduction by neglecting the internal dynamics.

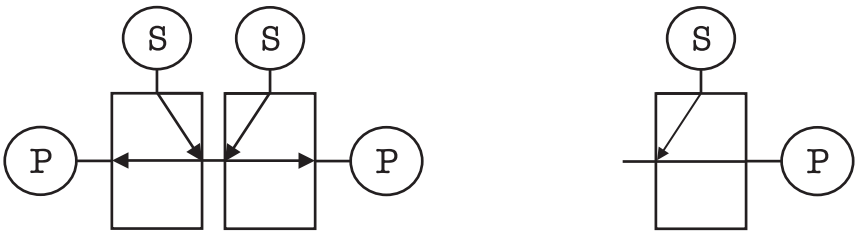

Figure 4.8: Schematic representation of the receptor molecules involved in dimerization process (described above). The extracellular binding site senses the signal S through binding, which effects the dimerization. The intracellular binding site is able to bind one effector P and effected by these dimerization. Monomeric and dimeric receptors have to be considered.

To decompose the system in three modules we consider the dimeric receptor as a molecule with four distinct binding domains, and treat symmetric dimeric complexes like DM[R[S, 0], R[0, 0]] and DM[R[0, 0], R[S, 0]] as different and distinguishable states. The model is set up as described in the beginning of this chapter taking all possible species and reactions into account, applying the law of mass action and leading to a microscopic description consisting of ordinary differential equations for the dynamics of all 20 species.

The transformation on macroscopic and mesoscopic states follows the levels of occupancy, where the dimeric receptor is treated like a monomeric receptor with four distinct binding domains. This means for instance, that we have two different but symmetric occupancies of P, one for the docking site of the left receptor half $occDM_2[P] = DM[R[*,P],R[*,*]]$, and one for the right $occDM_4 = DM[R[*,*],R[*,P]]$[2].

Such transformation leads to a modularization of the system in mesoscopic coordinates by the hierarchical decoupling of three sets of equations. Each module describes a particular process of the receptors signal transduction.

The first block describes signal docking and dimerization with two equa-

[2] $occDM_i[E_k]$ denotes the occupancy level of the receptor dimer, having bound the effector E_k on the i-th docking site. Therefore, the docking sites of the dimer are counted from left to right: $DM[R[\overset{1}{*},\ldots,\overset{b_i}{*}],R[\overset{b_i+1}{*},\ldots,\overset{2b_i}{*}]]$ (see also definitions in Chapter 3)

tions for the occupancy levels of the monomeric receptor and four equations for the occupancy levels of the dimeric receptor.

$$\begin{aligned} R[*,*] &= occR[] \\ R[S,*] &= occR_1[S] \\ DM[R[*,*],R[*,*]] &= occDM[] \\ DM[R[S,*],R[*,*]] &= occDM_1[S] \\ DM[R[*,*],R[S,*]] &= occDM_3[S] \\ DM[R[S,*],R[S,*]] &= occDM_{1,3}[S,S] \end{aligned}$$

The asterisks denote the binding sites which are not specified, and the sum over this effectors are build. So $R[S,*]$ for instance is the sum of all receptor monomers having bound S. The state $R[*,*]$ presents the 0st-order occupancy for the monomer, where no binding site is specified, as does $DM[R[*,*],R[*,*]]$ for the dimer. Then we have the 1st-order occupancies of one bound signal $R[S,*]$ for the monomer and $DM[R[*,*],R[S,*]]$, $DM[R[S,*],R[*,*]]$ for the dimer. As can be seen the information necessary for this module only corresponds to the signal molecule S. There is no information about the level of bound P necessary.
We can define the level of total bound S as output for this module just as

$$y_1 = R[S] = DM[R[*,*],R[S,*]] + DM[R[S,*],R[*,*]]$$

where all molecules carrying one S appear ones and all molecules carrying two S appear twice. This output presents a macroscopic value equivalent to the total S-R complex occurrence $O[SR]$ within the system and can be interpreted as the signal strength of the signal S effecting the receptor.

