Automated kinetic simulation of molecular interaction networks

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Biowissenschaften
der Johann Wolfgang Goethe Universität
in Frankfurt am Main

von
Jakub Pijewski
aus Warschau

Frankfurt 2007
(D30)

vom Fachbereich Biowissenschaften der
Johann Wolfgang Goethe-Universität
als Dissertation angenommen

Dekan:	Prof. Dr. Rüdiger Wittig
Erster Gutachter:	Prof. Dr. Bernd Schürmann
Zweiter Gutachter:	Prof. Dr. Gisbert Schneider
Datum der Disputation:	

Contents

Acknowledgements

I am grateful to many persons who have contributed to making this work a success. Especially, I would like to thank:

Prof. Dr. Bernd Shürmann for supervising and evaluating this work and for his generous support.

Prof. Dr. Gisbert Schneider for his commitment and evaluation of this work.

Prof. Dr. Horst Stöcker for his generous support.

Special thanks to Dr. Martin Stetter for his constant support, supervision, inspiration and enthusiasm.

Dr. Michael Meyer-Hermann and his group at FIAS and to Dr. Martin Stetter and his group at Siemens for insightful discussions, very pleasant working atmosphere and good time.

Last but not least, to my family and friends, especially to Andy for his heart-and-soul engagement and to Tatiana for the greatest PhD 101 ever.

Chapter 1

Introduction

The biological complexity comprises of several levels, starting with single molecules, going through biochemical reactions, cells, tissues, organisms and ending up at ecosystems [121]. All of these levels are at the interest of theoretical research in the domain of Systems Biology and Computational Biology and can be modeled mathematically [49], [68], [158] and simulated with computational methods [76], [135], [201] in order to unveil the mechanisms determining the organization at each level and support experimental research [122] [123].

The level of molecular interaction networks (MIN), is a good example of how the theoretical research can contribute to understanding vital biological and medical problems, such as regulation of gene expression [25], intracellular signaling [97], regulation of metabolism [211] or drug discovery [139]. Thus, it is of general interest to develop efficient approaches for modeling of MIN [45] [128], [135], [190]. This work presents one such approach.

1.1 Features of MIN in the cell

1.1.1 Definition of molecular interaction

Living cells are filled with a number of molecule types, proteins in particular, that can potentially interact with each other [24], [145]. Pairwise interactions can be combined into pathways which generate complex cellular responses [166], [167].

There are 3 established standards for interaction data exchange and modeling: Proteomics Standards Initiative Molecular Interaction XML for-

mat (PSI MI) [104] [105], Systems Biology Markup Language (SBML) [83], [109] and Biological Pathways Exchange format (BioPAX) [28]. All these standards consider interactions to be always based on simple physical contact of molecules [200].

Furthermore, the mentioned standards differentiate between several basic types of physical interaction, such as binding or enzymatic modification. The PSI MI defines only one type of physical interaction - aggregation (binding), formalized description of other interaction types, such as enzymatic modification is under development [162]. The SBML is able to describe binding, transformation and transport, which can be related to kinetic rate laws. The BioPAX offers the broadest descriptive scope with such categories as: complex assembly, catalysis, modulation, transport (see [200] for review).

Finally, the basic interaction types can be further classified into numerous sub-types, depending on different existing structural motifs of proteins [26],[46].

A modeling framework can consider any of the above levels of detail (physical contact - biochemical character of interaction - structure of interacting protein regions). In this work, we will define **molecular interaction** as any physical contact between two molecules in the cell, especially: binding, enzymatic modification or transport. We will further assume, that any regulatory relationships between interactions (like cooperative binding or inhibition of enzymatic reactions), can be also exerted only by means of physical contact. We will put no further limitations on the biochemical nature of the interacting molecules (type of molecule, shape etc.).

1.1.2 Detection of interactions and networks

Molecular interactions in the cell can be detected using several biochemical methods (see [127] fo review). Interaction detection on systematical basis is possible thanks to the advancement of high throughput methods (Hi-Tru), such as: *a*) yeast two-hybrid assay (Y2H) [57], [81], *b*) tandem affinity purification (TAP) [170], [176] coupled with mass spectrometry (MS) [148], [164], *c*) synthetic lethality [205] and development of bioinformatical approaches, such as *d*) literature mining (see [177] for review) or *e*) comparative genomics [175]. The molecular interaction data resulting from these heterogeneous sources is stored in numerous databases (see [85], [151] for reviews).

The binary interactions between the molecules in the cell can be put together into a network [165]. Large-scale protein interaction maps covering

substantial parts of the genome exist already for some model organisms, such as:

- budding yeast *Saccharomyces cerevisiae* - several maps based on various methods: Y2H [111], [210], Y2H combined with literature research [189], TAP/MS [88], [107] and synthetic lethality [206].
- fruit fly *Drosophila melanogaster* - several maps, all based on Y2H [84], [91] or modified Y2H [194].
- nematode *Caenorhabditis elegans* - single map based on Y2H and independently confirmed with TAP [142].
- human - several maps based on various methods: Y2H [178], [197], literature mining [171], manual curation of literature and databases [86] and *in silico* comparisons with other species [50], [141], [168].

For a specific organism, the above maps represent only subsets of a complete MIN and they show little overlap witch each other, even if obtained using the same method [84], [111]. For instance, different human interaction maps share only about 10% of interactions [54]. This incompleteness, together with relatively noisy character of Hi-Tru interaction data [215] contribute to partially contradictory results when analyzing network properties of different maps, as will be described in the next section.

First developments towards complete, genome wide MINs enabling a consistent structural analysis have been made by combining and curating the fragmentary maps into more exhaustive MIN for budding yeast [174] and human [54].

1.1.3 Topological characteristics of MIN

One of key structural properties of any MIN its the density of connections, which can be measured by **degree distribution**. A degree of a node is the number k of its connections to other nodes, in our case it is the number of interactions that a molecular species is involved in. In a random network, the degree distribution over nodes has a Poisson character and is centered around an average value of k in this network, whereas MIN are characterized by a non-random distribution, with most of the nodes having only a few links and only a few nodes, having many links [30].

The highly connected nodes are called hubs and can be divided into two types - **"party" hubs** and **"date" hubs** [99]. Party hubs are co-expressed

and co-localized with their partners, and thus can be involved in many simultaneous interactions [99]. This is facilitated by the fact, that party hub proteins are usually long and contain many binding domain repeats [69]. Date hubs interact with partners expressed at different times and locations and thus are involved in many transient interactions [99] which is facilitated by a feature called **intrinsic disorder** [67]. Intrinsically disordered proteins and protein regions lack an unique 3-D structure and exist in a dynamic ensemble of conformations [102]. Binding of a partner to a disordered region invokes an disorder-to-order transition which enables highly reversible interactions while maintaining high specificity [192]. Date hub proteins have significantly more disordered regions compared to party hubs [192].

The existence of hubs influences network **robustness**, i.e. ability to maintain function upon removal of nodes - the MIN are robust to removal of peripheral nodes but sensitive to knock-out of hubs. In one study, the likelihood that removal of a protein will prove lethal to the cell clearly correlates with the number of interactions the protein has [113]. However, in another study such correlation was not found, instead, number of interactions a hub has with other hubs turned to be a good predictor of its influence on network robustness[31].

Another apparent feature of MIN is the **small-world organization** - a travel between any pair of nodes in the network requires hopping through a low number of intermediate nodes [216]. This is related to the fact that MIN usually contain locally dense regions (clusters) that are sparsely connected to other clusters [216]. These clusters are thought to correspond to functional modules responsible for a specific process [87], [99]. However, the **modular structure** of MIN is also open to discussion, since an stratic (layer-like) structure, also empirically founded, has been proposed as an alternative [31].

As evident from above examples, the exact characterization of structural features of MIN still needs to be elaborated, since the current results may be biased by data incompleteness of existing interactome maps. This fact makes definition of precise modeling requirements difficult. However, at least one apparently well-founded implication of the above considerations is that a successful modeling approach for MIN must be able to depict hub nodes with their potentially many simultaneous connections.

1.1.4 Combinatorial complexity in MIN

The term 'combinatorial explosion' origins from administration and computing, where it means a rapidly accelerating increase in lines of communication as organizations are added in a process [6]. In analogy, in a system of interacting molecules various combinations of multimolecular complexes with different properties can occur [34]. This feature is called 'combinatorial complexity' [106] and it results from both multivalent binding (different structures of multimolecular assemblies) and multivalent enzymatic modification (different properties of these assemblies) [130]. For instance, a protein with n binding or covalent modification sites can have up to 2^n distinct states [73]. Thus, the number of possible species can exceed 10^6 for $n > 20$ which clearly poses a difficulty on modeling of MIN.

Combinatorial complexity has been observed experimentally in many biological systems, like signaling pathways, where multimolecular complexes occur (signalosomes) [48]. For instance, T-cell receptor (TCR) signaling involves assembly of various multimolecular complexes, both transient [51] and stabilized with multivalent cooperative interactions [52]. Similar relationships occur in metabolism due to many-to-many relationships between enzymes and substrates [101]. Thus, combinatorial complexity implies existence of highly branched, physiologically meaningful signaling and metabolic pathways as opposite to the traditional, linear approach [37] [96].

Thus, successful modeling needs to treat combinatorial complexity both in terms of large variable numbers and branched, interconnected pathway structure. Several approaches exist to achieve these goals:

- **rule-based description**. The system is described using a limited set of rules related to binary interactions from which all combinatorial molecular assemblies and related chemical reactions can be automatically derived by the simulation software [73], [74]. This approach relies on parameter extrapolation [147], thus only a limited parameter set specified in the rules is sufficient to derive the combinatorial description [37]. However, the rule-based approach can potentially lead to a problem of unwanted polymerization, where the simulation software generates erroneous polymeric structures if multiple binary interactions are possible between the same pair of molecules [47], [147].

- **domain-exclusion**. The total number of possible combinations of molecules is reduced by excluding all combinations containing molecules targeting the same physical site on a scaffold or receptor protein [60].

For instance, in a molecule with single site to which potentially n partners can bind, the total number of molecular combinations is reduced from 2^n to $(n+1)$, where 1 stands for the completely unbound state.

- **domain-separation**. Two physical sites located on the same protein but not influencing each other in any way (e.g. by regulation) are dissected and treated as parts of two independent molecules [42] [43]. For instance, in a molecule with two mutually independent groups of n and m sites, respectively, the total number of molecular combinations is reduced from 2^{n+m} to $2^n + 2^m$.

- **on-the-fly combination**. At a given time point of the simulation, only the combinations of molecules that are actually involved in the currently simulated reactions enter the variable list [147]. This approach relies on an observation that under most conditions only a small portion of the MIN is active [75]. However, this active portion can change dramatically with conditions [75].

1.2 Models of MIN

A model can be defined as any representation of a system [70]. In this work, we will narrow this definition to mathematical representation. For MIN, this representation can be constructed at different levels of abstraction, from structural (top-down) to mechanistic (bottom-up) approaches [110], [198].

1.2.1 Overview of modeling strategies

Structural representations

The top-down models describe components of the MIN (e.g. molecular species) and connections between them (e.g. biochemical interactions). At this level of abstraction, the components can be generalized to higher-level entities, e.g. pathway modules [35]. Similarly, the connections can be abstracted from direct physical interactions to higher level relations, like gene-gene relationships [198]. Such generalized connections can be described in terms of conditional probabilities [64].

Mechanistic representations

The bottom-up models describe both the network structure and dynamics, i.e. behavior in time. Thus, the MIN components are described with

time-dependent variables (e.g. concentrations of molecular species) and the relations between them are described in terms of coupling functions having these variables as arguments:

$$X_i(t_{n+1}) = f_i[\mathbf{X}(t_n)] \quad \text{if time is discrete} \tag{1.1}$$

$$dX_i/dt = f_i[\mathbf{X}(t)] \quad \text{if time is continuous} \tag{1.2}$$

Where $\mathbf{X}$ is the set of variables and $\mathbf{f}$ is the set of coupling functions.

Within the mechanistic models, also different levels of abstraction are possible. A relatively abstract category of mechanistic models is based on Boolean networks, where variables X_i have a binary value (e.g. 'active' - 'inactive') and change according to logical rules [115]. More detailed description is possible using kinetic models, where variables can have real-number (concentrations) or integer values (number of molecules) and change according to deterministic or probabilistic rules [71], [129] [160].

The selection of most suitable modeling abstraction level depends on the intended scope of the model. In this work we present a detail-oriented modeling framework that is based on the kinetic approach. Thus, we will describe the category of kinetic models in a more detail below.

1.2.2 Kinetic models

The kinetic models rely on the **law of mass action** formulated in 1864 by C. Guldberg and P. Waage based on experimental results of M. Berthelot and P. St. Gilles on ethylacetic ester synthesis [134]. It states, that the rate (velocity) v of a reaction is proportional to the quantity Q of the reacting substances [62], [214]:

$$v = k\Pi_{i=0}^{n} Q_i \tag{1.3}$$

Where: k - rate constant, n-number of interacting molecular species ($n > 2$ is considered to be highly improbable).

Reaction rate constants in kinetic models

The rate constant k depends on several physical factors as outlined in the Eyring equation [72]:

$$k = \frac{RT}{N_A h} e^{\frac{\Delta S^{\ddagger}}{R} - \frac{\Delta H^{\ddagger}}{RT}} \tag{1.4}$$

where: T - absolute temperature in K (Kelvin degrees), R - molar gas constant = 8.314472 J mol^{-1} K^{-1} [156], N_A - Avogadro constant = 6.0221415 10^{23}

mol^{-1} [156], h - Planck constant = 6.6260693 10^{-34} J s [156], $\Delta S^{\ddagger}$ - activation entropy in J mol^{-1} K^{-1}, i.e. the probability that reaction occurs because substrates are properly arranged in space [62], $\Delta H^{\ddagger}$ - activation enthalpy in J mol^{-1}, i.e. the energy necessary for the substrates to form the intermediate complex [62].

Since the value of k can be influenced by the above factors, a term 'rate coefficient' would be more appropriate [196]. However, this work deals primarily with kinetic models that assume constant temperature and pressure, thus we will keep the traditional nomenclature of k as 'rate constant'.

The values of constants k are often unknown and need to be assumed, which is a major limitation of the kinetic models [110]. For some reactions, experimentally measured values of k or related parameters (like Michaelis constants of enzymatic reaction) are available from several databases, such as: BRENDA [187], [188], TECRDB [92] or SABIO-RK [217]. This data is, however, not always applicable in a straightforward manner due to discrepancies between the *in vivo* conditions assumed in the simulation and the *in vitro* conditions of the measurement [143]. Values of k can be also estimated using various computational methods, such as numerical optimization, neural networks, genetic and evolutionary algorithms (see [116], [157], [207], [212] for reviews).

Stochastic and deterministic kinetic models

The molecular quantity Q can be represented in two ways [147]:

- as an integer **molecule number** changing in a discrete manner by probabilistic rules (stochastic description) [89]. The stochastic approach is based on the notion, that interacting molecules move randomly through the reaction volume and if a reaction takes place, the molecular quantities can only change by integer amounts [89]. Thus, the time evolution of the system is probabilistic and discrete in time, which allows for modeling of systems with low molecular quantities, where the continuity of concentration changes can not be assumed [118].

- as a real-valued **concentration** changing in a continuous manner according to a differential equation that leads to a defined state in the future (deterministic description) [158]. The deterministic approach is a widely accepted way of describing many cellular processes in terms of temporary changes of molecular concentrations due to biochemical

reactions [196]. We assume, that a realistic unit range for concentration is $1nM$ to $1\mu M$ based on experimental observations for some MIN [41]. This implies the number of molecule in a given reaction volume in the order of 10^{14} - 10^{17} (based on the Avogadro constant), which is compliant with the assumption of infinite molecular quantities underlying the deterministic description. Assuming such concentration units implies that the units of rate constants k can be set to $\mu M^{-1}s^{-1}$ for bimolecular and s^{-1} for unimolecular reactions (taking s as a time unit is somewhat arbitrary but commonly practiced).

Spatial and non-spatial kinetic models

The cell can be assumed to be a homogenous, well-stirred reactor and thus diffusion is not taken into account in the model [103]. Alternatively, the cell is assumed to be a collection of compartments with fluxes of molecules between them [193]. Thus, the concentrations of molecular species change both in time and space, due to biochemical reactions and diffusion [71]. This allows for simulation of systems, where the spatial resolution plays an important role, for instance signal transduction in neurons, that has been modeled with both deterministic [146] and stochastic [199] spatial approaches.

1.2.3 Elementary and approximated mass action approach

The differentiation between deterministic/stochastic and spatial/non-spatial kinetic models leads to four possible model classes [147], from which the non-spatial, deterministic approaches based on ordinary differential equations (ODE) require the least computational power and / or parameter set and thus are a popular approach to modeling of systems where spatial resolution and low numbers of molecules do not have to be considered (see [71], [129], [160] for reviews). Henceforth, whenever speaking about ODE models of MIN, we will mean the non-spatial, deterministic, ODE-based models.

However, classical ODE formulations, with the most prominent example of Michaelis-Menten kinetics, rely on several approximations, e.g. concerning the steady state of enzyme-substrate complex and enzyme saturation [62], which are not necessarily valid for MIN [41], [59]. We will henceforth call such approximated kinetic ODE formalisms **Approximated Mass Action** (AMA) as opposed to non-approximated ones, which we will call **Elementary Mass Action** (EMA).

Definition of EMA and AMA

We will define EMA as a non-spatial, kinetic ODE formalism describing molecular interactions in terms of elementary association-dissociation reactions in the form of law of mass action (Eq. 1.4). Of course, other general approximations underlying the ODE kinetic modeling in general (Sec.1.3) do still hold for EMA. That is why prefer calling this model class 'elementary' but not 'exact'.

We will further define AMA as a non-spatial, kinetic ODE formalism describing molecular interactions with forms derived from an EMA description based on additional assumptions simplifying the conservation relationships (like negligibility of the enzyme-substrate complex) or dynamics of some variables in the system (like the quasi-steady state approximation for the enzyme substrate complex).

Description of enzymatic systems with EMA and AMA

The difference between EMA and AMA can be illustrated with the standard example of an enzymatic reaction, where a substrate S is converted into a product S^* by an enzyme E. This reaction can be seen as a composition of an elementary association-dissociation reaction $[E] + [S] \leftrightharpoons [ES]$ (rate constants: k_1 and k_{-1}) with an added transformation step $[ES] \rightarrow [E]+[S^*]$ (rate constant k_{11}) [62].

The EMA description following from this scheme is: (cf. Eq. 2.6)

$$\begin{aligned} d[S]/dt &= -k_1[S][E] + k_{-1}[ES] \\ d[ES]/dt &= +k_1[S][E] - k_{-1}[ES] - k_{11}[ES] \\ d[S^*]/dt &= +k_{11}[ES] \\ d[E]/dt &= -k_1[S][E] + (k_{-1} + k_{11})[ES] \end{aligned} \tag{1.5}$$

With following conservation equations:

$$\begin{aligned} [\mathrm{S}_T] &= [S] + [ES] + [S^*] \\ [\mathrm{E}_T] &= [E] + [ES] \end{aligned} \tag{1.6}$$

The AMA description of this system is based on two additional assumptions. First, the enzyme concentration is assumed to be at least one order of magnitude smaller than the concentration of the substrate: $[E_T] << [S_T]$. This allows to neglect the $[ES]$ within the total concentration of S, i.e.: $[S_T] \approx [S] + [S^*]$. Second, the $[ES]$ is assumed to reach steady state much

quicker than other variables in the system, thus $d[ES]/dt \approx 0$. Based on the two above assumptions, an AMA description of the enzymatic reaction can be derived:

$$d[S^*]/dt = \frac{V_{max}[S_T]}{[S_T] + KM} \tag{1.7}$$

Where: $V_{max} = k_{11}[E_T]$ and the Michaelis constant $KM = (k_{-1} + k_{11})/k_1$.

The Eq. 1.7 is the standard Michaelis-Menten description of an enzymatic reaction [62], which is the canonical way of describing enzymatic systems, such as metabolic networks. However, the assumptions underlying AMA are considered to be impropriate for modeling of enzymatic cascades [41], [59], especially because in such case both enzyme and substrate of a given reaction are proteins and can come in concentrations of the same order of magnitude [23]. Thus, the AMA and EMA applied to MIN modeling can produce significantly different results [59] [155].

However, describing MIN in terms of EMA increases both the number and length of ODE for a given system. Thus, it is difficult to unify the descriptive detail of EMA with the combinatorial complexity of larger systems.

1.3 Summary of modeling requirements and alternatives

The biological background and existing mathematical approaches applicable to modeling of MIN outlined in previous sections can be summarized to following requirements for a novel and advantageous MIN modeling framework:

- Description of different types of molecular interactions like binding or enzymatic reactions. This description should be based on the assumption, that interactions and regulations rely on physical contact of molecules.
- Description of molecular species having many simultaneous interactions in order to depict network hubs.
- Management of combinatorial complexity, i.e. automatic derivation of all possible combinations of interacting molecular species in the

system and related reaction pathways while avoiding artificial creation of non-existing polymers. This can be eventually achieved using rule-based system description and domain-oriented definition of molecular species.