The second module describes the binding of P through additional ten

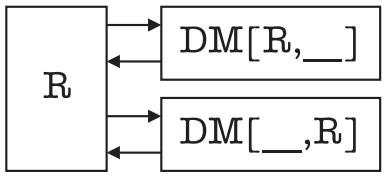

Figure 4.9: Schematic representation of the signal routing of the P binding process. Macro states: R monomers, $DM[*,R]$ right side of the dimer, $DM[R,*]$ left side of the dimer.

equations.

$$\begin{aligned}
DM[R[*,*],R[*,P]] &= occDM_4[P] \\
DM[R[S,*],R[*,P]] &= occDM_{1,4}[S,P] \\
DM[R[*,*],R[S,P]] &= occDM_{3,4}[S,P] \\
DM[R[S,*],R[S,P]] &= occDM_{1,3,4}[S,S,P] \\
R[*,P] &= occR_2[P] \\
R[S,P] &= occR_{1,2}[SP] \\
DM[R[*,P],R[*,*]] &= occDM_2[P] \\
DM[R[S,P],R[*,*]] &= occDM_{1,2}[S,P] \\
DM[R[*,P],R[S,*]] &= occDM_{2,3}[P,S] \\
DM[R[S,P],R[S,*]] &= occDM_{1,2,3}[S,P,S]
\end{aligned}$$

The necessary information in this block constitutes the levels of bound S in combinations with bound P for each receptor half (S-P correlation), but not the double P levels. This means exemplarily that for the effector binding of the first receptor in the dimer it his not necessary to know wether the effector domain of the second receptor in the dimer is occupied or not. Interestingly the phosphorylation or P-binding process of either receptor half is just indirectly coupled to the other one through the monomer states as illustrated in Figure 4.9. Analogous to the first block we define the total level of bound P as output concentration.

$$y_2 = Occ[RP] = DM[R[*,*],R[*,P]] + DM[R[*,P],R[*,*]] + R[*,P]$$

The last block consists of the three states with double P occupancy levels. It is not shown here in detail, because no biological relevant processes are described by them. Further, within the scope of the discussed model reduction method, these equations are neglected.

Lumping symmetric states

The distinction of the symmetric dimer species is only of theoretical interest, since each pair has the same biological effect, and further, cannot be distinguished by state-of-the-art measurement techniques. Thus lumping these concentrations will result in a dimensionally smaller system meeting the desired requirements on the mathematical description. The new macroscopic and mesoscopic states are:

$$\begin{aligned}
\tilde{z}_0 &= R[*,*] \\
\tilde{z}_1 &= R[S,*] \\
\tilde{z}_2 &= DM[R[*,*],R[*,*]] \\
\tilde{z}_3 &= DM[R[S,*],R[*,*]] + DM[R[*,*],R[S,*]] \\
\tilde{z}_4 &= DM[R[S,*],R[S,*]] \\
\tilde{z}_5 &= R[*,P] \\
\tilde{z}_6 &= R[S,P] \\
\tilde{z}_7 &= DM[R[*,P],R[*,*]] + DM[R[*,*],R[*,P]] \\
\tilde{z}_8 &= DM[R[S,P],R[*,*]] + DM[R[*,*],R[S,P]] \\
\tilde{z}_9 &= DM[R[S,*],R[*,P]] + DM[R[*,P],R[S,*]] \\
\tilde{z}_{10} &= DM[R[S,P],R[S,*]] + DM[R[S,*],R[S,P]]
\end{aligned}$$

They present the effector occupancy levels of monomeric and dimeric receptors as defined in Chapter 3.3 (Definition 7 on page 56). This makes sense since the binding domains of either receptor part in a dimer cannot be distinguished, and thus, its corresponding effector levels are lumped. The module $z_0, \ldots, z_4$ can be decoupled further, if we introduce the total amount of receptor R (monomeric and dimeric) and the total level of bound S instead of

the monomer concentrations. For consistency we use the total level of bound P and SP as well, i.e.

$$\begin{aligned} z_0 &= Occ[R] = \tilde{z}_0 + 2\tilde{z}_3 \\ z_1 &= Occ[RS] = \tilde{z}_1 + \tilde{z}_4 \\ z_5 &= Occ[RP] = \tilde{z}_5 + \tilde{z}_7 \\ z_6 &= Occ[RSP] = \tilde{z}_6 + \tilde{z}_8 \end{aligned}$$

These concentrations are the occurrence levels of the receptor-effector complexes as we have defined in Chapter 3 (Definition 8 on page 58). The state z_0 is the occurrence level of the receptor, where all dimeric receptor species have to be counted twice. Building the sum of occupancy levels $z_1 + z_2$ and $z_5 + z_7$ results in the occurrence levels of the receptor-effector complexes $Occ[RS]$ and $Occ[RP]$ respectively. Thereby, the dimeric receptors having bound the same effector twice are counted twice, since these species occur in both occupancy levels (for the docking site of the first and second receptor-monomer of the dimer). The states z_2 to z_4 and z_7 to z_{10} present the occurrence levels of the receptor-receptor complexes already. Now we have four modules possessing a hierarchical structure like sketched in Figure 4.10 and describing the following processes:

- $\{z_0, z_1\}$ signal docking
- $\{z_2, z_3, z_4\}$ dimerization
- $\{z_5, \ldots, z_{10}\}$ effector binding (phosphorylation)
- $\{z_{11}, z_{12}, z_{13}\}$ double effector levels

We can solve each block independently from the followings ,and thus, just use as input for the next. The most biological relevant states are the second order occurrence levels, since they present the total amount of signal bound to the receptor (incoming signal strength), the total amount of receptor dimers (dimerization status) and the total amount of by the receptor engaged downstreaming effector (outgoing signal strength). To solve

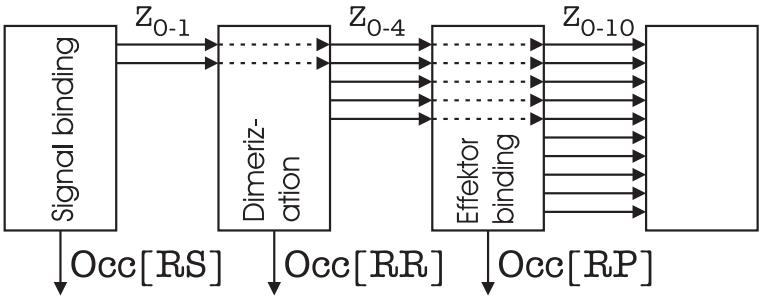

Figure 4.10: Schematic representation of the signal routing processed by symmetric dimer receptors. After application of the occurrence levels, the set of ordinary differential equations has a modular structure analogous to the receptors dependency pattern. $Occ[RS]$, $Occ[RR]$, $Occ[RP]$ total level of bound signal S, dimerized receptor and bound P (phosphorylation).

this block, the last block composed of the third order occurrence levels of the receptor-effector-effector complexes is not necessary and can be omitted, resulting in a model reduction of three ordinary differential equations. Like we have already seen in the discussion of the asymmetric model, the effector binding site of the first receptor in the dimer-complex does not demand any information about the status of effector-occupancy of the second receptor and vice versa.

The first block, describing the extracellular signal binding, consists of two states: The total concentration of receptors in the system $Occ[R]$, which is a conserved moiety, and the concentration of receptors having bound the signal molecule $Occ[RS]$. No information about any intracellular domains and their effectors or the dimerization is necessary. The second block contains the information about the dimerization status $Occ[RR]$ and the dimer-signal complexes $Occ[RRS]$ and $Occ[RRSS]$ and describes the dimerization process. Therefore, no information about the downstreaming intracellular effectors is necessary. Finally, the third block describes the intracellular effector binding. It includes the receptor-signal-effector occurrence levels and the dimer-signal-effector occurrence levels, but no information about double-effector complexes is necessary.