- Modeling abstraction level suitable for the intended scope of the model and computational efficiency.
- Possibly far-reaching automation of model construction for MIN of arbitrary size and complexity.

We believe that the above requirements can be met by applying ODE-based modeling approaches, since this class of models allows relatively detailed, mechanism-oriented description in terms of molecular concentrations that can be well related to physiological states of the cell [208]. However, we have to be aware of following limitations of the ODE-based approaches:

- Assumption of isobaric and isothermal conditions (constant temperature and pressure).
- Deterministic description in terms of concentrations is unable to correctly describe systems with low molecular quantities.
- Assumption of spatial homogeneity (all concentrations are uniform in the volume of the study). Thus, no spatial resolution can be included into the model.
- Existing ODE formulations are based on several approximations (AMA) which are not compliant with the nature of MIN. Application of an EMA formalism in this respect is more appropriate.

Through this work, we will develop an EMA-based formalism for kinetic modeling of MIN with ODE (Chapter 2) that allows automated model construction for MIN of arbitrary size and complexity (Chapter 3). We will also apply this formalism to investigation of several biological systems and compare the outcomes with results of a classical AMA description (Chapter 4).

Chapter 2

Formal description of molecular interaction networks

This chapter presents a novel formal description of MIN addressing the requirements and limitations discussed in Chpt. 1. The presented approach relies on automated combination of EMA ODE modules based on a set of simple, user-defined rules. This allows for a far-reaching automation of the modeling process and treating network complexity without compromising mathematical precision. The descriptive scope of our formalism covers following aspects: *a*) basic interaction types (binding and enzymatic reactions, synthesis and degradation), *b*) regulation of those (including cases of multiple regulators and logical relationships between them) and *c*) abstract biological functions, as outlined in the following sections.

2.1 Conceptual framework

The most common approach to depict MIN is to treat each molecular species (for example proteins) as a network node and the interactions between species (like binding or enzymatic modification) as the network links (Sec. 5.1.1). In such a description, capturing regulation, i.e. relationships between different interactions of the same species, is either impossible or requires introduction of many additional nodes representing various combinations of the regulated species with its regulators. The same combinatorial work is required for translating the network structure into a set of variables. In the case of species with a high number of interactions, like the network

hubs typical for MIN (Sec. 1.1.3), such an approach turns out to be especially inefficient and error-prone due to combinatorial explosion of variables and parameters. Here we propose an alternative approach based on a two-level description of MIN as outlined in Fig. 2.1.

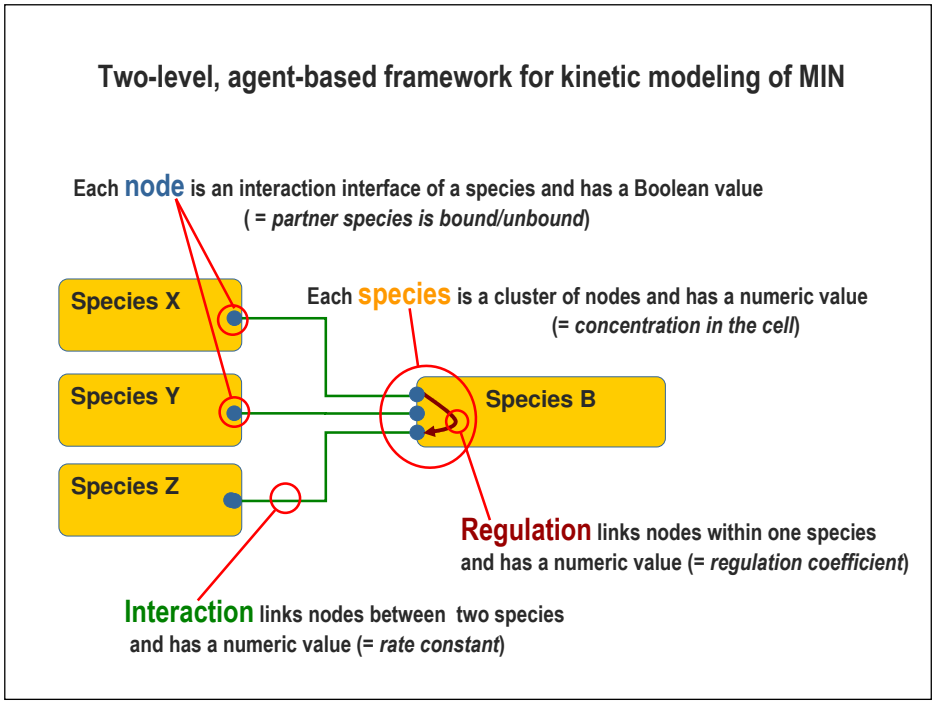


Figure 2.1: Conceptual framework of the proposed approach to MIN modeling. In this example, a species B can bind three different partners: X, Y, and Z. Each of these partners has a binding interface to B and B has three binding interfaces, one for each partner. The binding of X to B can regulate the binding of Z to B.

2.1.1 Two-level network representation

The key idea of the proposed approach is to split the network representation into two levels, where each molecular species is not a single node but a network cluster composed of several nodes. We define following main model components: species, interfaces, interactions, regulations and phenomena. These components are described in detail in the Sec. 2.1.2 and the relations between them are summarized in Tab. 2.1.

	main entity	**entity name**	**entity-related value**
level 1	cluster of nodes	species	concentration
level 2	single node	interface	Boolean
	link type	**link name**	**link-related value**
level 1	connects interfaces of 2 species*	interaction	rate constant
level 2	connects interfaces of 1 species	regulation	regulation coefficient

Table 2.1: Main components of the proposed network representation. * - with exceptions (see text for details).

2.1.2 Components of the model

Species.

We will define **species** as any basic type of molecule (for example a type of protein) included in the MIN. A numerical concentration value is assigned to all species. This concentration is a variable of an ODE system describing behavior of the MIN in time. As outlined in Sec. 1.2.2, such ODE can be derived based on the law of mass action, which assumes a large number of molecules of each given species in the reaction volume. We will refer to species as an abstract category and when talking explicitly about an amount of molecules we will use the term **species concentration**. In case of single molecule of a given species we will refer to it as **species molecule** later in text.

Interface.

Species can bind to each other in various combinations. To capture these combinations, we describe a species as a collection of **interaction interfaces** to other species. Each interface has a Boolean value telling if a given interaction is taking place or not in a specific combination of molecules. We will denote interfaces with the Greek letter σ. Interfaces typically refer to physical regions of the molecule (e.g. binding sites) but for mathematical consistency need to be considered as abstract terms. Especially, for two species X and Y binding at the same physical site to species Z, we will assign two distinct interfaces of Z - because they are related to different binding rate constants. For the same reason, if n molecules of X can bind to a single molecule of Z at different physical sites, the species X will receive n different interfaces.

Sub-species.

We will define **sub-species** of a given species as any possible combination of its interface values. Precisely this representation of sub-species as Boolean vectors allows for efficient and automated treating of combinatorial complexity. Henceforth we will denote sub-species by adding to the species name an index specifying values of all interfaces in this sub-species. For example, for a species X that can interact with 3 other species and thus has 3 interfaces σ^X_{1-3}, the symbol X_{010} means a sub-species of X, where $\sigma^X_1 = false$, $\sigma^X_2 = true$ and $\sigma^X_3 = false$, i.e. a bimolecular complex of X with its second interaction partner. Alike species, sub-species are also abstract categories and will be distinguished from **sub-species concentration** and **sub-species molecule** later in text. All sub-species have also a numerical concentration value that is a variable of the ODE system.

Interaction.

We will define **interaction** as any type of relation that can be measured with a biochemical rate constant, in particular: binding of two molecules, enzymatic modification, synthesis and degradation. Defined in such way, an interaction does not necessarily involve two species; specific cases of binding (homopolymerization) and enzymatic reaction (autocatalysis) and most cases of synthesis and degradation involve molecules of only one species. In general, interactions are translated into ODE by multiplication of the rate constant with a species concentration term (Sec. 2.2). The interaction rate constants are parameters of the model.

Regulation.

We will define **regulation** as any type of relation between two interfaces of the same species, especially activation or inhibition of one interaction by the other. Regulations are translated into ODE by multiplication of a rate constant with a regulation coefficient (Sec. 2.4) Alike rate constants, the regulation coefficients are parameters of the model.

Phenomena.

Finally, we define a special category of model components that allows incorporating meaningful biological functions such as 'cell mass' or 'cell division' into the model. We will call these components **phenomena**. They can have either a quantitative or a qualitative character and thus correspond to

a real-valued or a Boolean system variable. Translation of phenomena and their links with species into differential equations is explained in detail in Sec. 2.5.

2.1.3 Agent-like view of the network

The outlined conceptual framework allows to treat each molecular species as an individual agent, with a specific interaction menu and a resulting set of sub-species (Sec. 2.2). We refer here to a very broad definition of 'agent' as an 'an entity that is capable of perception and action', not to the specific term related to 'agent-based computing' [21].

The above interaction menu is defined as a list of rules, which allows an automatic creation of possible sub-species and reaction pathways in the system and thus efficient treatment of combinatorial complexity as will be presented in Chpt. 3.

Such agent-like representation relies on the biologically founded assumption that all interactions and regulations are based on direct physical contact between molecules (Sec. 1.1.1). Thus, the only sub-species of A that can be tracked in the system refer to combinations of A with its direct interaction partners, irrespectively of further combinatorial status of these partners. With such approach, we cannot track some polymeric structures composed of repetitive units. For example, if A has two interfaces to B, we can only track sub-species A, AB and ABB but not $ABABAB\ldots$ etc. This is, however, consistent with the assumption about direct physical contact, from which it follows that in this case any potential function of A can be regulated only by 0, 1 or 2 interactions with B irrespectively if a given molecule of A is a part of a longer chain or not. Actually, artificial creation of such polymer chains is regarded as a methodological problem reported for other rule-based modeling formalisms [147]. In our approach, this problem does not occur by definition.

Moreover, the agent-like approach allows for a substantial reduction of the number of combinatorial variables. In a system with n species each having k interactions to each other, the number of possible sub-species per species $= 2^k$. The resulting total number of sub-species in the system in a non-agent approach can potentially equal to 2^{nk}, whereas in our approach it is reduced to $n2^k$.

2.2 Representation of basic interactions

Graphically, species are represented as rectangles, interfaces as dots placed on the edges of rectangles, interactions and regulations as links between these dots as shown in Fig. 2.2. To facilitate future representation of regulatory influence on synthesis and degradation, we introduce 'empty' nodes linked to synthesis and degradation constants as shown in Fig. 2.2 C-E. However, these nodes do not depict any interaction interfaces.

Mathematically, each link translates into an ODE module which, according to the law of mass action, consists of coefficients describing the interaction (e.g. rate constants) and species concentration terms. We will define both a graphical and mathematical description for several following basic interaction types: binding (Sec. 2.2.1), enzymatic modification (Sec. 2.2.2), synthesis (Sec. 2.2.3) and degradation. (Sec. 2.2.3).

2.2.1 Representation of binding

Binding is the most basic interaction. In our example case from Fig. 2.2 A, species A and B bind to each other with the rate constant k_1 and unbind with the constant k_{-1}. Thus, A and B have each one interface, σ_1^A and σ_1^B, respectively, that can be either true or false. The only possible sub-species for A and B are:

- A_0 ($\sigma_1^A = 0$) - free fraction of A.
- A_1 ($\sigma_1^A = 1$) - fraction of A bound to B (complex AB).
- B_0 ($\sigma_1^B = 0$) - free fraction of B.
- B_1 ($\sigma_1^B = 1$) - fraction of B bound to A (complex BA).

Note that in this simple case, the sub-species A_1 and B_1 correspond to an identical biochemical entity (complex AB = complex BA). This artificial distinction results from the modeling assumptions necessary for our agent-like approach (as outlined in Sec. 2.1). Such correspondence of sub-species proves useful for treatment of more complicated cases where degradation or regulation occur (Sec. 2.2.3).

Thus, for future reference we will define here a **corresponding sub-species set of A towards B** as such set of sub-species of A that corresponds to an identical biochemical entity as some specific set of sub-species of B. From this it immediately follows, that the concentrations of sub-species in corresponding sets must be identical. In the case of basic binding

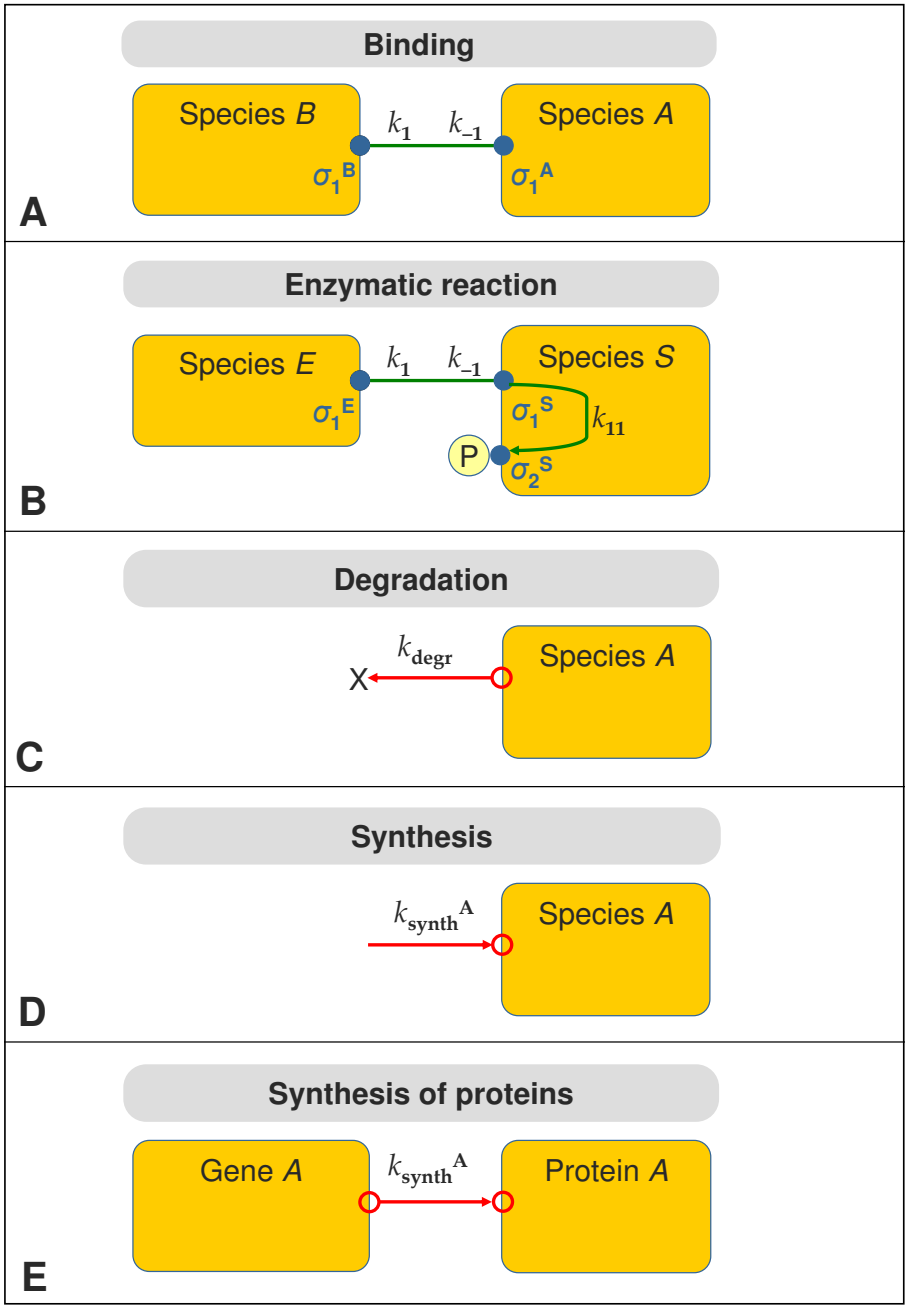


Figure 2.2: Graphical representation of basic molecular interactions. Rectangles represent species, lines and arrows represent interactions, k_n denote rate constants, filled dots represent interfaces, σ_n^X denote interfaces, empty circles are introduced only for visual convenience (see text for details). The enzymatic reaction (**B**) refers in particular to phosphorylation and dephosphorylation.

of A to B, there is only one correspondence relation (quadratic brackets indicate concentrations): $[A_1] = [B_1]$.

The differential equations describing the basic binding interaction have the following EMA form (rate constants as in Fig. 2.2 A):

$$\begin{aligned} d[A_0]/dt &= -k_1[A_0][B_0] + k_{-1}[A_1] \\ d[A_1]/dt &= +k_1[A_0][B_0] - k_{-1}[A_1] \\ d[B_0]/dt &= -k_1[B_0][A_0] + k_{-1}[B_1] \\ d[B_1]/dt &= +k_1[B_0][A_0] - k_{-1}[B_1] \end{aligned} \tag{2.1}$$

With following conservation equations:

$$[\mathrm{A}_T] = [A_0] + [A_1] \tag{2.2}$$

$$[\mathrm{B}_T] = [B_0] + [B_1] \tag{2.3}$$

Note, that the conditions in Eq. Sys. (2.3) are automatically fulfilled because the Eq. Sys. (2.1) sums up to zero. This is an universal feature of our modeling approach resulting from consequent application of EMA and it has a fundamental meaning, since it allows modular composition of differential equations for basic motifs when describing the more complicated cases without violating the mass conservation relationships in the system. Such violation can often occur when applying AMA in a modular way, which may affect the calculated results. We analyze this problem in detail in Chpt. 4.

2.2.2 Representation of enzymatic reactions

Enzymatic reaction is also a basic reaction type defined in our approach, however a bit more complicated one. In the most general case, one can describe the enzymatic reaction between an enzyme E transforming substrate S into a product S^* as follows [62] (quadratic brackets indicate concentrations, $\leftrightharpoons$ indicate bidirectional reactions):

$$[E] + [S] \leftrightharpoons [ES] \leftrightharpoons [ES^*] \leftrightharpoons [E] + [S^*] \tag{2.4}$$

Where $[ES]$ and $[ES^*]$ are the enzyme-substrate and enzyme-product complexes, respectively. This scheme can be simplified based on following assumptions:

- The step $[ES] \leftrightharpoons [ES^*]$ is much faster than other steps in the scheme 2.4 and thus we can omit this step in the mathematical representation.

- The step $[ES^*] \leftrightharpoons [E] + [S^*]$ is mostly irreversible in biological conditions, especially if we consider reactions of covalent modifications of proteins, like phosphorylation. For simplicity, we will assume this step to be always irreversible. We further assume a conversion of $[S^*]$ back to $[S]$ as perfectly possible, but only in a separate reaction carried out by a different enzyme.

- The concentration of ATP and related nucleotides consumed in phosphorylation reactions and concentration of phosphate groups produced in dephosphorylation reactions is assumed to be in an excess and thus can be parameterized within the rate constants for the substrate-product conversion.

The above assumptions yield a following simplified enzymatic reaction scheme [62]:

$$[E] + [S] \leftrightharpoons [ES] \rightarrow [E] + [S^*] \tag{2.5}$$

In this representation, the enzymatic reaction can be seen as a binding reaction $[E] + [S] \leftrightharpoons [ES]$ (rate constants: k_1 and k_{-1}) with an added transformation step $[ES] \rightarrow [E] + [S^*]$ (rate constant k_{11}). This scheme is depicted in Fig. 2.2 B.

For the binding part, both species S and E have one interface, σ_1^S and σ_1^E, respectively. The substrate species S has a second interface σ_2^S corresponding to the enzymatic modification site at which $[E]$ transforms $[S]$ ($\sigma_2^S = 0$) into $[S^*]$ ($\sigma_2^S = 1$). Depending o the combinations of states of σ_1 and σ_2, the possible sub-species of E and S are:

- E_0 ($\sigma_1^E = 0$) - free fraction of enzyme E.
- E_1 ($\sigma_1^E = 1$) - enzyme-substrate complex ES
- S_{00} ($\sigma_1^S = 0$, $\sigma_2^S = 0$) - free fraction of substrate S.
- S_{10} ($\sigma_1^S = 1$, $\sigma_2^S = 0$) - enzyme-substrate complex ES.
- S_{01} ($\sigma_1^S = 0$, $\sigma_2^S = 1$) - product S^*.

Here again we have one sub-species correspondence relation $[E_1] = [S_{10}]$. Note that from the assumptions leading to the simplified enzymatic reaction scheme 2.5 it follows immediately, that it is not allowed to create a sub-species of S where both the E is still bound and the S is already transformed into S^* ($\sigma_1^S = 1$, $\sigma_2^S = 1$).