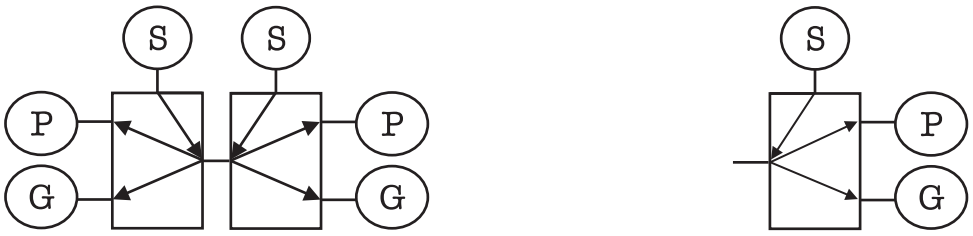

Figure 4.11: Schematic representation of the receptor molecules involved in dimerization process for a receptor with two effector binding sites.

4.2.2 Decoupling distinct binding sites

For monomeric receptors having independent binding domains or one controlling domain macroscopic description leads to decoupling of distinct effector binding sites (see [Borisov et al. 2005; Conzelmann et al. 2006]). In this section it is shown that macroscopic description using receptor-occurrence-levels as above provides analogous decoupling for dimeric receptors with one controlling domain. As described previously in section 4.2.1 the controlling domain influences the dimerization process, which in sequence influences the effector binding. The general concept is exemplarily shown at the simplest possible example of a dimeric receptor with two intracellular effector docking sites. The results can be easily extended to receptors having multiple effector binding domains, applying the occurrence-levels to all docking sites analogously.

Let us consider a dimer-receptor with two distinct binding sites each able to bind a different effector (see Figure 4.11). This leads to a combinatorial complexity of 44 species. Through macro state description, we can decompose this system in smaller modules and reach a reduction of 28 equations without any error in the output. Like in the section above, we have the same equations describing the dimerization (block one) and the signal binding (block two). Additionally, we have two blocks describing the effector binding of P and G separately (see figure 4.12).

More generally, we can state that to each effector domain with distinct effector there corresponds a set of six additional equations describing the

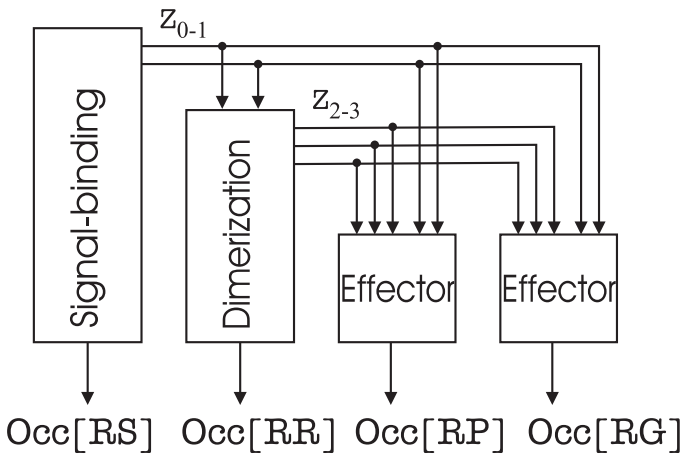

Figure 4.12: Schematic representation of the signal routing processed by symmetric dimer receptors with two distinct effector binding domains. $Occ[R]$, $Occ[RS]$, $Occ[RP]$, $Occ[RPG]$ total level of receptor R, bound signal S and bound effector P and G respectively.

effector's binding. These six states are all levels of occurrence of a single-effector-receptor complex, including all possible distinguishable combinations of the single effector and the signal. Crossover or miscellaneous occupancy combinations of the effector between the two receptor-monomers of the dimer or even between distinct domains of one receptor are not necessary.