The ODE describing the basic enzymatic interaction with EMA are as follows (quadratic brackets indicate concentrations, rate constants as in Fig. 2.2 E:

$$\begin{aligned} d[S_{00}]/dt &= -k_1[S_{00}][E_0] + k_{-1}[S_{10}] \\ d[S_{10}]/dt &= +k_1[S_{00}][E_0] - k_{-1}[S_{10}] - k_{11}[S_{10}] \\ d[S_{01}]/dt &= +k_{11}[S_{10}] \\ d[E_0]/dt &= -k_1[S_{00}][E_0] + (k_{-1} + k_{11})[E_1] \\ d[E_1]/dt &= +k_1[S_{00}][E_0] - (k_{-1} + k_{11})[E_1] \end{aligned} \tag{2.6}$$

With following conservation equations (again automatically fulfilled):

$$\begin{aligned} [\mathrm{S}_T] &= [S_{00}] + [S_{10}] + [S_{01}] \\ [\mathrm{E}_T] &= [E_0] + [E_1] \end{aligned} \tag{2.7}$$

Note that for the enzyme, the transformation step $[ES] \rightarrow [E] + [S^*]$ also means a release from the enzyme-substrate complex, thus the rate constants of both processes, i.e. ES dissociation (k_{-1}) and transformation (k_{11}) are added up. A reaction of enzyme unbinding from substrate is an example of **composed rate** problem. Composed rates need special treatment in the case of enzymatic regulation, as will be outlined in the Sec. 2.4.1, with the use of the sub-species correspondence concept explained in the Sec. 2.2.1.

2.2.3 Representation of synthesis and degradation

In the most basic case (Fig. 2.2 C and D), both synthesis and degradation refer to a single species A and thus do not involve any second interaction partner. This results from a semi-phenomenological treatment of these processes, where we assume the possibly participating species such as RNA polymerase, tRNA, ribosomes (for synthesis) or proteasomes, ubiquitin etc. (for degradation) to be in the cell in an abundant quantity and thus their concentrations and catalytic activity can be parameterized within a single synthesis or degradation rate constant, k_{synth} and k_{degr}, respectively. Different values of these constants for different species capture the species-specific efficiency of both processes (e.g. highly expressed vs. low expressed proteins, unstable vs. stable molecules etc.).

For this reason, the basic mathematical representation of synthesis and of degradation involves only a single rate constant:

$$d[A_{0\ldots0}]/dt = +k_{synth} \tag{2.8}$$

$$d[A_n]/dt = -k_{degr}[A_n] \tag{2.9}$$

The parameterization of the synthesis/degradation machinery components concentrations is an efficient modeling approach but it disables the incorporation of regulation of these two processes into the model. Within our framework, we propose two ways of handling the regulation of synthesis/degradation: on the entirely phenomenological level, as presented in Sec. 2.5 or on the level of molecular interactions as presented in Sec. 2.4.2 and Sec. 2.4.3.

Note, that we allow a synthesis term (2.8) only in the equation of the first, most basic sub-species $A_{0...0}$ of the species A, where all $\sigma_n^A = 0$. This important modeling assumption comes from biological reasons, as we assume that any molecule needs to be synthesized in its basic form before it can get involved in any interaction. On the other hand, a degradation term (2.9) is allowed in the equation of any sub-species A_n of the species A as we assume that a molecule can be degraded having any interaction status.

Note, that in any of the proposed approaches, both synthesis and degradation terms by definition disobey the conservation relationship $[A_T] = const.$ since they directly influence the total concentration of A.

An important calculative aspect of degradation is that it causes composed rates introduced in Sec. 2.4.1 - an unbinding rate from any partner needs to be summed up with the degradation rate of this partner.

The composed rate constants for basic interaction types are shown in Fig. 2.3. Due to introduction of degradation rates, the EMA description of these motifs need to be modified as follows:
- for binding reactions (cf. Eq.(2.1):

$$\begin{aligned} d[A_0]/dt &= -k_1[A_0][B_0] + (k_{-1} + k^B_{degr})[A_1] - k^A_{degr}[A_0] \\ d[A_1]/dt &= +k_1[A_0][B_0] - (k_{-1} + k^B_{degr})[A_1] - k^A_{degr}[A_1] \\ d[B_0]/dt &= -k_1[B_0][A_0] + (k_{-1} + k^A_{degr})[B_1] - k^B_{degr}[B_0] \\ d[B_1]/dt &= +k_1[B_0][A_0] - (k_{-1} + k^A_{degr})[B_1] - k^B_{degr}[B_1] \end{aligned} \quad (2.10)$$

- for enzymatic reactions (cf. Eq.(2.6):

$$\begin{aligned} d[S_{00}]/dt &= -k_1[S_{00}][E_0] + (k_{-1} + k^E_{degr})[S_{10}] - k^S_{degr}[S_{00}] \\ d[S_{10}]/dt &= +k_1[S_{00}][E_0] - (k_{-1} + k_{11} + k^E_{degr})[S_{10}] - k^S_{degr}[S_{10}] \\ d[S_{01}]/dt &= +k_{11}[S_{10}] - k^S_{degr}[S_{01}] \\ d[E_0]/dt &= -k_1[S_{00}][E_0] + (k_{-1} + k_{11} + k^S_{degr})[E_1] - k^E_{degr}[E_0] \\ d[E_1]/dt &= +k_1[S_{00}][E_0] - (k_{-1} + k_{11} + k^S_{degr})[E_1] - k^E_{degr}[E_1] \end{aligned} \quad (2.11)$$

In the case of degradation regulation, the composed rates need special treatment with the use of sub-species correspondence concept explained in Sec. 2.2.1. This procedure will be outlined in the Sec. 2.4.3.

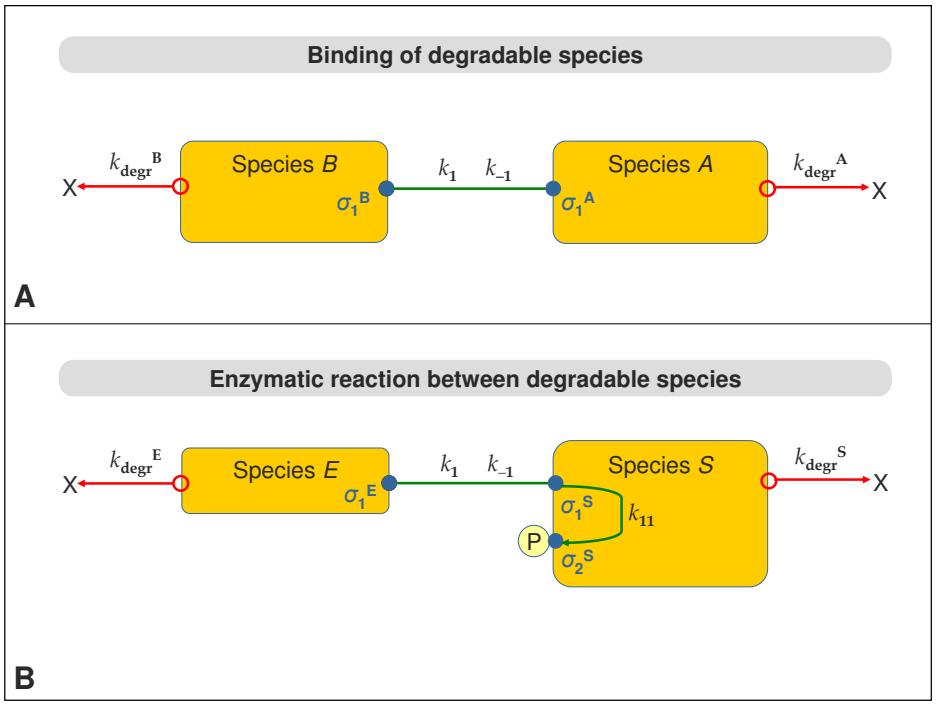
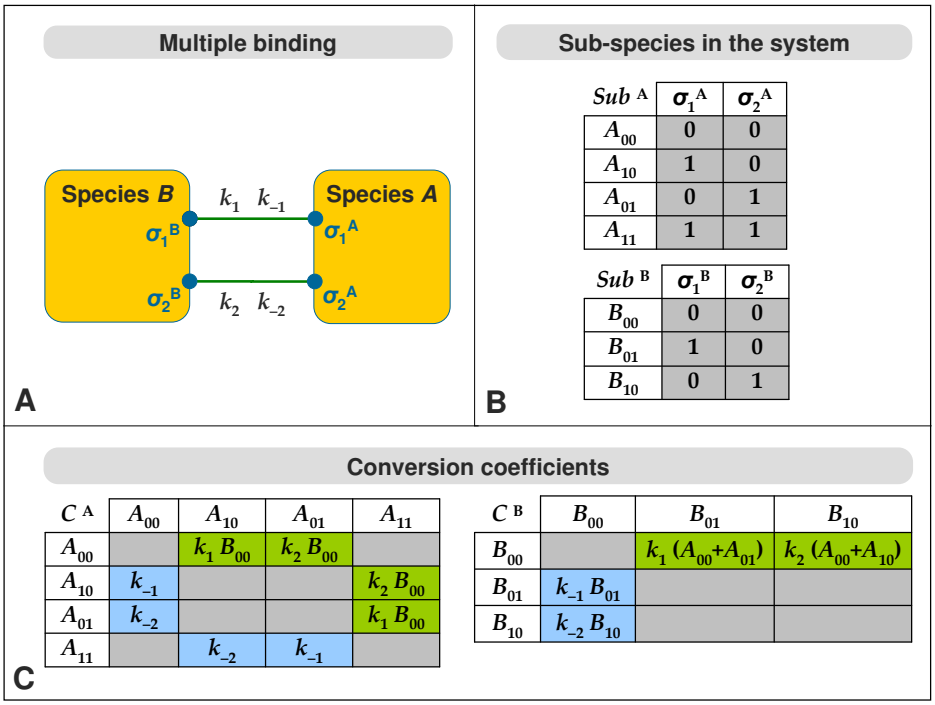


Figure 2.3: Combination of basic interactions with degradation requires summation of unbinding and degradation rate constants. Rectangles represent species, lines and arrows represent interactions, k_n denote rate constants, filled dots represent interfaces, σ_n^X denote interfaces, empty circles are introduced only for visual convenience.

2.3 Composition of basic interactions into complex modules

2.3.1 Multiple binding

The basic binding motif is easily extendable to cases, where a given species A has more than one physical binding site for the partner B, i.e. there are $n > 1$ molecules of B binding to 1 molecule of A. In this case, both interacting species (A and B) have each n interfaces. In the example shown in Fig. 2.4 A, species A has 2 binding sites for B and thus 2 interfaces σ_1^A and σ_2^A. Note, that we also assign 2 interfaces to B, although 1 molecule



Sub A	σ_1^A	σ_2^A
A_{00}	0	0
A_{10}	1	0
A_{01}	0	1
A_{11}	1	1

Sub B	σ_1^B	σ_2^B
B_{00}	0	0
B_{01}	1	0
B_{10}	0	1

C A	A_{00}	A_{10}	A_{01}	A_{11}
A_{00}		$k_1 B_{00}$	$k_2 B_{00}$	
A_{10}	k_{-1}			$k_2 B_{00}$
A_{01}	k_{-2}			$k_1 B_{00}$
A_{11}		k_{-2}	k_{-1}	

C B	B_{00}	B_{01}	B_{10}
B_{00}		$k_1 (A_{00}+A_{01})$	$k_2 (A_{00}+A_{10})$
B_{01}	$k_{-1} B_{01}$		
B_{10}	$k_{-2} B_{10}$		

Figure 2.4: Interaction scheme of partners binding at two different sites per molecule. The k_n denote rate constants, σ_n^X interfaces. **A** - graphical representation of the system, **B** - possible sub-species in the system (Sub^X) , **C** - conversion coefficients of transition between sub-species belonging to a given species (C^X) based on rate constants depicted in **A**.

of B can bind maximally 1 molecule of A. This distinction is necessary in the cases where the binding sites of A differ in terms of rate constants

or regulation mechanisms. However, the fact that B has only one physical binding site is accounted for by not allowing both interfaces of B to be simultaneously true. Thus, we can differentiate 4 sub-species of A and 3 sub-species of B as specified in Fig. 2.4 B.

Due to the number of sub-species, the easiest way to represent the EMA ODE describing the multiple binding interaction is to use the **conversion coefficient matrices** C depicted in Fig. 2.4 C. These matrices are composed for each species based on its sub-species library (Fig. 2.4 B) and interaction menu (Fig. 2.4 A). An algorithm for automatic composition of these matrices will be presented in the Sec. 3.4. In short, for every sub-species of A we identify possible conversion links to all other sub-species based on following rules:

- Conversion of sub-species A_i into sub-species A_j can only occur in the course of single interaction. Consequently, A_i and A_j can only differ by opposite values of one single interface (e.g. partner unbound $\rightarrow$ bound). One exception is the second step of enzymatic reaction, where the linked A_i and A_j must differ by opposite values of 2 interfaces (enzyme bound $\rightarrow$ unbound and enzymatic site unmodified $\rightarrow$ modified).

- The rates of conversion of A_i into A_j depend on the rate constants k_n of corresponding interactions (Fig. 2.4 A).

- For all binding reactions, the rate constants k_n are multiplied by the concentrations of partner species participating in binding. Note, that in the matrix C^B in Fig. 2.4 C, specific summation terms ($[A_{00}]+[A_{10}]$) and ($[A_{00}] + [A_{01}]$) occur. Such summation terms can be expected whenever a partner species is involved in several interactions and thus has many sub-species. Then for each interaction it has to be decided which of the many sub-species can participate in it as a sum like just described. Determination, which sub-species of the interaction partner are allowed to interact and thus contribute their concentration to the conversion coefficient is an algorithmic task, based on the conversion matrix of the partner, and it is described in the Sec. 3.4.4 in a more detail.

Based on the conversion coefficient matrices C^A and C^B from Fig. 2.4 C., we can compose following EMA ODE for each sub-species A_i (quadratic brackets indicate concentrations; they have been omitted in the C^A matrices

for visual convenience only):

$$d[A_i]/dt \quad = \quad -\Sigma_j C^A_{ij}[A_i] + \Sigma_j C^A_{ji}[A_j] \tag{2.12}$$

Note, that in this representation (Eq. 2.12), each term from the conversion matrix C_A enters the equation twice - once with a positive and once with a negative sign. This illustrates best, that all the EMA equations sum up to zero and thus automatically fulfill conservation relationships that can be defined in the most general form as follows:

$$[\mathrm{A}_T] \quad = \quad \Sigma_i A_i \tag{2.13}$$

Another important relationship that we can derive algorithmically from the structure of conversion matrices C^A and C^B from Fig. 2.4 C. is the correspondence of sub-species sets between A and B. Here we have following correspondence relationships: $[B_{10}] = [A_{10}] + [A_{11}]$ and $[B_{01}] = [A_{01}] + [A_{11}]$. The algorithm to derive such relationships is described in Sec. 3.4.4.

2.3.2 Coupled forward and backward enzymatic reactions

Description of a basic enzymatic reaction $[E] + [S] \leftrightharpoons [ES] \rightarrow [E] + [S^*]$ is outlined in Sec. 2.2.2. As mentioned there, the conversion of product $[S^*]$ back to substrate $[S]$ by an enzyme different from E is perfectly possible, and actually very common in nature, with the most prominent example of phosphorylation and dephosphorylation. A basic motif consisting of two coupled enzymatic reactions proceeding in two opposite directions is depicted in the Fig. 2.5 where the forward reaction enzyme (e.g. a kinase) is labeled $E1$ and the backward reaction enzyme (e.g. a phosphatase) is labeled $E2$.

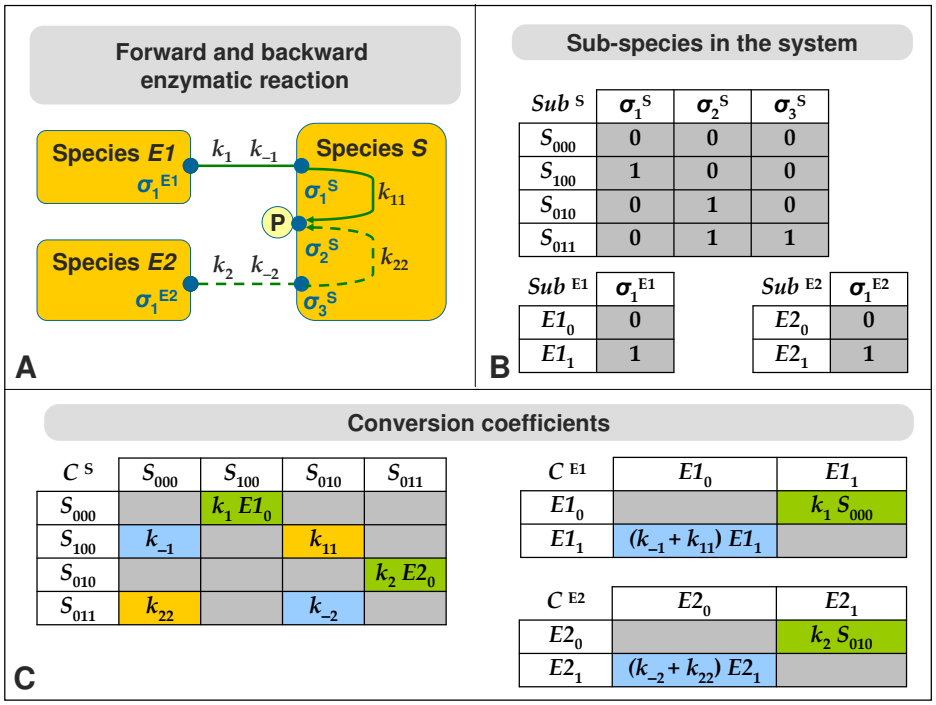


Figure 2.5: Coupled forward and backward enzymatic reactions. The k_n denote rate constants, σ_n^X interfaces. Dashed lines depict a backward enzymatic reaction, e.g. dephosphorylation. **A** - graphical representation of the system, **B** - possible sub-species in the system (Sub^X) , **C** - conversion coefficients of transition between sub-species belonging to a given species (C^X) based on rate constants depicted in **A**.

The reaction scheme can be set up analogically to (2.5):

$$[\mathrm{E1}] + [\mathrm{S}] \quad \leftrightharpoons \quad [E1S] \rightarrow [E1] + [S^*] \tag{2.14}$$

$$[\mathrm{E2}] + [\mathrm{S}^*] \quad \leftrightharpoons \quad [E2S^*] \rightarrow [E2] + [S] \tag{2.15}$$

As shown in Fig. 2.5 B, the species $E1$ has sub-species $E1_0$ and $E1_1$ identical to E_0 and E_1 in the example of a basic enzymatic reaction in Sec. 2.2.2. Analogically, the 2 sub-species of $E2$ correspond to free $E2$ ($E2_0$) and to the enzyme-substrate complex $E2S^*$ ($E2_1$, since the product S^* is actually a substrate of $E2$). The addition of $E2$ into the system requires adding a new interface σ_3^S to the species S, which results in 4 sub-species of S:

- S_{000} ($\sigma_1^S = 0$, $\sigma_2^S = 0$, $\sigma_3^S = 0$) - free fraction of substrate S.
- S_{100} ($\sigma_1^S = 1$, $\sigma_2^S = 0$, $\sigma_3^S = 0$) - enzyme-substrate complex $E1S$.
- S_{010} ($\sigma_1^S = 0$, $\sigma_2^S = 1$, $\sigma_3 = 0$) - transformed fraction of substrate S^*.
- S_{011} ($\sigma_1^S = 0$, $\sigma_2^S = 1$, $\sigma_3^S = 0$) - enzyme-product complex $E2S^*$.

Here we have following correspondence relationships: $[E1_1] = [S_{100}]$ and $[E2_1] = [S_{011}]$. Note, that sub-species $E1S^*$ ($\sigma_1^S = 1$, $\sigma_2^S = 1$) is excluded for the same reasons as in the previous example, and for $E2$ we analogically do not allow a sub-species $E2S$ ($\sigma_1^S = 0$, $\sigma_3^S = 1$). We also additionally assume, that two enzymes targeting the same enzymatic modification site cannot bind simultaneously to the same substrate, which excludes states $E1E2S$ ($\sigma_1^S = 1$, $\sigma_2^S = 0$, $\sigma_3^S = 1$) and $E1E2S^*$ ($\sigma_1^S = 1$, $\sigma_2^S = 1$, $\sigma_3^S = 1$. The algorithm for composing sub-species lists (Sec. 3.4.2) and conversion matrices (Sec. 3.4.3) takes these restrictions into account. The EMA ODE describing the whole system can be derived from the conversion coefficient matrices presented in Fig. 2.5 B using the universal equation 2.12.

2.3.3 Multiple coupled enzymatic reactions

In analogy to extending the basic binding reaction scheme (Sec. 2.2.1) to multiple binding (Sec. 2.3.1), the forward-backward enzymatic reaction motif (Sec. 2.3.2) is extensible to several cases discussed below.

Substrate with independent enzymatic reactions.

The simplest extension is to pool several independent enzymatic modification sites within one substrate. An example in Fig. 2.6 shows such substrate with two modification sites.

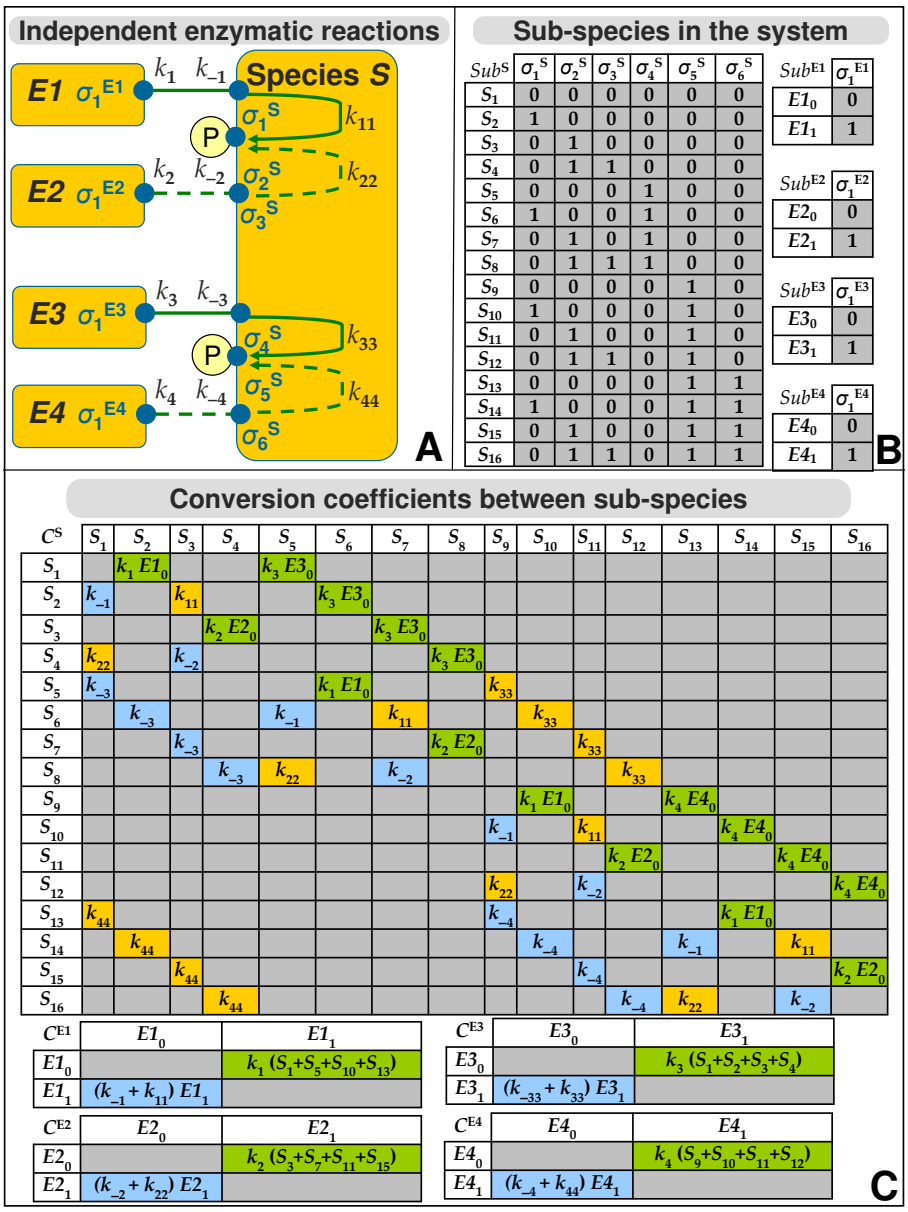


Figure 2.6: Substrate with several independent enzymatic modification sites. The k_n denote rate constants, σ_n^X interfaces. Dotted lines depict a backward enzymatic reaction, e.g. dephosphorylation. **A** - graphical representation of the system, **B** - possible sub-species in the system (Sub^X) , **C** - conversion coefficients of transition between sub-species belonging to a given species (C^X) based on rate constants depicted in **A**.

The substrate S has 16 possible sub-species that are independent combinations of the four basic sub-species $S_1 - S_4$ of the forward-backward motif presented in Sec 2.3.2, which already demonstrates combinatorial complexity of even relatively simple systems. A regular pattern of conversion matrix C_S (Fig. 2.6 C) indicates the independent character of the two enzymatic sites. This pattern can be broken by adding a single regulatory relation between the enzymatic modification sites as presented in the matrix C^S from (Fig. 2.6 C) describing a allosteric regulation of enzymatic reactions.

Competition between enzymes over a substrate.

Another possible combination of enzymatic reactions is a system where several enzymes target the same modification site (Fig. 2.7).

In this case, the library of allowed sub-species is less numerous as in the previous example (6 vs. 16), because some substrate sub-species are excluded due to following rules:

- Binding of a modifying enzyme to an already modified substrate (esp. a kinase to a phosphorylated substrate) or an unmodifying enzyme to the substrate (esp. a phosphatase to an unphosphorylated substrate), as explained in Sec 2.3.2.

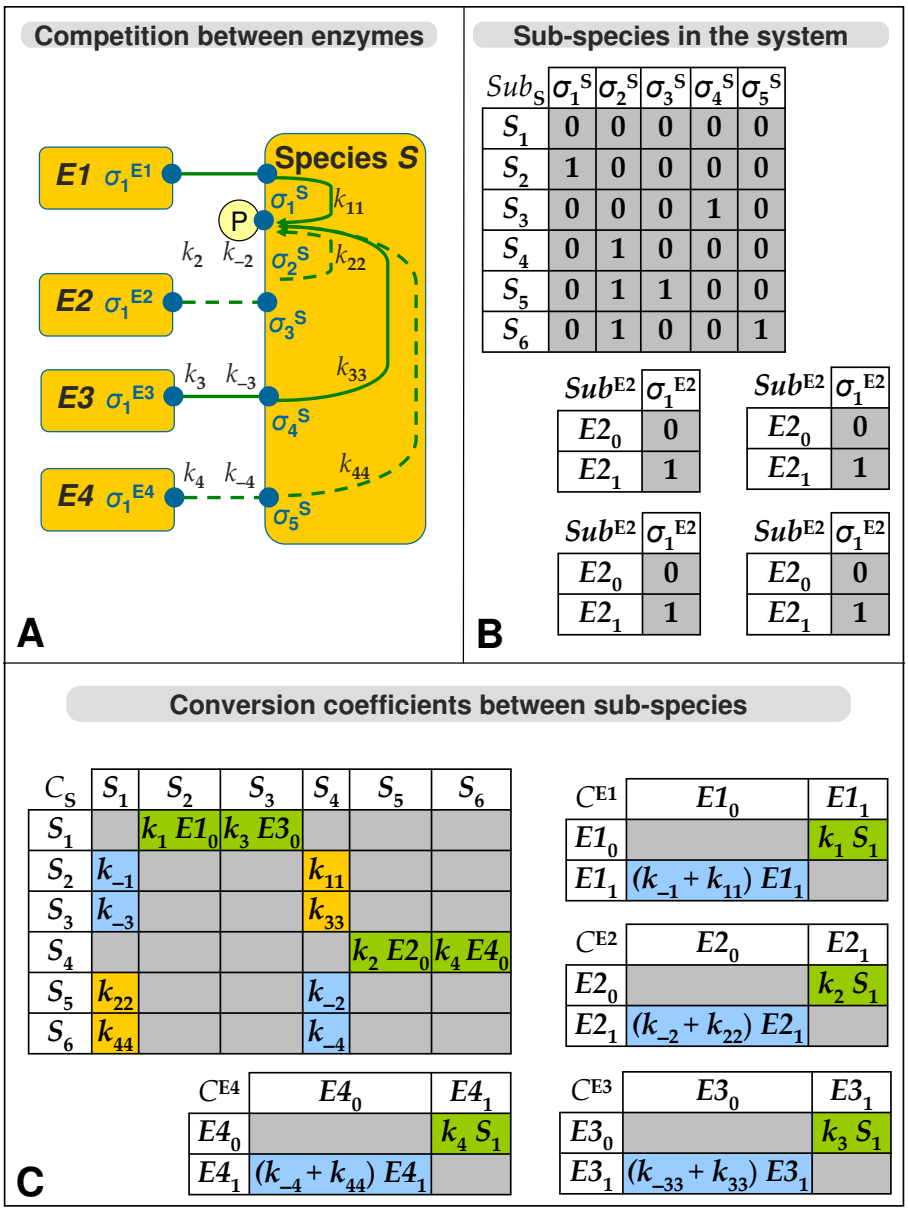
- Simultaneous binding to the substrate of more than one enzyme acting on the same modification site. If that was allowed, one of the competing enzymes would modify the substrate, unbind and leave other enzymes still attached to an already modified substrate, which was forbidden in the previous point.

These simple exclusion rules reduce combinatorial complexity of systems containing competing enzymes and they are taken into account by the equation-composing algorithm described in Sec. 3.4.

Competition between substrates over an enzyme.

A reverse situation is an enzyme targeting several substrates (Fig. 2.8).

This case automatically fulfills the scheme of competitive inhibition, especially, when the affinity of enzyme to substrate S_1 is significantly higher than to the substrate S_2. This means, that the enzyme-S_1 complex is more stable: $k_1 >> k_3$ and $(k_{-1} + k_{11}) << (k_{-3} + k_{33})$ and thus blocks the S_2 from being processed. This can especially happen when S_1 is 'mimicking' S_2 having a very similar spatial structure of the modification site that 'locks'



Sub_S	σ_1^S	σ_2^S	σ_3^S	σ_4^S	σ_5^S
S_1	0	0	0	0	0
S_2	1	0	0	0	0
S_3	0	0	0	1	0
S_4	0	1	0	0	0
S_5	0	1	1	0	0
S_6	0	1	0	0	1

Sub^{E2}	σ_1^{E2}
$E2_0$	0
$E2_1$	1

Sub^{E2}	σ_1^{E2}
$E2_0$	0
$E2_1$	1

Sub^{E2}	σ_1^{E2}
$E2_0$	0
$E2_1$	1

Sub^{E2}	σ_1^{E2}
$E2_0$	0
$E2_1$	1

C_S	S_1	S_2	S_3	S_4	S_5	S_6
S_1		$k_1\ E1_0$	$k_3\ E3_0$			
S_2	k_{-1}			k_{11}		
S_3	k_{-3}			k_{33}		
S_4					$k_2\ E2_0$	$k_4\ E4_0$
S_5	k_{22}			k_{-2}		
S_6	k_{44}			k_{-4}		

C^{E1}	$E1_0$	$E1_1$
$E1_0$		$k_1\ S_1$
$E1_1$	$(k_{-1} + k_{11})\ E1_1$	

C^{E2}	$E2_0$	$E2_1$
$E2_0$		$k_2\ S_1$
$E2_1$	$(k_{-2} + k_{22})\ E2_1$	

C^{E4}	$E4_0$	$E4_1$
$E4_0$		$k_4\ S_1$
$E4_1$	$(k_{-4} + k_{44})\ E4_1$	

C^{E3}	$E3_0$	$E3_1$
$E3_0$		$k_3\ S_1$
$E3_1$	$(k_{-33} + k_{33})\ E3_1$	

Figure 2.7: Competition between enzymes targeting the same modification site within a single substrate. The k_n denote rate constants, σ_n^X interfaces. Dotted lines depict a backward enzymatic reaction, e.g. dephosphorylation. **A** - graphical representation of the system, **B** - possible sub-species in the system (Sub^X) , **C** - conversion coefficients of transition between sub-species belonging to a given species (C^X) based on rate constants depicted in **A**.

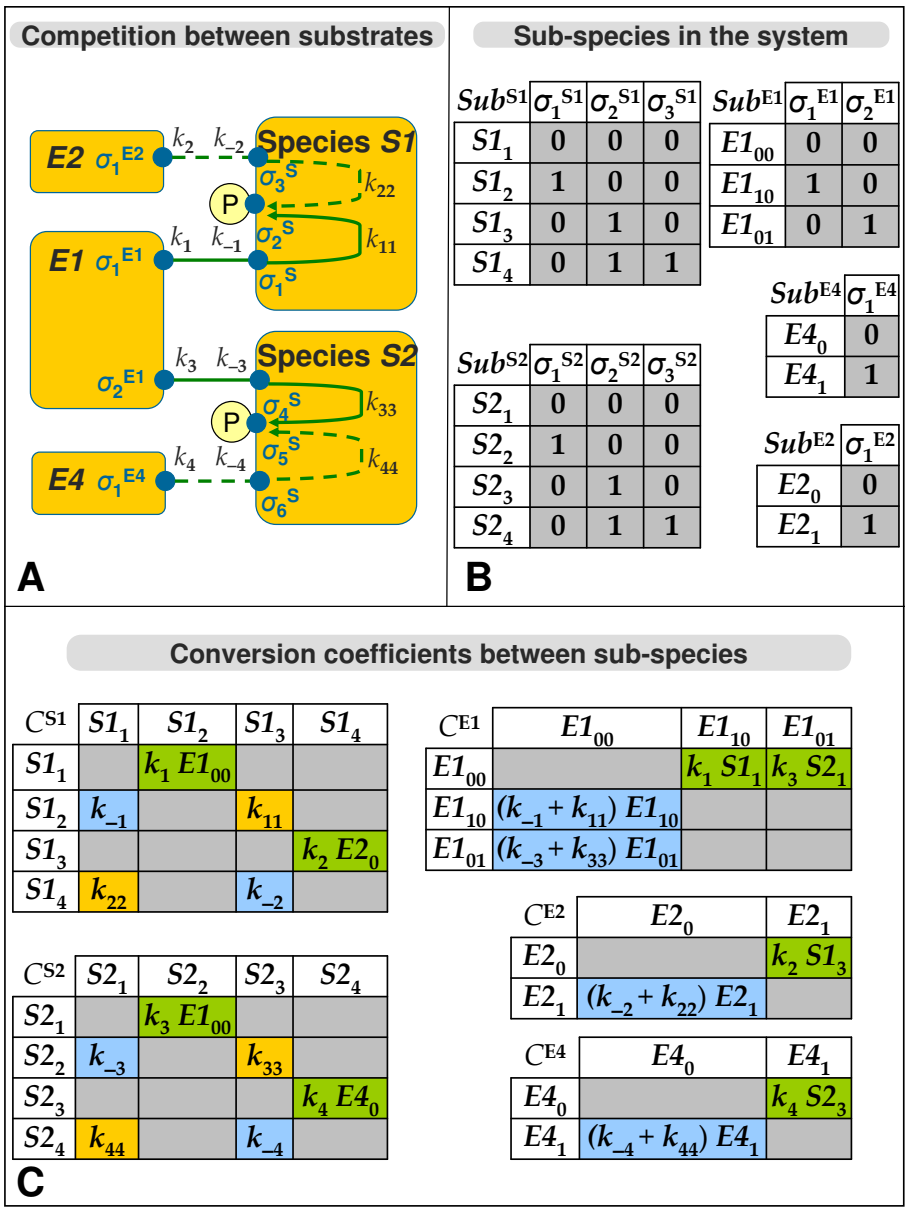


Figure 2.8: Competition between substrates over a single enzyme. The k_n denote rate constants, σ_n^X interfaces. Dotted lines depict a backward enzymatic reaction, e.g. dephosphorylation. The k_n denote rate constants. **A** - graphical representation of the system, **B** - possible sub-species in the system (Sub^X) , **C** - conversion coefficients of transition between sub-species belonging to a given species (C^X) based on rate constants depicted in **A**.

the active site of the enzyme but disables its dissociation from S_1, in such case of perfect inhibition $k_{11} = 0$. In an even more extreme case, both $k_{11} = 0$ and $k_{-1} = 0$ and then the inhibition is irreversible.

In the example presented in (Fig. 2.8 B), the multi-substrate enzyme $E1$ has thus 3 possible sub-species: E_{00} - free enzyme, E_{10} - enzyme-substrate complexes with S_1, E_{01} - enzyme-substrate complexes with S_2. A simultaneous binding of the enzyme to more than one substrate is not allowed, since we assume this is also commonly the case in nature.

Multi-site enzymatic reaction.

Finally, we can imagine a situation when a single enzyme targets several sites within a single substrate (Fig. 2.9).

Note, that we have created two different interfaces σ_1^S and σ_4^S for the same enzyme, since it can potentially bind at the two different sites of the substrate with two different constants ($k_1 \neq k_1$). However, for simplicity, we will not allow sub-species where such binding occurs simultaneously.

This situation is a hybrid of two already described cases. The substrate S with two sites has almost the same sub-species as in the independent site case (Fig. 2.6 B), with the exclusion of the sub-species S_6 there, which has a not allowed interface value combination $\sigma_1^S = 1$ and $\sigma_4^S = 1$. On the other hand, the enzyme has the same sub-species as in the competition over enzyme case (Fig. 2.8 B) and if reaction constants for both sites also invoke substantially different affinities, like: $k_1 >> k_3$ and $(k_{-1} + k_{11}) << (k_{-3} + k_{33})$ or the other way round), a specific case of competitive inhibition is also possible.

However, an opposite situation occurs often in nature, where an enzymatic modification on one site is activating, not inhibiting the modification of the other site of the same substrate via a positive regulatory link. This situation is presented in (Sec. 2.4.1, Par. 'Cooperative multi-site enzymatic reaction').

This and previous examples illustrate the potential of our approach to describe complex network motifs by modular composition of basic motifs with use of simple rules. Following, we will show that our framework can also capture further important features of biological systems, such as regulation.

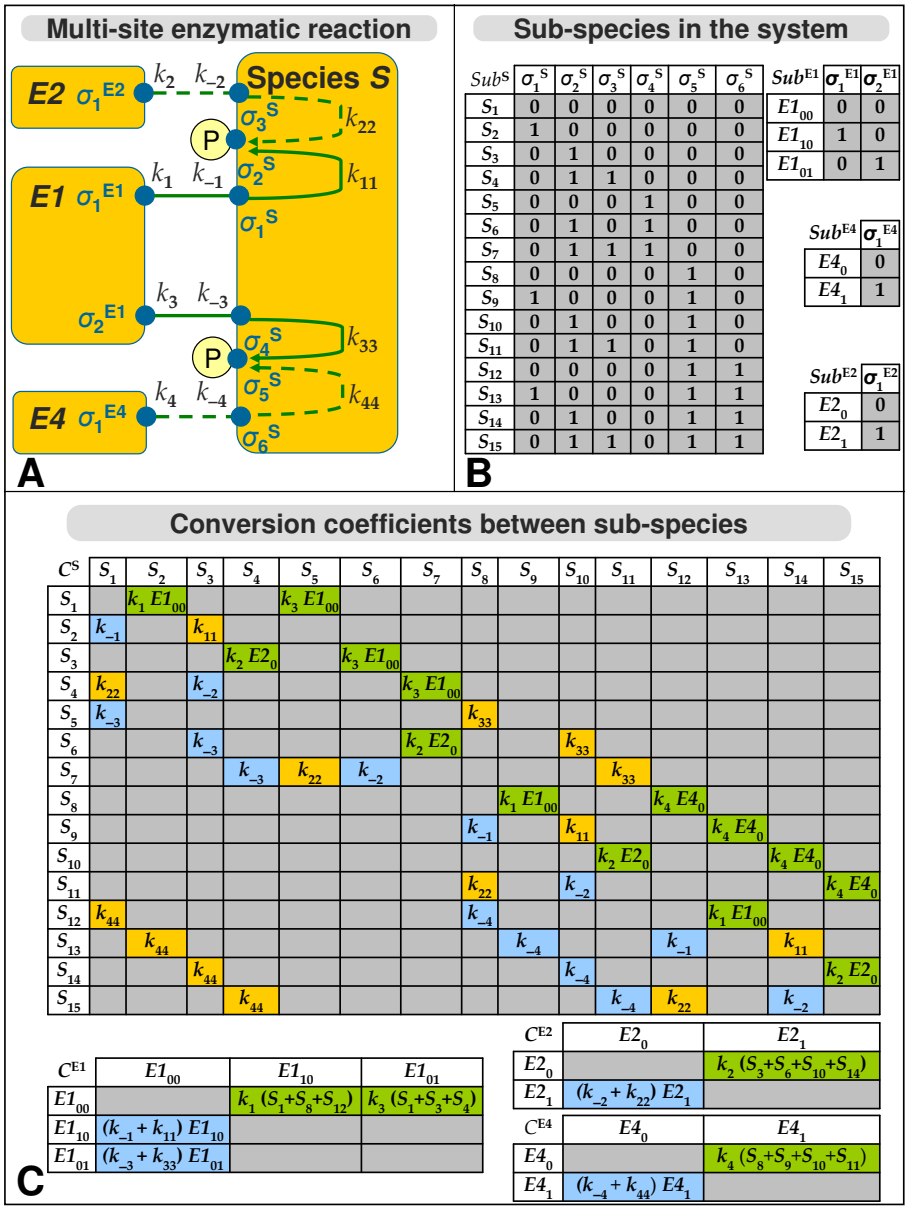


Figure 2.9: Multi-site enzymatic reaction, where a single substrate has several enzymatic modification sites that compete over a single enzyme. The k_n denote rate constants, σ_n^X interfaces. Dotted lines depict a backward enzymatic reaction, e.g. dephosphorylation. **A** - graphical representation of the system, **B** - possible sub-species in the system (Sub^X) , **C** - conversion coefficients of transition between sub-species belonging to a given species (C^X) based on rate constants depicted in **A**.