Summarizing there are two states $Occ[R]$,$Occ[RS]$ describing the signal docking, three additional states $Occ[RR]$, $Occ[RRS]$, $Occ[RRSS]$ corresponding to the process of dimerization and further for each effector binding site six additional states describing the binding of the domain specific effector, indeed independently from the others. The remaining set of miscellaneous and double occupancy levels can be omitted since the above mentioned blocks are decoupled from them and we are only interested in the outputs, namely the occurrence levels of the receptor-effector complexes. In general, this presents a high level and exact reduction increasing exponentially in degree with the number of binding sites and the number of effectors.

4.2.3 Summary and conclusions

In this section, aspects of dimeric receptors have been analyzed, revealing a hierarchical modular structure of the dimerization process and confirming the dissection of independent binding sites by macroscopic states. Even though occupancy levels are suitable to achieve modular model structures, further model reduction could be reached by the use of occurrence levels due to the fact that they deal with dimer-inherent symmetries efficiently and lump biological indistinguishable species together. Despite this advantage they do not fail to decouple signaling events at distinct independent binding sites.

The exact definition and application of occupancy levels for such complex processes had not been clear as I started this work, and an iterative process of modeling, macroscopic states definition and application was necessary to solve the symmetry problem. This finally resulted in the two different occupancy level definitions presented in Chapter 3.3. This strongly supports the statement, that in addition to the advantages given in Section 4.1.4, the levels of occurrence should further be the preferred method as soon as symmetric processes, as for instance dimer formations, are involved.

Further, since occupancy levels refer to a carrier molecule, their application to a system that considers monomeric and dimeric receptors is not intuitive. As we have seen they can be applied to monomeric and dimeric receptors separately, and then fused together. The concept of occurrence levels however, deals perfectly with different carrier molecules and circumvents such special treatment. It can be applied directly straightforward to the modeled micro-description system. The detour made in its application above, involving the occupancy levels, is not necessary, but was outlined here in order to show the close relationship between the two definitions and to make the advantages of the levels of occurrence clear.

Chapter 5

Outline

Domain oriented modeling is a holistic approach in modeling and analyzing signal transduction pathways, that integrates qualitative biological knowledge and full combinatorial complexity of signaling pathways in terms of a detailed mathematical model and a systematic model reduction method (Figure 5.1). As has been shown, it is capable of revealing system inherent structures and dissect complex systems into smaller modules in a straightforward derivation process.

In this work we have generalized former modeling methods to cover all kind of signaling pathways and applied it on two biological relevant domain-specific processes, namely signaling-protein assembly and phosphorylation, as well as to receptors capable of dimerization. The results showed that exact systematic model reduction is possible in all analyzed cases possessing a parallel structure. Further analysis of other signaling events, as for example kinase activation and signals merging, will show if with the introduced levels of occurrence modularization or even model reduction is possible.

The presented concept of domain oriented modeling using the here first defined occurrence levels present a general concept of modeling and model reduction for biological signaling networks. In advantage to former macroscopic states, the occurrence levels are not restricted to any kind of carrier molecule, and thus applicable to any kind of biological reaction system. They possess higher reduction feasibilities and reveal system inherent structures as well as conserved moieties more efficient than the earlier definition of occu-