2.4 Representation of regulation

In Sec. 2.1.2, we have defined regulation as a relation between two interfaces of the same species, σ_i and σ_j, such that the rate constant k_j related to σ_j can change depending on the state of σ_i. We will call a species containing σ_i and σ_j a **host**, a species interacting with host at σ_i a **master** and a species interacting at σ_j a **slave**. Restriction of regulation to a process taking place within one host comes from an assumption that it can be only exhibited by direct (though not necessarily simultaneous) physical contact between the molecules of host and master and between the molecules of host and slave. In most cases, the regulatory influence corresponds to a physical process. For example, interaction with master at one physical site (corresponding to σ_i) might invoke conformational changes of the host molecule that affect the shape and thus properties of its other physical site where the slave interacts resulting in an altered binding affinity at σ_j and thus a new value of k_j.

In our approach, regulation is translated into EMA ODE by multiplication of k_j with some regulation coefficient α. A given k_j can be potentially regulated by a number of masters. This is specified by a Boolean matrix $\mathbf{R}$ and a corresponding matrix α of dimensions (n), $(n+2)$, where n equals to the host's number of interfaces and 2 stands for the possibility to assign a slave status to a synthesis or a degradation rate of host. An entry $R_{ij} = true$ means that the interaction at σ_i regulates the interaction at σ_j by a coefficient α_{ij}. We propose distinguishing 4 types of regulatory links, as presented in the Fig. 2.10 A:

- exclusion: $\alpha_{ij} = 0$, if: $\sigma_i R_{ij} = 1$, else: $\alpha_{ij} = 1$.
- inhibition: $\alpha_{ij} = \alpha < 1$, if: $\sigma_i R_{ij} = 1$, else: $\alpha_{ij} = 1$.
- activation: $\alpha_{ij} = \alpha > 1$, if: $\sigma_i R_{ij} = 1$, else: $\alpha_{ij} = 1$.
- necessity: $\alpha_{ij} = 0$, if: $(1 - \sigma_i) R_{ij} = 1$, else: $\alpha_{ij} = 1$.

2.4.1 Regulation of binding and enzymatic reactions

Regulation of binding is depicted in (Fig. 2.10). For binding reactions we assume that the regulatory modification affects the binding affinity and thus the binding rate k_1 (Sec. 2.2.1) to a slave partner ($k_1 = k_j$):

$$k_1^* = k_1 \Pi_i \alpha_{ij} \tag{2.16}$$

The unbinding rate k_{-1} is left unchanged, otherwise the regulation would influence reaction velocity but not affinity.

Regulation of enzymatic reactions is depicted in (Fig. 2.11). For enzymatic reactions, we take a simplifying assumption that both binding rate k_1 and modification rate k_{11} belonging to a slave enzymatic reaction are multiplied by the same regulation coefficient α, which changes both the enzyme-substrate affinity and the overall speed of enzymatic conversion:

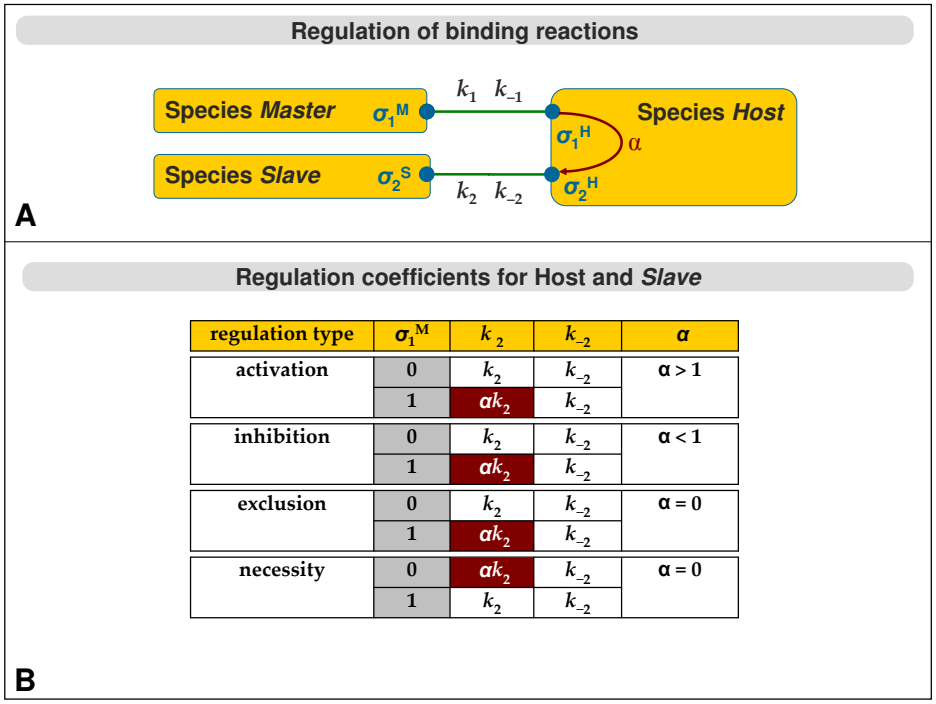
$$k_1^* = k_1 \Pi_i \alpha_{ij} \tag{2.17}$$

$$k_{11}^* = k_{11} \Pi_i \alpha_{ij} \tag{2.18}$$

Again, the backward rate k_{-1} remains unchanged. Thus, to keep consistency with the mathematical description of enzymatic reactions defined in Section 2.2.2, the unbinding rate of the enzyme, $(k_{-1} + k_{11})$, is transformed under regulatory influence to $(k_{-1} + k_{11}^*)$.

Further, since enzymatic reactions involve two interfaces of the substrate: for enzyme binding σ_1^S and for enzymatic modification σ_2^S, we make a simplifying assumption that an enzyme can execute a master regulatory role only via the modification site, thus only $\sigma_2^S = \sigma_i$ and the state of σ_1^S is irrelevant here, i.e. the regulation does not occur in enzyme-substrate complexes before the modification step. Note, that it automatically implies, that if more enzymes target the same modification site σ_2^S, they can all play the same master role towards an interaction with a given slave. This involvement is irrespective of if such enzymes act in a forward or a backward modification reaction, like phosphorylation and dephosphorylation, respectively.

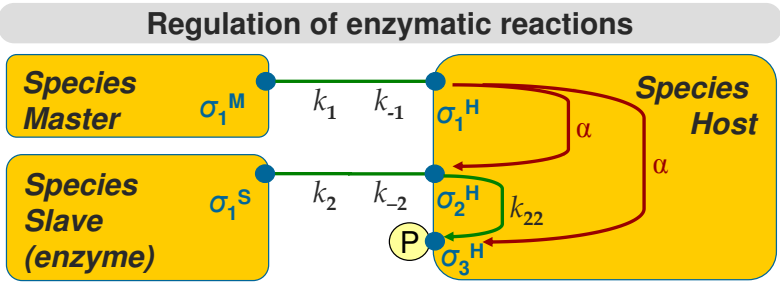
The host-slave-master approach outlined above enables an uniform treatment of various specific regulation cases common in nature, such as cooperativity, allostery, regulation of transcription and regulation of degradation, as presented in the following sections.



regulation type	σ_1^M	k_2	k_{-2}	α
activation	0	k_2	k_{-2}	α > 1
	1	αk_2	k_{-2}	
inhibition	0	k_2	k_{-2}	α < 1
	1	αk_2	k_{-2}	
exclusion	0	k_2	k_{-2}	α = 0
	1	αk_2	k_{-2}	
necessity	0	αk_2	k_{-2}	α = 0
	1	k_2	k_{-2}	

Figure 2.10: Graphical representation of regulation of binding reactions. Rectangles represent species, green lines represent interactions, k_n denote rate constants, purple arrows represent regulation, α_n denote regulation coefficients, filled dots represent interfaces, σ_n^X denote interfaces.

A



B

Regulation coefficients for *Host*

regulation type	σ_1^M	k_2	k_{-2}	k_{22}	α
activation	0	k_2	k_{-2}	k_{22}	α > 1
	1	αk_2	k_{-2}	αk_{22}	
inhibition	0	k_2	k_{-2}	k_{22}	α < 1
	1	αk_2	k_{-2}	αk_{22}	
exclusion	0	k_2	k_{-2}	k_{22}	α = 0
	1	αk_2	k_{-2}	αk_{22}	
necessity	0	αk_2	k_{-2}	αk_{22}	α = 0
	1	k_2	k_{-2}	k_{22}	

C

Regulation coefficients for *Slave*

regulation type	σ_1^M	k_2	k_{-2}	α
activation	0	k_2	$(k_{-2} + k_2)$	α > 1
	1	αk_2	$(k_{-2} + \alpha k_2)$	
inhibition	0	k_2	$(k_{-2} + k_2)$	α < 1
	1	αk_2	$(k_{-2} + \alpha k_2)$	
exclusion	0	k_2	$(k_{-2} + k_2)$	α = 0
	1	αk_2	$(k_{-2} + \alpha k_2)$	
necessity	0	αk_2	$(k_{-2} + \alpha k_2)$	α = 0
	1	k_2	$(k_{-2} + k_2)$	

Figure 2.11: Graphical representation of regulation of enzymatic reactions. Rectangles represent species, green lines and arrows represent interactions, k_n denote rate constants, purple arrows represent regulation, α_n denote regulation coefficients, filled dots represent interfaces, σ_n^X denote interfaces.

Cooperative binding

Cooperative binding (Fig.2.12) is an extension of the multiple binding case presented in Sec. 2.3.1 by linking the interaction sites σ_1^A and σ_2^A with a regulatory relationship, such that $\sigma_1^A = \sigma_i$ (master) and $\sigma_2^A = \sigma_j$ (slave) (Fig.2.12 A). Note, that in this case both σ_i and σ_j refer to the same interaction partner B which is thus simultaneously a master and a slave.

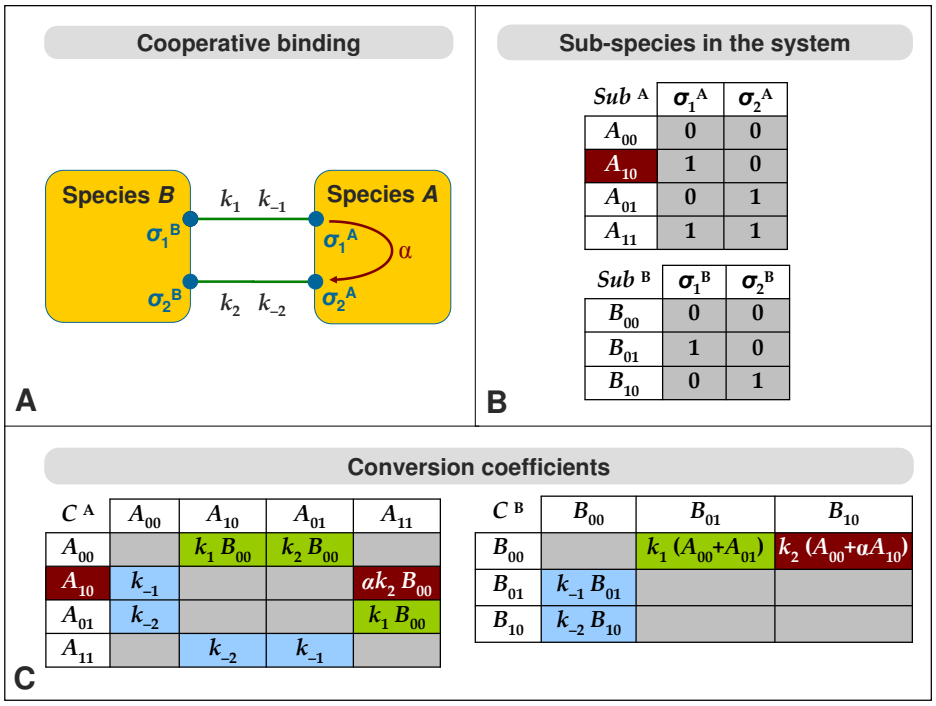


Figure 2.12: Cooperative regulation of binding. The k_n denote rate constants, α_n regulation coefficients, σ_n^X interfaces. **A** - graphical representation of the system, **B** - possible sub-species in the system (Sub^X) , **C** - conversion coefficients of transition between sub-species belonging to a given species (C^X) based on rate constants and regulation coefficients depicted in **A**.

Potentially, a regulatory behavior can occur in all those sub-species of host A where:

- $\sigma_i = 1$, so that the master interaction has taken place.
- $\sigma_j = 0$, so that the slave interaction can potentially take place in the course of a transition to another sub-species where $\sigma_j = 1$.

In the given example, only one sub-species, A_{10} meets these requirements (Fig.2.12 B). Consequently, all binding rate constants related to A_{10} in the conversion matrix C^A shown in Fig. 2.12 C are modified. In this case, there is only one such term: $C^A_{2:4} = \alpha k_2 B_{00}$. It is obtained by multiplying term $C^A_{2:4}$ from the conversion matrix describing a case of multiple binding without regulation (Fig. 2.4 C) with the regulation coefficient α. Note, that for visual convenience we omit in this example regulation by necessity, where we would simply have to consider the sub-species $_{00}$ ($\sigma_i = 0$ and $\sigma_j = 0$) instead of A_{10}.

Another important notion is that α needs also to enter the conversion matrix C^B whenever A_{10} appears too, in this case in the term $C^B_{1:3} = k_2(A_1 + \alpha A_2)$ (Fig. 2.12 C).

Cooperative multi-site enzymatic reaction

similarly to cooperative binding, cooperativity can also occur in enzymatic reactions with multiple modification sites, as shown in Fig. 2.13 A. The multi-site enzymatic reaction described in Sec. 2.3.1 is supplemented here with a regulatory link between the enzymatic modification sites σ^S_2 and σ^S_5, such that $\sigma^S_2 = \sigma_i$ (master) and $\sigma^S_5 = \sigma_j$ (slave). From this automatically follows, that $\sigma^S_4 = \sigma_j$ too, as explained in Sec. 2.4.1).

In this example case, there are several sub-species of S that fulfill the condition $\sigma_i = 1$ and $\sigma_j = 0$; these are: S_3, S_4, S_6 and S_7 as specified in Fig. 2.13 B. Accordingly, rate constants associated to these sub-species are multiplied with α. This is evident from a comparison OF the conversion matrix terms $C^S_{3:6}$, $C^S_{4:7}$, $C^S_{6:10}$ and $C^S_{7:11}$ shown in Fig. 2.13 C to the respective terms of matrix C^S in Fig. 2.9 C, which describes a corresponding, unregulated case (Sec. 2.3.3).

Since the regulation is also related to the master-slave enzyme $E1$, the regulatory coefficient α also needs to enter the conversion matrix C^{E1}. For the regulated $E1 - S$ binding reactions, α appears simply whenever the regulated sub-species S_3 and S_4 do, i.e. in the term $C^{E1}_{1:3}$, in analogy to the cooperative binding case shown in Fig. 2.12.

This is, however, more complicated for the regulated $E1 - S$ unbinding reaction in the conversion matrix term $C^{E1}_{3:1} = (k_{-3} + \beta k_{33})$, since this case represents a composed rate problem. The unbinding rates of enzymes are composed of two terms: substrate unbinding rate constant k_{-3} and substrate modification rate constant k_{33}, from which only k_{33} is regulated (Sec. 2.2.2). The regulation of such composed rates can't be described in terms of EMA

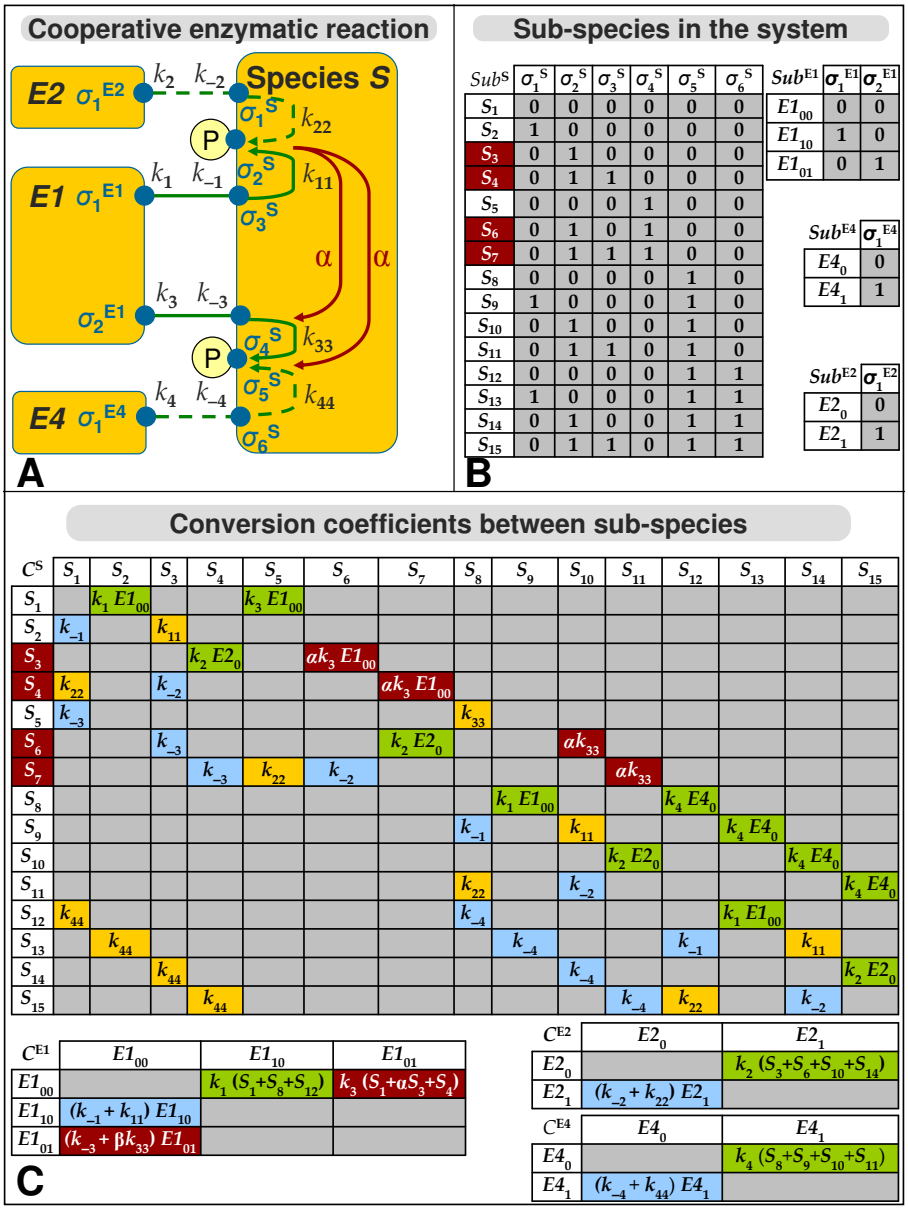


Figure 2.13: Cooperative multi-site enzymatic reaction. The k_n denote rate constants, α_n regulation coefficients, σ_n^X interfaces. Dotted lines depict a backward enzymatic reaction, e.g. dephosphorylation. **A** - graphical representation of the system, **B** - possible sub-species in the system (Sub^X) , **C** - conversion coefficients of transition between sub-species belonging to a given species (C^X) based on rate constants and regulation coefficients depicted in **A**.

using the original regulation coefficient α, because that would violate mass conservation. Namely, in the term $C^{E1}_{3:1}$ the regulated rate k_{33} is multiplied with a concentration term $[E1_{01}]$. Based on the species correspondence concept outlined in Sec. 2.2.1: $[E1_{01}] = [S_5] + [S_6] + [S_7]$, where only $[S_6]$ and $[S_7]$ are regulated via α. Based on this, we can calculate a new EMA-compliant regulatory coefficient β as follows:

$$\begin{aligned}
[\mathrm{E1_3}]\ (\mathrm{k}_{-5} + \beta k_{55}) &= [E1_3]k_{-5} + [E1_3]\beta k_{55} \\
[\mathrm{E1_3}]\ \mathrm{k}_{-5} + [E1_3]\beta k_{55} &= [E1_3]k_{-5} + ([S_5] + \alpha[S_6] + \alpha[S_7])k_{55} \\
([S_5] + [S_6] + [S_7])\beta k_{55} &= [E1_3]k_{-5} + ([S_5] + \alpha[S_6] + \alpha[S_7])k_{55} \\
\\
\beta &= \frac{[S_5] + \alpha[S_6] + \alpha[S_7]}{[S_5] + [S_6] + [S_7]} \qquad (2.19)
\end{aligned}$$

From the above derivation we immediately see, that β is simply a dimensionless weighted average of α where the weights are the concentrations of corresponding sub-species. This situation illustrates well the application of sub-species correspondence concept introduced in Sec. 2.2.1. The composed unbinding rates resulting from regulated degradation are handled in identical way as will be shown in Sec. 2.4.3. The algorithm for calculation of β coefficients and their incorporation into the equation is presented in Sec. 3.4.4 and Sec. 3.5.2.

Allosteric regulation

Allosteric regulation is another regulation mechanism of enzymatic activity common in nature, it relies on binding a regulatory molecule (master) to an enzyme (slave) at a site different than the enzyme's active site. Thus, it is similar to cooperativity with the difference that the regulatory molecule is always a species different from the slave enzyme. It can be either a binding partner or a modifying enzyme; here we present the later example (Fig.2.14).