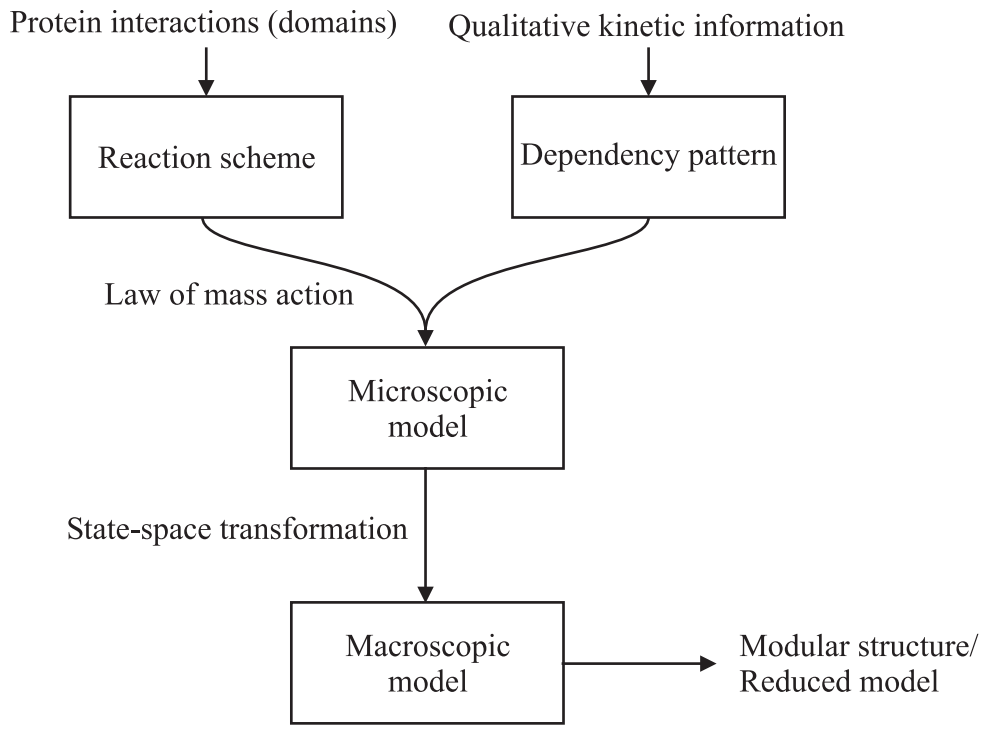

Figure 5.1: Systematic approach of domain oriented modeling. Information about protein interactions and established signal transduction pathways and qualitative kinetic knowledge are used to derive a full mathematical description, taking the combinatorial complexity into account. After application of the macroscopic states the system's inherent modular structure is revealed.

pancy levels.

Signal transduction is a highly complex issue, and different pathways are connected through key elements as for example receptors, scaffolds, second messengers and protein kinases. The greater goal of the macroscopic approach is not only to understand these different biological signaling processes in detail, but also in context with each other. Thereby it aims in providing a modular tool-set of quantitative mathematical descriptions of these processes without isolating them from each other and taking its multilevel cross-talk into account. This work has succeeded in modularization and reduction in several cases, which indicates that this may be a reachable goal, and further work under this aspect, can be recommended.

Bibliography

Aebersold, R. and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, 422:198 207.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *Molecular Biology of the Cell*. Garland Publishing.

Batzer, A., Rotin, D., Urena, J., Skolnik, E., and Schlessinger, J. (1994). Hierarchy of binding sites for grb2 and shc on the epidermal growth factor receptor. *Mol Cell Biol*, 14:5192 5201.

Blagoev, B., Ong, S., Kratchmarova, I., and Mann, M. (2004). Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat. Biotechnol.*, 22:1139 1145.

Blinov, M. L., Faeder, J. R., Goldstein, B., and Hlavacek, W. S. (2004). Bionetgen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*, 20:3289–3291.

Borisov, N., Markevich, N., Hoek, J., and Kholodenko, B. (2005). Signaling through receptors and scaffolds: independent interactions reduce combinatorial complexity. *Biophysical Journal*, pages 951–966.

Bunnell, S. C., Hong, D. I., Kardon, J. R., Yamazaki, T., McGlade, C. J., Barr, V. A., and Samelson, L. E. (2002). T cell receptor ligation induces the formation of dynamically regulated signaling assemblies. *J Cell Biol*, 158(7):1263–1275.

Chong, L. and Ray, B. L. (2002). Introduction to special issue whole-istic biology. *Science*, 295(5560):1661.

Conzelmann, H., Saez-Rodriguez, J., Sauter, T., Bullingera, E., Allgwer, F., and Gilles, E. D. (2004). Reduction of mathematical models of signal transduction networks: simulation-based approach applied to egf receptor signalling. *IEE Syst. Biol.*, 1:159–169.

Conzelmann, H., Saez-Rodriguez, J., Sauter, T., Kholodenko, B., and Gilles, E. (2006). A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. *BMC Bioinformatics*, 7(1):34.

Cooper, G. M. (2000). *The Cell - A Molecular Approach*. Sinauer Associates, Inc, Sunderland (MA), 2 edition.

Csete, M. E. and Doyle, J. C. (2002). Reverse engineering of biological complexity. *Science*, 295(5560):1664 – 1669.

Egan, S., Giddings, B., Brooks, M., Buday, L., Sizeland, A., and Weinberg, R. (1993). Association of sos ras exchange protein with grb2 is implicated in tyrosine kinase signal transduction and transformation. *Nature*, 363:45 51.

Elion, E. (2001). The ste5p scaffold. *J Cell Sci*, 114:3967 3978.

Faeder, J. R., Hlavacek, W. S., Reischl, I., Blinov, M. L., Metzger, H., Redondo, A., Wofsy, C., and Goldstein, B. (2003). Investigation of early events in fcri-mediated signaling using a detailed mathematical model. *J. Immunol.*, 170:3769–3781.

Ferguson, K., Berger, M., Mendrola, J., Cho, H.-S., Leahy, D., and Lemmon, M. (2003). Egf activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Mol Cell*, page 507 517.

Garrett, T., McKern, N., Lou, M., Elleman, T., Adams, T., Lovrecz, G., H-J, Z., Walker, F., Frenkel, M., Hoyne, P., Jorissen, R., Nice, E., Burgess, A.,

and CW, W. (2002). Crystal struture of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor a. *Cell*, 110:763773.

Hlavacek, W., Faeder, J., Blinoc, M., Perelson, A., and Goldsten, B. (2003). The complexity of complexes in signal in signal transduction. *Biotechnol Bioeng*, 84:782–94.

Isidori, A. (1995). *Nonlinear Control Systems*. Springer Verlag.

Jiang, X. and Sorkin, A. (2002). Coordinated traffic of grb2 and ras during epidermal growth factor receptor endocytosis visualized in living cells. *Mol Biol Cell*, 13:1522 1535.

Jorissen, R., Walker, F., Pouliot, N., Garrett, T., Ward, C., and Burgess, A. (2003). Epidermal growth factor receptor: Mechanisms of activation and signalling. *Exp Cell Res*, 284:31 53.

Kholodenko, B., Demin, O., Moehren, G., and Hoek, J. (1999). Quantification of short term signaling by the epidermal growth factor receptor. *J Biol Chem*, 274(42):30169–30181.

Kinet, J.-P. (1999). The high-affinity ige receptor (fceri): from physiology to pathology. *Ann Rev Immunol*, 17:931 972.

Kitano, H. (2002a). Computational systems biology. *Nature*, 420:206 210.

Kitano, H. (2002b). Systems biology: A brief overview. *Science*, 295(5560):1662–1664.

Li, N., Batzer, A., Daly, R., Yajnik, V., Skolnik, E., Chardin, P., Bar-Sagi, D., Margolis, B., and Schlessinger, J. (1993). Guanine-nucleotide-releasing factor hsos1 binds to grb2 and links receptor tyrosine kinase to ras signalling. *Nature*, 363:85 88.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., and Darnell, J. E. (2000). *Molecular Cell Biology*. W. H. Freeman & Co, New York, 4 edition.

Lowenstein, E., Daly, R., Batzer, A., Li, W., Margolis, B., Lammers, R., Ullrich, A., Skolnik, E., Bar-Sagi, D., and Schlessinger, J. (1992). The sh2 and sh3 domain containing protein grb2 links receptor tyrosine kinases to ras signaling. *Cell*, 70:431 442.