An allosteric regulation mechanism of one enzymatic reaction by another can be well derived from the basic example of substrate with independent enzymatic reactions presented in Sec. 2.3.1 that was modified by linking the enzymatic modification sites σ_2 and σ_5 with a regulatory relationship, such that $\sigma_2 = \sigma_i$ and $\sigma_5 = \sigma_j$ (additionally $\sigma_4 = \sigma_j$ too, as explained in Sec. 2.4). This results in modifying the values of some rate constants in the conversion matrices shown in Fig. 2.14 C in comparison to the original matrices from Fig. 2.6 C. The terms $C^S_{3:7}$, $C^S_{4:8}$, $C^S_{7:11}$, $C^S_{8:12}$ and $C^{E3}_{1:2}$ are

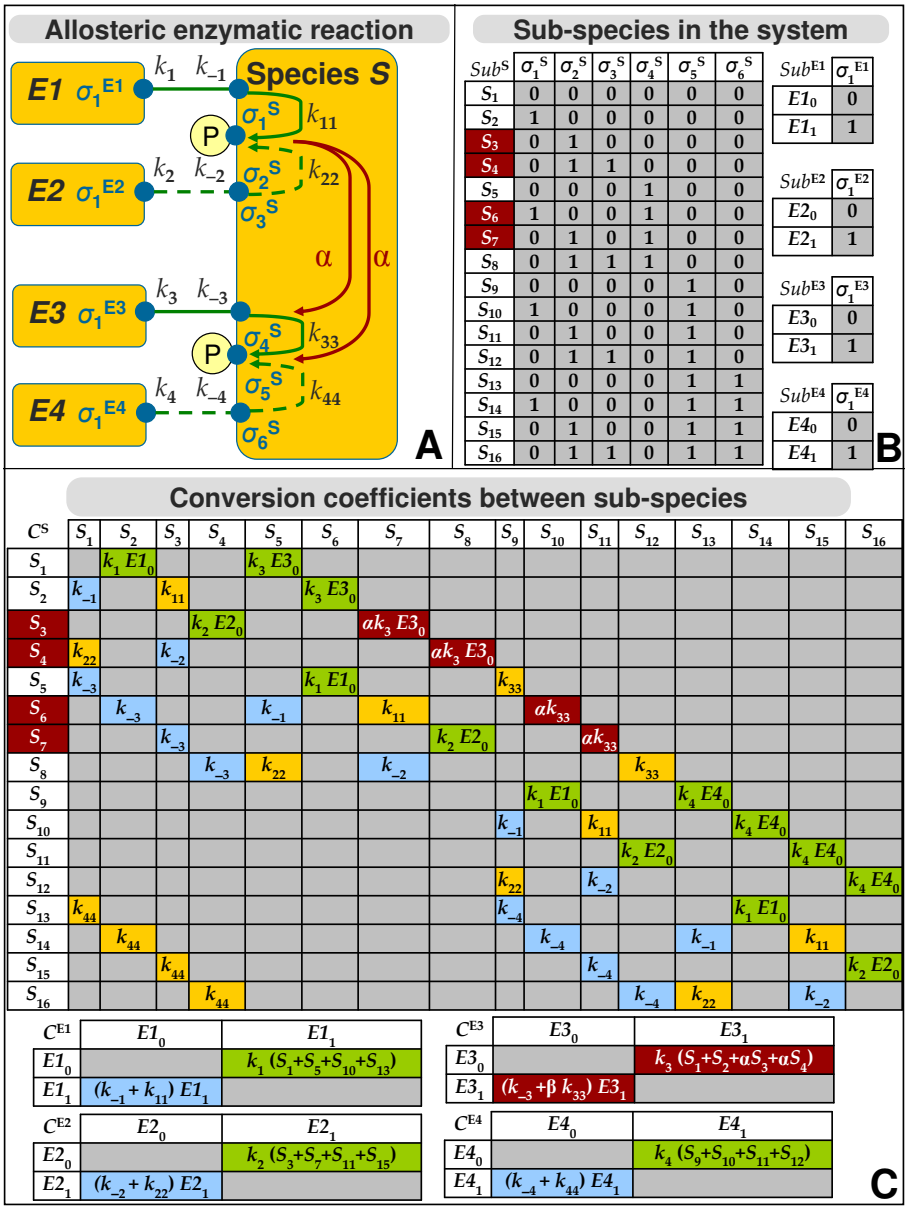


Figure 2.14: Allosteric regulation of enzymatic reactions. The k_n denote rate constants, α_n regulation coefficients, σ_n^X interfaces. Dotted lines depict a backward enzymatic reaction, e.g. dephosphorylation. **A** - graphical representation of the system, **B** - possible sub-species in the system (Sub^X) , **C** - conversion coefficients of transition between sub-species belonging to a given species (C^X) based on rate constants and regulation coefficients depicted in **A**.

modified with the regulation coefficient α and the term $C_{2:1}^{E3}$ is modified with the regulation coefficient β. In analogy to Eq. (2.19), $\beta = ([S_5] + [S_6] + \alpha[S_7] + \alpha[S_8])/([S_5] + [S_6] + [S_7] + [S_8])$ here.

2.4.2 Regulation of protein synthesis

The regulation of protein synthesis can also be described using the host-slave-master approach introduced in Sec. 2.4.3. For this purpose, for a given protein species A, we introduce a corresponding **gene** species gA that is a host of regulation. This approach relies on following assumptions:

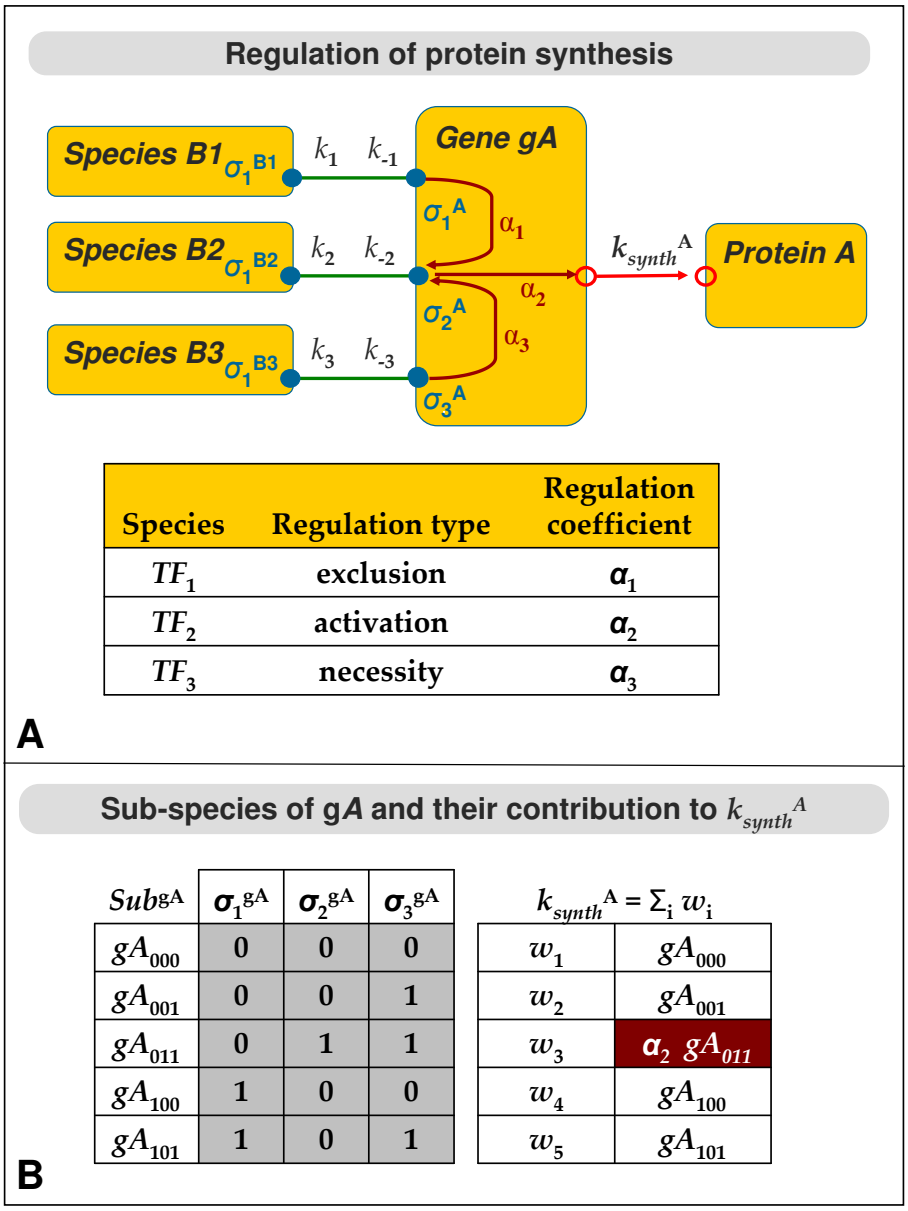
- For a given protein species A, there is only one corresponding gene species gA with fixed total concentration $[gA_T] = 1.0$. Thus, the concentration of of any subspecies $[gA_j]$ can be considered as percentage fraction: $[gA_j] = [gA_j]/[gA_T]$.
- Synthesis of A, as described in the Eq. 2.8, does not necessarily involve any interactions of A itself and can only happen through physical contact of gA with the protein synthesis machinery. Thus, only gA can be a host of the regulation of synthesis of A, the basic synthesis rate of A, k_{synth}^{A}, is a slave and different interaction partners of gA can be masters here.

The k_{synth}^{A} is modified depending on the interaction status of gA. Different combinations of gA with regulators (e.g. transcription factors) result in many possible sub-species gA_j. The fractional concentrations $[gA_j]/[gA]$ act as percentage weights used to calculate the overall synthesis rate of A:

$$k_{synth}^{A*} = \Sigma_j([gA_j]/[gA]k_{synth}^{A}\Pi_i\alpha_{ij}) \tag{2.20}$$

An example of multiple regulated synthesis with following weights contributing to the modified k_{synth}^{A*} is presented in the Fig. 2.15.

The outlined approach allows a detailed though uniform description of various protein synthesis regulation motifs in terms of EMA. However, the application of a gene concentration term $[gA]$ is questionable in terms of mass action law, since in the reality there are too few copies of the gene in the cell for the mass action approximation to be valid. Thus, the $[gA]$ term should be seen as a parameterization of several factors, such as promoter activity and mRNA concentration. An alternative solution could be either to use stochastic modeling algorithms (Sec. 1.2.2) or a phenomenological description as presented in Sec. 2.5.



Species	Regulation type	Regulation coefficient
TF_1	exclusion	α_1
TF_2	activation	α_2
TF_3	necessity	α_3

Sub^{gA}	σ_1^{gA}	σ_2^{gA}	σ_3^{gA}
gA_{000}	0	0	0
gA_{001}	0	0	1
gA_{011}	0	1	1
gA_{100}	1	0	0
gA_{101}	1	0	1

$k_{synth}^A = \Sigma_i\, w_i$	
w_1	gA_{000}
w_2	gA_{001}
w_3	$\alpha_2\ gA_{011}$
w_4	gA_{100}
w_5	gA_{101}

Figure 2.15: Regulation of protein synthesis. The k_n denote rate constants, α_n regulation coefficients, σ_n^X interfaces. Empty dots are introduced for visual convenience (see text for details). **A** - graphical representation of the system, **B** - possible sub-species of the regulated gene gA and their contribution to the synthesis rate of protein A.

2.4.3 Regulation of degradation

The regulation of degradation can also be described with the host-slave-master approach introduced in Sec. 2.4, based on following assumptions:

- Degradation of any molecule of species A can only happen by spontaneous disintegration or by physical contact with the degradation machinery molecules (ubiquitin, proteasomes etc.). Thus only A itself can be the host of regulation and its basic degradation rate, k^A_{degr} a slave of regulation.
- Regulation of degradation can happen only by physical contact between host and its partners. Thus, any interaction partner of host can be a master of regulation.
- If a regulation of degradation in a given sub-species A_j occurs, its degradation rate is modified depending on the values of all interfaces related to master partners in this sub-species.
- The regulation of degradation can have any of four kinds: activation, inhibition, necessity and exclusion, where the latest means, that a related master prevents A from degradation.

The basic degradation rate k_{degr} of species A is modified for a given sub-species A_j depending on its interaction status:

$$k^{A_j}_{degr} = k_{degr}\Pi_i\alpha_{ij} \tag{2.21}$$

An example of multiple regulated degradation with following individual k_{degr} for each A_j is presented in the Fig. 2.16.

The introduction of degradation rates for host results also in composed unbinding rates for its interaction partners, as shown in Fig. 2.16. Handling of such composed rates in the case of degradation regulation is identical to the enzymatic regulation case presented in the Sec. 2.4.1. Following from there, the regulation coefficients β_1 and β_2 are calculated using the basic regulation coefficients α_1 and α_2 and the species correspondence relationships: $[B1_2] = [A_4] + [A_5] + [A_6]$ and $[B2_2] = [A_3] + [A_6]$. From feeding those terms into the weighted average formula 2.19, we obtain:

$$\beta_1 = \frac{\alpha_1[A_4] + \alpha_1[A_5] + \alpha_1\alpha_2[A_6]}{[A_4] + [A_5] + [A_6]}$$

$$\beta_2 = \frac{\alpha_2[A_3] + \alpha_1\alpha_2[A_6]}{[A_3] + [A_6]}$$

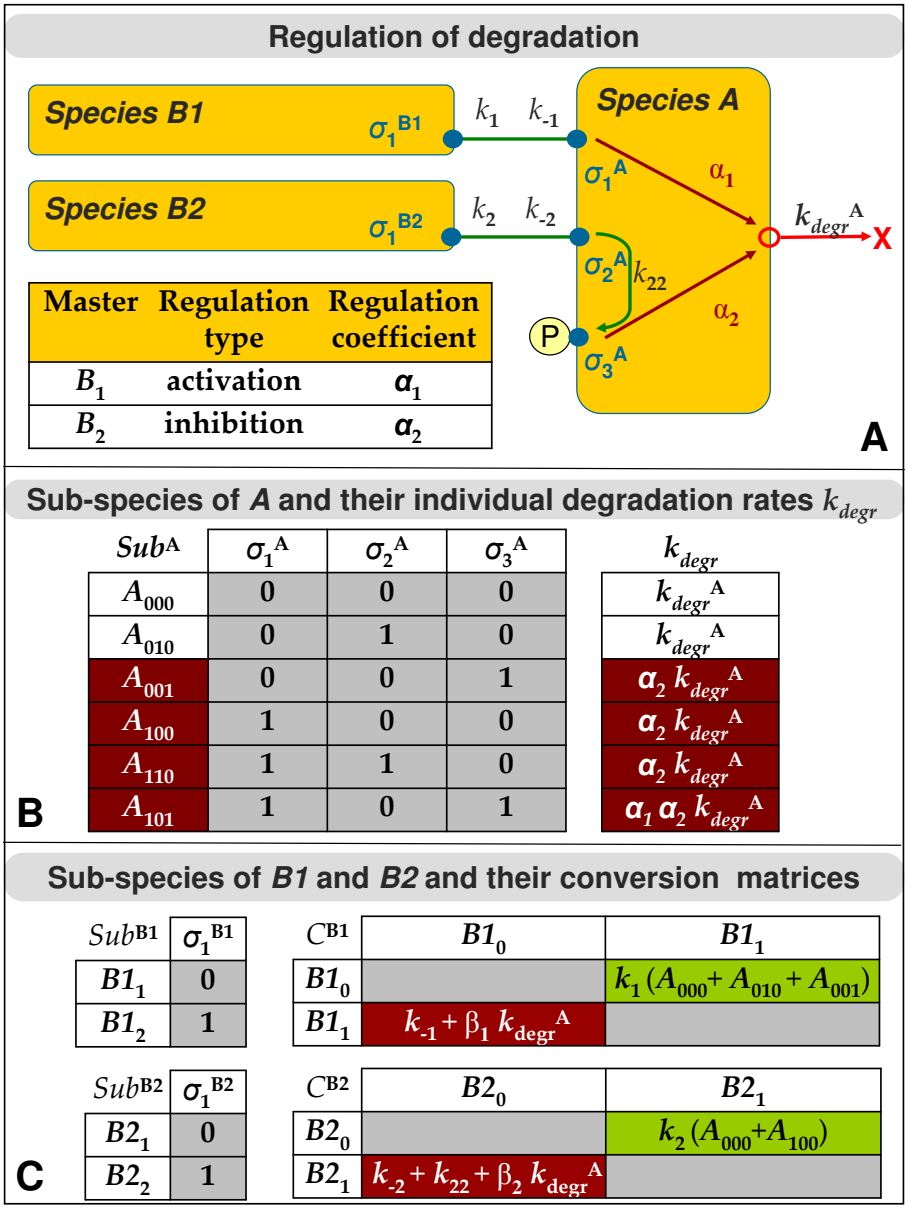


Figure 2.16: Regulation of degradation. The k_n denote rate constants, α_n regulation coefficients, σ_n^X interfaces. **A** - graphical representation of the system, **B** -possible sub-species of the regulated species A and their individual degradation rates **C** - possible sub-species of interaction partners $B1$ and $B2$ and their conversion coefficient matrices containing composed rates resulting from the degradation of A.

The outlined approach allows a detailed though uniform description of various degradation regulation motifs. An alternative approach is given on the phenomenological level as presented in Sec. 2.5.

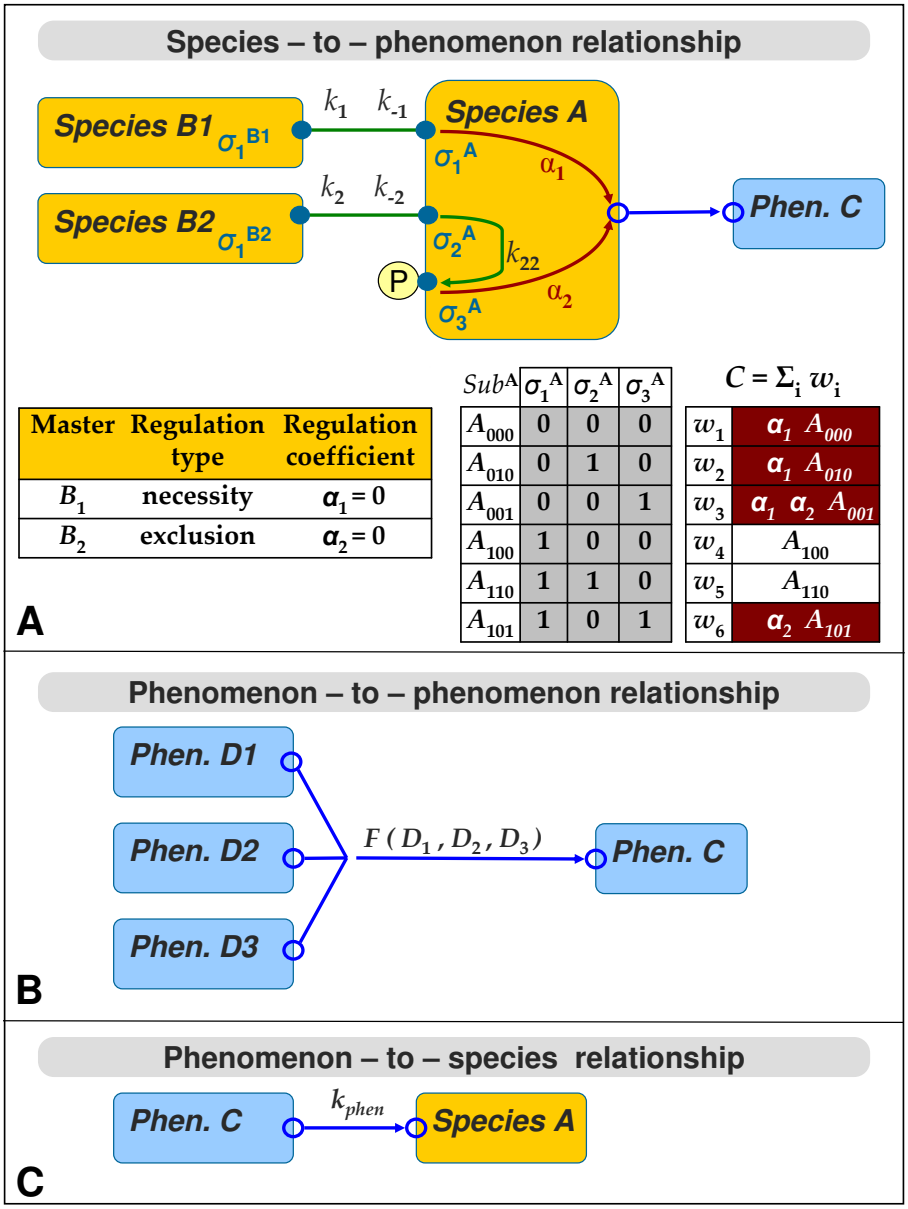
2.5 Representation of phenomena

Phenomena are an abstract category of species representing meaningful biological functions such as 'cell mass' or 'phosphorylated A'. Their specific character results from following assumptions:

- Any phenomenon species is an abstract term describing some aspect of the system's behavior. Thus, they can not be involved in physical interactions and consequently posses interaction interfaces.
- Prohibition of interaction interfaces to phenomena means that they can not have any sub-species. Consequently, each phenomenon species corresponds to only one variable of the system, which can be either real-valued (e.g. 'cell mass' = 1.0) or Boolean (e.g. 'cell division' = true).
- Phenomena can be linked to each other by means of algebraic-Boolean functions (Sec. 2.5.2) and to molecular species (especially proteins) by simple algebraic rules (Sec. 2.5.1 and Sec. 2.5.3). These couplings determine the value of phenomena-related variables.