Maas, U. and Pope, S. (1992). Simplifying chemical kinetics: Intrinsic low-dimensional manifolds in composition space. *Combustion and Flame*, 88:239–264.

Mann, M. and Jensen, O. (2003). Proteomic analysis of post-translational modifications. *Nat. Biotechnol.*, 21:255 261.

March, M. and Ravichandran, K. (2002). Regulation of the immune response by ship. *Semin Immunol*, 14:37 47.

Meyer, T. and Teruel, M. (2003). Fluorescence imaging of signaling networks. *Trends Cell Biol*, 13:101 106.

Nijmeijer, H. and Van der Schaft, A. (1990). *Nonlinear Dynamical Controll Systems*. Springer Verlag.

Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.-H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., and Yokoyama, S. (2002). Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell*, 110:775 787.

Okabayashi, Y., Kido, Y., Okutani, T., Sugimoto, Y., Sakaguchi, K., and Kasuga, M. (1994). Tyrosines 1148 and 1173 of activated human epidermal growth factor receptors are binding sites of shc in intact cells. *J Biol Chem*, 269:1867418678.

Pacini, S., Valensin, S., Telford, J., Ladbury, J., and Baldari, C. (2000). Temporally regulated assembly of a dynamic signaling complex associated with the activated tcr. *Eur. J. Immunol.*, 30:26202631.

Pawson, T. (2003). Organization of cell-regulatory systems through modular-protein-interaction domains. *Philos Transact A Math Phys Eng Sci*, 361(1807):1251–1262.

Pawson, T. (2004). Specifity in signal transduction: From phosphotyrosine-sh2 domain interactions to complex cellular systems. *Cell*, 116:191–203.

Pawson, T. and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. *Science*, 300(5618):445–452.

Pelicci, G. and Lanfrancone L, Grignani F, M. J. C. F. F. G. N. I. G. F. P. T. P. P. (1992). A novel transforming protein (shc) with an sh2 domain is implicated in mitogenic signal transduction. *Cell*, 70:93 104.

Roussel, M. R. and Fraser, S. J. (2001). Invariant manifold methods for metabolic model reduction. *Chaos*, 11:196–206.

Rozakis-Adcock, M., Fernley, R., Wade, J., Pawson, T., and Bowtell, D. (1993). The sh2 and sh3 domains of mammalian grb2 couple the egf receptor to the ras activator msos1. *Nature*, 363:83 85.

Salomon, A. R., Ficarro, S. B., Brill, L. M., Brinker, A., Phung, Q. T., Ericson, C., Sauer, K., Brock, A., Horn, D. M., Schultz, P. G., and Peters, E. C. (2003). Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. *Proc Natl Acad Sci U S A*, 100(2):443–448.

Sauter, T. and Bullinger, E. (2004). Detailed mathematical modeling of metabolic and regulatory networks. *BIOforum Europe*, 2:62–64.

Schlessinger, J. (2000). Cell signaling by receptor tyrosine kinases. *Cell*, 103:211– 225.

Schoeberl, B., Eichler-Jonsson, C., Gilles, E. D., and Mller, G. (2002). Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol*, 20(4):370–375.

Stelling, J. and Gilles, E. D. (2004). Mathematical modeling of complex regulatory networks. *IEEE Trans Nanobioscience*, 3(3):172–179.

Toth, J., Li, G., Rabitz, H., and Tomlin, A. S. (1997). The effect of lumping and expanding on kinetic differential equations. *SIAM J. Appl. Math.*, 57(6):1531–1556.

Turner, H. and Kinet, J.-P. (1999). Signalling through the high-affinity ige receptor fceri. *Nature*, 402(suppl):B20– B30.

Wiley, H., Shvartsman, S., and Lauffenburger, D. (2003). Computational modeling of the egf-receptor system: A paradigm for systems biology. *Trends Cell Biol*, 13:43– 50.