The introduction of phenomena species allows to enrich the modeling framework with following features:

- Incorporation of meaningful **biological functions** such as 'cell mass' or 'cell division' into the model.
- Selection of a meaningful **subset of sub-species** of a given species (Sec. 2.5.1), e.g. 'active A'. This allows to reduce the simulation output to an user-defined shortlist of variables.
- **Phenomenological description of synthesis and degradation** (Sec. 2.5.3).
- **Phenomenological description of different processes** if the exact molecular mechanisms or parameters are unknown or beyond the scope of a given model.



Master	Regulation type	Regulation coefficient
B_1	necessity	$\alpha_1 = 0$
B_2	exclusion	$\alpha_2 = 0$

Sub^A	σ_1^A	σ_2^A	σ_3^A
A_{000}	0	0	0
A_{010}	0	1	0
A_{001}	0	0	1
A_{100}	1	0	0
A_{110}	1	1	0
A_{101}	1	0	1

$C = \Sigma_i\, w_i$	
w_1	$\alpha_1\ A_{000}$
w_2	$\alpha_1\ A_{010}$
w_3	$\alpha_1\ \alpha_2\ A_{001}$
w_4	A_{100}
w_5	A_{110}
w_6	$\alpha_2\ A_{101}$

Figure 2.17: Graphical representation of phenomena. Orange rectangles represent species, blue rectangles represent phenomena, green lines represent interactions, k_n denote rate constants, purple arrows represent regulation, α_n denote regulation coefficients, blue arrows represent links with phenomena, filled dots represent interfaces, σ_n^X denote interfaces, empty circles are introduced only for visual convenience.

There are three ways of connecting phenomena to other species as shown in Fig. 2.17 and discussed below. similarly to the graphical representation of synthesis and degradation (Sec. 2.2), we use 'empty' nodes that allow a convenient visual linking of phenomena to each other and to other species. Again, these nodes do not depict any interaction interfaces.

2.5.1 Species-to-phenomenon relationship.

This relation allows reducing the simulation output to an user-defined short-list of meaningful variables like: 'active A' = 'A bound to B but not to C'. The user can determine, which interactions of species A should or should not contribute to a given phenomenon P using the same input format as for regulation. In the given example: host = 'A', slave = 'active A' , master1 = 'B', type1 = 'necessity', master2 = 'C', type2 = 'exclusion'.

Based on the values of master interfaces σ_i in every sub-species A_j, it can be decided if this A_j can contribute its concentration to the value of P or not. The only regulatory relationship allowed to be used here is 'necessity' or 'exclusion'. The value of P is simply calculated as a sum of all the concentrations $[A_j]$ that are allowed to contribute to P by the regulation coefficient α_{ij}:

$$P \quad = \quad \Sigma_j([A_j]\Pi_i\alpha_{ij}) \tag{2.22}$$

An example of a species A linked to a phenomenon P by means of regulation along with the sub-species A_j allowed and not-allowed to contribute to P is presented in Fig. 2.17 A.

2.5.2 Phenomenon-to-phenomenon relationship.

The relationships between phenomena, both real-valued and Boolean ones, can be described with user-defined, Boolean and arithmetic rules, like: 'IF concentration of 'phosphorylated A' exceeds threshold T_A, phenomenon 'exocytosis' is true'.

The interconnection of phenomena allows capturing various relationships, such as logical gates, threshold behavior etc. As shown in Fig. 2.17 B, a phenomenon P can be linked to any set of other phenomena $D_1 \ldots D_n$ by means of arbitrary Boolean-algebraic coupling terms F with any set of assistant Boolean or numerical arguments $X_1 \ldots . X_k$:

$$P \quad = \quad F(D_1 \ldots D_n, \quad X_1 \ldots . X_k) \tag{2.23}$$

Where $F()$ can for instance mean: 'IF $D_1 > threshold\ X_2$, THEN $P = P/D_4$. The structure of expressions is formalized in Sec. 3.2.4.

Such phenomenological treatment can be an useful alternative for handling unknown mechanisms or rate constants in some specific cases. However, this feature shifts the focus away from molecular mechanisms of interaction towards a phenomenological system description.

2.5.3 Phenomenon-to-species relationship.

As shown in Fig. 2.17 C, a real number phenomenon P can be linked to species A via a phenomenological rate constant k_{phen} which is treated as an additional degradation ($k_{phen} < 0$) or synthesis ($k_{phen} > 0$) rate of A that decreases/increases the concentration of A:

$$\begin{aligned} d[A]/dt &= k_{phen}P \qquad \text{if } k_{phen} > 0 \\ d[A]/dt &= k_{phen}P[A] \quad \text{if } k_{phen} < 0 \end{aligned} \tag{2.24}$$

This relation allows the phenomenological description of synthesis and degradation and their regulation. For instance, if we know that some forms of species B influence the synthesis of species A, e.g. 'phosphorylated B promotes synthesis of A' we can create a phenomenon P for the desired subset of B (Sec. 2.5.1) and link this phenomenon back to A by means of k_{phen}. For instance, if the phenomenon 'phosphorylated B' corresponds to 2 sub-species B_a and B_b, then $d[A]/dt = +k_{phen}([B_a] + [B_b])$.

Chapter 3

Automated model construction and simulation

We have implemented the formalism presented in Chpt. 2 with $Java^{TM}$ as described below. The resulting software, called **aceSim**, allows an automatic MIN model construction, simulation and analysis based on a limited input set of interaction rules. The acronym 'ace' refers to '**a**utomated, **c**ombinatorial, **e**lementary (mass action)', which we believe to be the key characteristics of the presented simulator.

In the aceSim framework, a MIN is described with a set of simple, user-defined rules containing names of interacting species, reaction rate constants and optional regulation coefficients. These parameters are extrapolated into ODE modules, which are automatically combined into an ODE system describing all combinatorial reaction pathways possible in the modeled network. Thus, no further parameters need to be entered nor manual modification of the existing parameters is required to obtain the system description.

The ODE modules have an EMA form, which ensures compliance with mass conservation laws even for large and complicated systems and thus greater mathematical precision compared to AMA-based descriptions. The purely automated parameter extrapolation and module combination is facilitated by an agent-like, Boolean representation of combinatorial molecular species.

3.1 General algorithm

The general algorithm of the automated model construction and simulation can be summarized to following steps:

1. Download the job file containing a tabular description of the system (Sec. 3.2).
2. Create a repository of all agents, i.e. interacting species in the system (Sec. 3.3).
3. Create several tables capturing relations between agents (Sec. 3.4):
   - Create an all-to-all Agent Connection Matrix .
   - Determine number of interfaces per agent based on the number of its interaction partners
   - Create Sub-Species Library, i.e. all possible combinations of interface states of each agent.
   - Create Conversion Matrix containing all possible conversions between sub-species resulting from interactions.
   - Create Phenomena Matrix capturing species-to-phenomenon and phenomenon-to-species relationships.
4. Define dynamical variables (Sec. 3.5.1).
5. Derive from the above tables ODE describing the system (Sec. 3.5.2).
6. Integrate equations (Sec. 3.5).
7. Analyze results with built-in analysis features (Sec. 3.5).

We will illustrate our implementation of this algorithm with an example of a kinase-phosphatase system (see Section 2.3.2) as outlined in Figures 3.1 and 3.3.

3.2 Structure of the job file

The job file has a worksheet structure divided into several tables as outlined in Fig. 3.2. The most important one is the interaction table *jobI*. It lists basic binding and enzymatic reactions, from which all combinatorial reaction pathways possible in the system will be automatically generated.

An example of such table for a kinase-phosphatase system is depicted in Figure 3.1 A. It contains 2 rows, corresponding to one phosphorylation reaction of substrate B by kinase $A1$ and one dephosphorylation reaction of B by phosphatase $A2$, respectively.

As depicted in Fig. 3.2 A, a *jobI* record contains names of species participating in a given interaction, interaction type and corresponding rate constants (see Sec. 3.2.1 for details).

The job file can also contain 3 optional tables listing: *a*) regulatory relationships (Sec. 3.2.2), *b*) initial values and basic synthesis/degradation rates for selected species (Sec. 3.2.3), *c*) phenomenon-to-phenomenon relations (Sec. 3.2.4). The compact form of the job file allows rule-based description of both interactions and their regulation and thus far-reaching automation of the modeling process. Below we characterize the fields of the job file tables in a more detail. An overview of these fields is presented in Fig. 3.2. All entries referred to as 'number' were implemented using the double precision numbering format (8 byte encoding) [8].

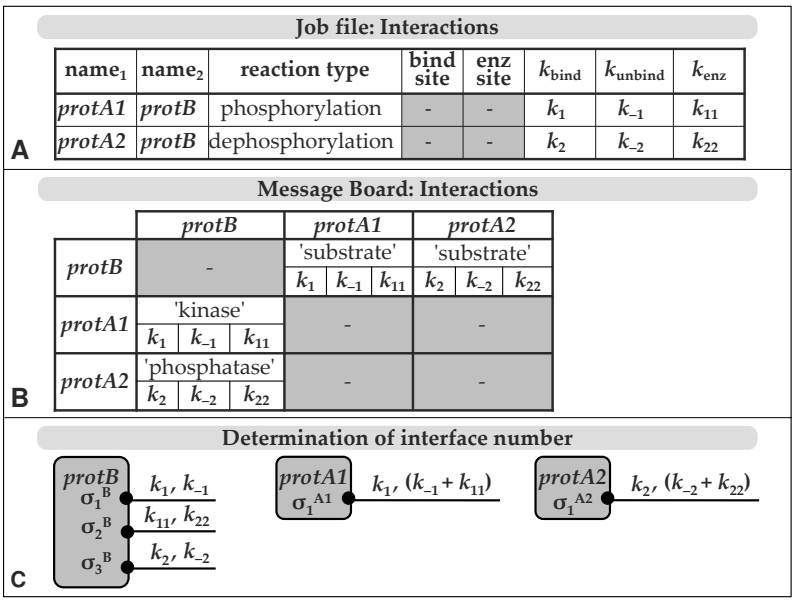


Figure 3.1: Implementation example based on a kinase-phosphatase motif. See text for details.

Structure of the job file

table:	Interaction Table *(jobI)*							
fields:	**name$_1$**	**name$_2$**	**type**	**bind site**	**enz site**	k_1	k_{-1}	k_{11}
type:	string	string	string	string	string	number	number	number
possible entries:	prot gene phen	prot gene	'B' 'E1' 'E2' 'V'	'@...'	'@...'	binding rate $k_{phen} > 0$	unbinding rate $k_{phen} < 0$	transform. rate

A

table:	Regulation Table *(jobR)*						
fields:	**host**	**master**	**slave**	**type**	**master site**	**slave site**	α
type:	string	string	string	string	string	string	number
possible entries:	prot gene	prot gene	prot gene phen 'synthesis' 'degradation'	'A' 'I' 'E' 'N'	'@...'	'@...'	regulation coefficient

B

table:	Initial Values Table *(jobV)*				
fields:	**name$_1$**	**ini**	k_{synth}	k_{degr}	**iniBoolean**
type:	string	number	number	number	Boolean
possible entries:	prot phen pheb	initial value	synthesis rate	degradation rate	initial value

C

table:	Phenomena Table *(jobP)*					
fields:	**cond. arg.**	**cond oper.**	**cond. param.**	**output arg.**	**output oper.**	**output param.**
type:	string	char.	string	string / Boolean	char.	string / number
possible entries:	phen pheb	'=' '<' '>'	phen threshold value	phen pheb	'+' '-' '*' '/'	phen modifier value

D

Legend:

- obligatory field
- optional field
- pre-defined entry format

Figure 3.2: Structure of the job file in a spreadsheet format. 'prot' - any protein species, 'gene' - any gene species, 'phen' - any quantitative phenomena species. See text Sec. 3.2 for further details.

3.2.1 Job file - Interactions

Structure of the Interaction Table (*jobI*) is based on the formal description of interactions presented in Sec. 2.2 and Sec. 2.3. The *jobI* contains following fields:

- **name$_1$** and **name$_2$** (string) - the names of interacting partners. All names have standardized first 4 characters defining the type of species, e.g.: *prot* for proteins, *gene* - for genes, *phen* - for real-valued phenomena, *pheb* - for Boolean phenomena. For instance, a cyclin D gene and protein could be named 'geneCyclinD' and 'protCyclinD', respectively. For enzymatic reactions, the enzyme is by convention entered as $name_1$ and the substrate as $name_2$.

- **type** (string) - the type of interaction can be specified as follows: **B** - binding reaction, **E1** - enzymatic modification reaction, esp. phosphorylation, in which case the $name_1$ is a kinase, **E2** - enzymatic demodification reaction, esp. dephosphorylation, in which case the $name_1$ is a phosphatase and **V** - abstract relation for linking phenomena to species, in which case $name_1$ is a phenomenon.

- **bind site** and **enz site** (string, optional) - explicit labels for binding and modification sites, respectively. By convention, the label starts with the character '@' (e.g.'@0' or '@Tyr115').

- **k$_n$**, **k$_{-n}$**, **k$_{nn}$** (number) - rate constants for, respectively: binding of $name_1$ to $name_2$, unbinding of $name_1$ from $name_2$ and modification of $name_2$ by $name_1$. In a more detail, the k_{nn} is a forward rate of an enzymatic transformation from enzyme-substrate complex to enzyme and product and it exists only if $int\,type = E1$ or $int\,type = E2$.

The *jobI* can list phenomenon-to-species relations with corresponding k_{phen} as described in Sec. 2.5.3. By convention, the phenomenon is entered as $name_1$, $k_{phen} \geq 0$ as k_n, $k_{phen} < 0$ as k_{-n} and *type* is set to 'V'.

The *jobI* can also list interactions where $name_2$ has $n > 1$ binding sites to $name_1$ or can be enzymatically modified at $n > 1$ sites, as described in Sec. 2.3.1 and Sec. 2.3.3, respectively (by convention, the species having such multiple sites is entered as $name_2$). In this case, each binding/enzymatic reaction related to a specific site is entered as a separate row of the *jobI* table; these rows contain the same pair of species names but differ in the values of *bind site* or *bind enz* fields.

3.2.2 Job file - Regulation

Structure of the optional Regulation Table (*jobR*) is based on the formal description of regulation presented in Sec. 2.4. The *jobR* contains following fields:

- **host**, **master** and **slave** (string) - the names of species involved in a regulatory relationship as described in Sec. 2.4. For a regulation between a given host, master and slave to be valid, There must be a host-master and a host-slave interaction listed in the (*jobI*) table.
- **reg type** (string) - the type of regulation can be specified as follows: **A** - activation, **I** - inhibition, **E** - exclusion, **N** - necessity.
- **master site** and **slave site** (string, optional) - site labels that were assigned in the *jobI* to host-master or host-slave interaction, respectively. By convention, for enzymatic reactions, the *master site* can only be an *enz site* and the *slave site* can only be a *bind site*. This is because we assume, that *master* enzymes can only execute regulation of their substrate after its enzymatic modification has taken place and that regulation of *slave* enzymes influences their activity already at the stage of substrate binding.
- α (number, optional) - the value of regulation coefficient α.

The *jobR* can list species-to-phenomenon relations as described in Sec. 2.5.3. The phenomenon is by convention entered as slave. This only refers to real-valued phenomena and the regulation type can only be 'N' or 'E'.

The *jobR* can also list host's synthesis and degradation regulation, as described in Sec. 2.4.2 and Sec. 2.4.3, respectively. In this case, the slave is entered as a pre-defined tag 'synthesis' or 'degradation'.

3.2.3 Job file - Initial Values

The optional Initial Values Table (*jobV*) contains following fields:

- **name$_1$** (string, optional) - the name of the species for which the subsequent fields of *jobV* are specified.
- **ini** (number, optional) - initial concentration of species $name_1$.
- **iniBoolean** (Boolean, optional) - initial value for phenomena with Boolean value.

- $\mathbf{k_{synth}}$, $\mathbf{k_{degr}}$ (number, optional) - respectively, the value of basic synthesis or degradation rate constant for $name_1$.

By convention, for all species appearing in $jobI$ but not in $jobV$, the default initial concentration, synthesis and degradation rates are set to 0, except for gene species where the concentration is fixed at 1 thorough the whole simulation. For Boolean phenomena, the default initial value is $false$. Note, that for real-valued phenomena we also allow synthesis and degradation rates, in order to simulate some dynamical phenomena, such as cell growth.

3.2.4 Job file - Phenomena

The optional phenomena table ($jobP$) lists phenomenon-to-phenomenon relations in form of user-defined, logic-algebraic expressions as described in (Sec. 2.5.2). Every record of the $jobP$ is a single rule in the form of Eq. 2.23. It can contain names of real-valued and Boolean phenomena from in the $jobI$ and $jobV$ tables, arithmetical operators and user-defined numbers. These components are combined using following formalized syntax:

IF(*condition argument*, *condition operator*, *condition parameter*)
THEN(*output argument*, *output operator*, *output parameter*)

For example, cell division can be simulated by setting a rule by which a phenomenon 'cell size' is halved whenever it reaches the threshold value T: IF(phenCellSize,=,T) THEN(phenCellSize,/,2).

The above syntax elements correspond to the following fields of $jobP$:

- **condition argument** (string) - name of any phenomenon.
- **condition operator** (character) - '=', '>', or '<'.
- **condition parameter** (string or number) - name of any real-valued phenomenon or an user-defined number to which the *condition argument* will be compared using *condition operator*.

The IF() part of the rule can always be evaluated in Boolean terms. If the result is *true*, the THEN() part will be executed. If *output argument* is a Boolean phenomenon, it will be simply set to *true* and that the fields *output operator* and *output parameter* do not need to be entered.

- **output argument** (string) - name of any phenomenon that will be modified by the rule.
- **output operator** (character, optional) - '+', '-', '*' or '/'.
- **output parameter** (string or number, optional) - name of any real-valued phenomenon or an user-defined number used together with *output operator* to modify the value of *output argument* if the IF() part of the rule returns *true*.

The above syntax allows representing some basic biological relationships like thresholds or logical gates. It can be further developed to increase the descriptive scope of our formalism on the phenomenological level.

3.3 Definition of agents

In the first step, the job file is parsed to automatically create the **global list** $aALL$, which is a repository of all agents, i.e. the interaction partners in the system. The $aALL$ contains following fields:

- **name** (string) - name of the agent. Field copied from the $jobI$ table.
- **type** (string) - first four characters of the agent's name. Can have values: 'prot' (proteins), 'gene' (genes), 'phen' (real-valued phenomena), 'pheb' (Boolean phenomena).
- **synth** (number, optional) - basic synthesis rate. Field copied from the $jobV$ table.
- **degr** (number, optional)- basic degradation rate. Field copied from the $jobV$ table.
- **ini** (number) - total initial value. Field copied from the $jobV$ table.

To simplify further data processing, $aALL$ is filtered according to the $aALL^{type}$ into following sub-lists:

- **global list of proteins and genes** (a) - contains *prot* and *gene* entries only. This list contains an additional field: **ref** (integer, optional) - reference of indexes on the a between each *prot* species and a corresponding *gene* species if such exists.
- **global list of real-valued phenomena** - (aP) - *phen* entries only.

- **global list of Boolean phenomena** - (aB) - *pheb* entries only.

The lists a, aP and aB are generated with following algorithm

1. Collect all distinctive names from the $jobI$ table fields $name_1$ and $name_2$ into a shortlist $aALL$.
2. For each entry $aALL_i$, set the $aALL_i^{type}$ to first 4 characters of $aALL_i^{name}$.
3. For each $aALL_i$, parse the $jobV$ and find an entry j such that $aALL_i^{name} = jobV_j^{name}$ and set following fields to values copied from $jobV$ if available or default values as indicated in the brackets: $aALL_i^{synth} = jobV_j^{k_{synth}}$ (0.0), $aALL_i^{degr} = jobV_j^{k_{degr}}$ (0.0), $aALL_i^{ini} = jobV_j^{ini}$ (1.0 for *prot* and *gene* species, 0.0 for *phen* species and $false$ for *pheb* species).
4. Filter the $aALL$ into following sub-lists based on the value of the *type* field as indicated in the brackets: a ($type = prot$ or $type = gene$), aP ($type = phen$), aB ($type = pheb$).
5. Parse the a for pairs of entries A_i and A_j where: $a_i^{name} = a_j^{name}$ (apart from the first 4 characters), $a_i^{type} =' gene'$, $a_j^{type} =' prot'$. If such pair A_i, A_j is found, set $a_j^{ref} = i$.

3.4 Communication between agents via message board

After a shortlist of all molecular species in the MIN has been created, relations between these agents are captured in several multidimensional matrices, from which equations describing the system's behavior can be derived (Sec. 3.5.2). These matrices are composed of following dimensions:

- **n** - number of *prot* and *gene* agents in the system (i.e. number of entries in the a list)
- **k** - number of distinctive sites of interaction between a given pair of agents A_i and A_j.
- **s** - number of sub-species of a given agent A_i. The dimension s is a function of n and k as described below.
- **p** - number of *phen* agents in the system (i.e. number of entries in the aP list)

These dimensions are combined in following five message board tables:

- **Message Board Interactions** (I) - $\{n, n, k\}$ (Sec. 3.4.1).
- **Message Board sub-species** (S) - $\{n, s\}$ (Sec. 3.4.2).
- **Message Board Conversions** (C) - $\{n, s, s\}$ (Sec. 3.4.3).
- **Message Board Relations** (R) - $\{n, n, k, s\}$ (Sec. 3.4.4).
- **Message Board Phenomena** (P) - $\{n, s, p\}$ (Sec. 3.4.5).

The automatic construction of message board tables is described below in detail.

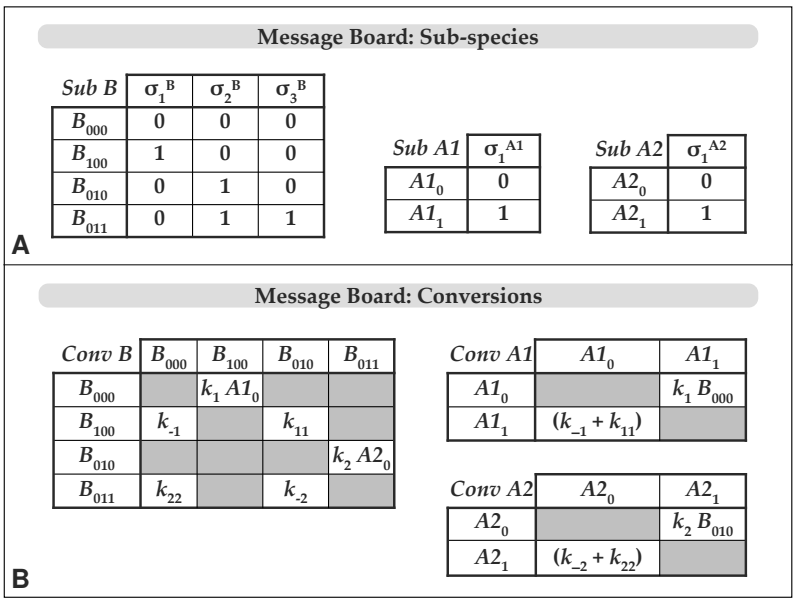


Figure 3.3: Implementation example based on a kinase-phosphatase motif (continuation). See text for details.

3.4.1 Message board - Interactions

The I table represents all interactions between species in the system, as exemplified in Fig. 3.1 B. An empty cell I_{ij} means, that there is no interaction between the agents A_i and A_j. Each non-empty cell I_{ij} corresponds to one

interaction between A_i and A_j and can contain k records that correspond to the sites of this interaction.

Each column of I corresponds to an individual interaction menu of a given agent A. Then, one interface of A is created per each non-empty entry in the column plus one common interface for all enzymatic reactions. Reactions with differently labeled binding or enzymatic sites receive separate interfaces. An example of interface derivation from the I table is depicted in Fig. 3.1 C.

Each record I_{ijk} contains following fields:

- **type** (string, optional) - type of interaction between A_i and A_j; field copied from $jobI$.
- **status** (string) - describes the status of A_i in respect to A_j. Depending on the field *type* and the number k of interaction sites between A_i and A_j, the field I_{ijk} *status* can have values: 'complex', 'multiple ligand', 'multiple receptor', 'enzyme', 'multiple enzyme', 'substrate', 'multiple substrate'.
- **bind** (string, optional) - label of the k-th binding interface of A_i to A_j; field copied from $jobI$.
- **bind index** (string) - a running number for all binding interaction interfaces of a given agent A_j. Exists also if the corresponding *bind* field is empty. Note, that for multiple partners binding to the same physical site, we assign individual interfaces and thus different indexes for each partner.
- **enz** (string, optional) - label of the k-th interface of enzymatic modification of A_j by A_i; field copied from $jobI$.
- **enz index** - a running number for all enzymatic modification sites of a given agent A_j. Exists also if the corresponding *enz* field is empty. For enzymatic sites targeted by $n > 1$ enzymes and thus appearing in several interaction entries of A_i, a single interface and thus an identical index number is created.
- $\mathbf{k}_1$, $\mathbf{k}_{-1}$, $\mathbf{k}_{11}$ (number) - rate constants of the reaction between A_i and A_j; fields copied from $jobI$.

The I table is created using following algorithm:

1. For each agent A_i listed in a search for interaction partners A_j in the table $jobI$.

2. If for the given A_i an interaction entry with partner A_j at site k is found, copy the corresponding fields of $jobI$ into I_{ijk}.

3. Set I_{ijk}^{status} depending on the value of I_{ijk}^{type} and number of sites in the I_{ij} entry (i.e. the value of dimension k).

4. For every column I_j, Set the subsequent entries $I_{ijk}^{bindindex}$ to a running number b starting from 0 end ending at b_{fin}.

5. Simultaneously, collect all distinctive names from I_{ijk}^{enz} and save them to a shortlist 'enzymatic sites'. For sites without name use one and the same string label e.g. 'noname'.

6. For every entry i in the column I_j, where I_{ijk}^{status} = 'enzyme' or 'mult enzyme', set $I_{ijk}^{enzindex} = b_{fin} + 1 + e$, where e is the index of I_{ijk}^{enz} on the 'enzymatic sites' shortlist.

3.4.2 Message board - sub-species

The S table represents all possible sub-species of each agent in the system. Thus, for a given species A, the S lists all possible combinations of its interface values in form of Boolean vectors, as described in Sec. 2.1 and exemplified in Fig. 3.3 A. This list is filtered by removing vectors corresponding to sub-species not allowed in enzymatic reactions (e.g. enzyme-product complexes) or due to regulation by necessity/exclusion (e.g. complexes of A with a slave and simultaneously with a master excluding this slave).

The S has a matrix form with 2 dimensions $\{n, s\}$. Each S_{ix} record corresponds to one sub-species A_{ix} of the agent A_i and contains following fields:

- **sub** (Boolean vector) - a possible combination of σ interface states of a given agent A_i; every such combination corresponds to one sub-species A_{ix}. The length of this vector equals the total number of interfaces that A_i has.

- **tot** (number) - initial total concentration of the A_{ix}.

- **perc** (number) - fraction of A_{ix} to A_i.

- **prot** (Boolean) - determines if a given A_{ix} corresponds to the initial state of a protein and thus can be influenced by a protein synthesis rate. The $S_{ix}\,prot = true$ only if a_i^{type} = 'prot' and $S_{ixp}^{sub} = false\,\forall\,p$

- **degr** - (number, optional) the degradation rate of A_{ix} depending on the a_i^{degr} and eventual regulation coefficients α listed in $jobR$. If a_i^{type} = 'gene', $S_{ix}^{degr} = 0.0$ by default.

- **synth** - (number, optional) the synthesis rate of A_{ix} depending on the a_i^{synth} and eventual regulation coefficients α listed in $jobR$. If a_i^{type} = 'prot', $S_{ix}^{synth} \neq 0$ only if $S_{ix}^{prot} = true$. If a_i^{type} = 'gene', a specific convention is used, by which S_{ix}^{synth} refers to the synthesis rate of a corresponding 'prot' species, as listed in the a_i^{ref}. Thus, a S_{ix}^{synth} of a 'gene' sub-species A_{ix} is a multiplication of all applicable regulation coefficients listed in $jobR$.

The S table is created using following algorithm:

1. For each agent A_i, determine the number of interfaces r which is the highest number in the column S_i in the fields $bind\,index$ or $enz\,index$.

2. Create $x = 2^r$ Boolean vectors of length r representing all possible combinations the r interface states of A_i, i.e. the sub-species A_{ix}.

3. Filter out all A_{ix} not allowed by rules for enzymatic interactions or due to regulation by necessity/exclusion listed in $jobR$.

4. Save the Boolean representation of each remaining A_{ix} to S_{ix}^{sub}.

5. Set the S_{ix}^{tot}, S_{ix}^{perc} and S_{ix}^{prot} as follows:

   - If any entry in the vector S_{ix}^{sub} is $true$, set the above fields to: 0.0, 0.0 and $false$, respectively.
   - If all entries in the vector S_{ix}^{sub} are $false$ and a_i^{type} = 'gene', set the above fields to: 1.0, 1.0 and $false$, respectively.
   - If all entries in the vector S_{ix}^{sub} are $false$ and a_i^{type} = 'prot', set the above fields to: a_i^{ini}, 1.0 and $false$, respectively.

3.4.3 Message board - Conversions

The conversion matrix C describes all possible conversions between sub-species A_{ix} and A_{iy} of each agent in the system using a matrix form with 3

dimensions: $\{n, s, s\}$. An example of such matrix for the kinase-phosphatase system is presented in Fig. 3.3 B.

An entry $C_{ixy} \neq 0$ means, that a conversion of sub-species A_{ix} into A_{iy} is possible. We assume that such conversion can only occur in the course of a single interaction. Consequently, A_{ix} and A_{iy} can only differ by opposite values of one single interface (e.g. partner unbound → bound). An exception is the second step of enzymatic reaction, where the linked A_{ix} and A_{iy} must differ by opposite values of 2 interfaces (enzyme bound → unbound and enzymatic site unmodified → modified).

The choice of an ODE module for describing conversion of A_{ix} into A_{iy} depends on the interaction type as presented in Chpt. 2.

An empty record C_{ixy} means that such conversion is not possible because: the sub-species A_{ix} and A_{iy} differ by more than 1 interface value or by 2 interface values not associated with the same enzymatic reaction or because participation in interaction is not allowed for these sub-species due to regulation.

Each C_{ixy} record contains following fields:

- **on** (number) - forward transition rate from A_{ix} to A_{iy} where the value of the responsible interface σ changes from 0 to 1. The value of *on* is copied from $jobI^{k_n}$ or $jobI^{k_{nn}}$.

- **off** (number) - backward transition rate from A_{ix} to A_{iy} where the value of the responsible interface σ changes from 1 to 0. The value of *off* is copied from $jobI^{k_{-n}}$.

- **reg** (number) - regulation coefficient associated with the *on* field. The value of reg is calculated as a product of all regulation coefficients α from $jobR$ that are applicable to the transition from A_{ix} to A_{iy}.

- **bind** (integer vector) - indexes i and k of the binding partner A_j listed in I_{ijk}. Exists if I_{ijk}^{type} = 'B'.

- **enz** (integer vector) - indexes i and k of the enzymatic partner A_j listed in I_{ijk}. Exists if I_{ijk}^{type} = 'E1' or 'E2'.

- **degr** (integer vector) - indexes i and k of the partner A_j listed in I_{ijk}. Exists if $a_j^{degr} \neq 0$.

The C table is created using following algorithm:

1. For each agent A_i, compare pairwise its all sub-species ($\{A_{ix}, A_{iy}\}$) represented as Boolean vectors in S_i^{sub} and check for following condition sets:

2. Condition set 1:

   - Vectors S_{ix}^{sub} and S_{iy}^{sub} differ only by value of 1 entry at the index p.
   - $S_{ixp}^{sub} = false$ and $S_{iyp}^{sub} = true$ (this is potentially a forward binding reaction).
   - Record I_{ijk} exists, where $I_{ijk}^{bindindex} = p$.

3. If the conditions listed in the condition set 1 are simultaneously fulfilled, set following fields:

   - $C_{ixy}^{on} = I_{ijk}^{k_n}$.
   - $C_{ixy}^{bind} = \{j, k\}$.
   - $C_{iyx}^{off} = I_{ijk}^{k_{-n}}$.
   - $C_{iyx}^{degr} = \{j, k\}$.

4. Condition set 2:

   - Vectors S_{ix}^{sub} and S_{iy}^{sub} differ only by value of 2 entries p and q
   - $p < q$ (potentially, p refers to an enzyme binding and q to an enzymatic modification interface).
   - $S_{ixp}^{sub} = true$ and $S_{iyp}^{sub} = false$ (the enzyme is unbinding).
   - Record I_{ijk} exists, where: $I_{ijk}^{bindindex} = p$, $I_{ijk}^{enzindex} = q$.
   - For the same record, I_{ijk}^{type} = 'E1' and $S_{ixq}^{sub} = false$ and $S_{iyq}^{sub} = true$ (the enzymatic modification occurs) OR I_{ijk}^{type} = 'E2' and $S_{ixq}^{sub} = true$ and $S_{iyq}^{sub} = false$ (the enzymatic modification is removed).

5. If the conditions listed in the condition set 2 are simultaneously fulfilled, set set following fields:

   - $C_{ixy}^{on} = I_{ijk}^{k_{nn}}$.
   - $C_{ixy}^{enz} = \{j, k\}$.

6. If a transition from A_{ix} to A_{iy} was validated by compliance with condition set 1 or 2, parse the table $jobR$ to check if records $jobR_r$ and I_{imn} exist, so that following condition set is fulfilled:

7. Condition set 3:

   - $jobR_r^{host} = A_i$, $jobR_r^{slave} = A_j$, $jobR_r^{master} = A_m$.
   - $jobR_r^{slave\,site} = I_{ijk}^{bind}$ or $jobR_r^{slave\,site} = I_{ijk}^{enz}$.
   - $jobR_r^{master\,site} = I_{imn}^{bind}$ or $jobR_r^{master\,site} = I_{imn}^{enz}$.
   - $S_{ixp}^{sub} = true$ at index p, such that:
   - I_{imn}^{type} = 'B' and $I_{imn}^{bindindex} = p$ OR I_{imn}^{type} = 'E1'('E2') and $I_{imn}^{enzindex} = p$ (a regulator is bound or a regulating enzymatic site is modified).

8. If the conditions listed in the condition set 3 are simultaneously fulfilled, set $C_{ixy}^{reg} = jobR_r^{\alpha}$.

9. If $n > 1$ records $jobR_r$ fulfilling the condition set 3 are found, set $C_{ixy}^{reg} = \Pi_{r=1}^{n}\alpha_r$.

3.4.4 Message board - Relations

The R table connects the I and S tables by specifying which sub-species of the interacting agents are taking part in the interaction. Thus, the R table uses a matrix form with 4 dimensions; $\{n, n, k, s\}$. Each record R_{ijkx} corresponds to one sub-species A_{ix} that interacts with an agent A_j at the site k and is constructed from following fields:

- **bind** (Boolean) - specifies if sub-species A_{ix} can participate in a binding interaction with A_j at the site k.

- **bindreg** (number) - eventual regulation coefficients α for *bind*. The default value is 1 (no regulation).

- **enz** (Boolean) - specifies if sub-species A_{ix} can participate in a enzymatic transformation of A_i at the site k.

- **enzreg** (number) - eventual regulation coefficients α for enz. The default value is 1 (no regulation).

- **degr** (Boolean) - specifies if sub-species A_{ix} can contribute with its degradation rate k_{degr} to the unbinding from A_j at the site k.

- **degrreg** (number) - eventual regulation coefficients α for *degr*. The default value is 1 (no regulation).

The R table is created using following algorithm:

1. For each record I_{ijk}, parse the conversion matrix of species A_i to check if a record C_{ixy} exists, such that following condition sets are fulfilled:
2. Condition set 1:
   - $C_{ixy}^{on} \neq 0.0$
   - $C_{ixy}^{bind} = \{j, k\}$
3. If the conditions in the condition set 1 are simultaneously fulfilled, set:
   - $R_{ijkx}^{bind} = true$
   - $R_{ijkx}^{bindreg} = C_{ixy}^{reg}$
4. Condition set 2:
   - $C_{ixy}^{on} \neq 0.0$
   - $C_{ixy}^{enz} = \{j, k\}$
5. If the conditions in the condition set 2 are simultaneously fulfilled, set:
   - $R_{ijkx}^{enz} = true$
   - $R_{ijkx}^{enzreg} = C_{ixy}^{reg}$
6. Condition set 3:
   - $C_{ixy}^{off} \neq 0.0$
   - $C_{ixy}^{degr} = \{j, k\}$
7. If the conditions in the condition set 3 are simultaneously fulfilled, set:
   - $R_{ijkx}^{degr} = true$
   - $R_{ijkx}^{degrreg} = S_{ix}^{degr}$. Note, that in this case we derive the sub-species-specific regulation coefficient from the S table, not the C table.

3.4.5 Message board - Phenomena

The P table connects the quantitative phenomena agent list aP with the protein and gene sub-species table S by specifying which sub-species of a given agent A_i contribute to the value of a given phenomena P_q (Sec. 2.5.1) and vice versa, i.e. which phenomena contribute to the changes of concentration of which agents. Thus, the P table has a matrix form with 3 dimensions: $\{n, s, p\}$ and contains following fields:

- **phen2prot** (number) - an entry $P_{ixq}^{phen2prot} = k_{phen} \neq 0$ means, that the phenomenon P_q influences the concentration changes of the sub-species A_{ix} by the coefficient k_{phen}. This approach allows description of synthesis ($k_{phen} > 0$) and degradation ($k_{phen} < 0$) on the phenomenological level, as described in Eq. (2.24). Consistent with this, we allow negative k_{phen} for any A_{ix} but positive k_{phen} only for those A_{ix}, where $S_{ix}^{prot} = true$. Values of k_{phen} are copied from $jobI^{k_n}$ and $jobI^{k_{-n}}$.

- **prot2phen** (Boolean) - an entry $P_{ixq}^{prot2phen} = true$ means, that the concentration of sub-species A_{ix} needs to be included into the value calculation of the phenomenon P_q as described in Eq. (2.22).

The P table is created using following algorithm:

1. Parse the $jobI$ table. For all entries I_a, where $I_a^{type} =' V'$ determine indexes q and i, where:
   - q = index of I_a^{name1} from the phenomena list aP.
   - i = index of I_a^{name2} from the gene and protein list a.
2. Parse the sub-species in S_i^{sub}. Set $P_{ixq}^{phen2prot} = jobI_a^{k_n}$ only if $S_{ix}^{prot} = true$. Set $P_{ixq}^{phen2prot} = jobI_a^{k_{-n}} \forall A_{ix}$.
3. Parse the $jobR$ table. For all entries, where R_a^{slave} is a phenomenon P_q, determine indexes q and i, where:
   - q = index of R_a^{slave} from aP.
   - i = index of R_a^{host} from a.
4. Parse the sub-species in S_i^{sub}. For each A_{ix}, parse the $jobR$ table again and for every entry, where P_q is a slave and A_i is a host, check if all applicable regulation rules are simultaneously fulfilled by the

Boolean representation S_{ix}^{sub} (in the same way as described in the C table construction algorithm Pt. 6 - 9). If this is the case, set $P_{ixq}^{prot2phen} = true$.

3.5 Simulation and analysis

3.5.1 Definition of dynamical variables

Following variables change in the course of simulation:

- $\mathbb{S}$ (number array of dimensions $\{n, s\}$). Each entry $\mathbb{S}_{ix}$ is a total concentration of a protein or gene sub-species A_{ix} with dynamics described in Eq. (3.5).
- $\mathbb{T}$ (number vector of length $\{n\}$). Each entry $\mathbb{T}_i$ is a total concentration of a protein or gene species A_i as described in Eq. (3.12).
- $\mathbb{P}$ (number vector of length $\{o\}$) - Each entry $\mathbb{P}_i$ is the value of a real-valued phenomenon with dynamics described in Eq. (3.10 - 3.11).
- $\mathbb{B}$ (number vector of length $\{p\}$) - Each entry $\mathbb{B}_i$ is the value of a Boolean phenomenon with dynamics described in Eq. (3.11).

Initial conditions:

$$\mathbb{S}_{ix} = S_{ix}^{tot} \tag{3.1}$$
$$\mathbb{T}_i = a_i^{ini} \tag{3.2}$$
$$\mathbb{P}_i = aP_i^{ini} \tag{3.3}$$
$$\mathbb{B}_i = aB_i^{ini} \tag{3.4}$$

In the kinase-phosphatase example, the dynamical variable $\mathbb{S}$ corresponds to the sub-species listed in Section 2.3.2 with following initial values: $[B_{000} = 1]$, $[B_{100} = 0]$, $[B_{010} = 0]$, $[B_{011} = 0]$, $[A1_0 = 1]$, $[A1_1 = 0]$, $[A2_0 = 1]$, $[A2_1 = 0]$.

3.5.2 Derivation of equations from message board

The equations (3.5) - (3.12) describing the dynamical behavior of a simulated system are derived automatically from the message board tables in

following way:

Elementary mass action equations:

$$\begin{aligned} d\mathbb{S}_{ix}/dt &= -\Sigma_y \mathbb{S}_{ix}\left\{C^{on}_{ixy}\, C^{reg}_{ixy}\, \beta + C^{off}_{ixy} + \delta\right\} \\ &\quad + \Sigma_y \mathbb{S}_{ix}\left\{C^{on}_{iyx}\, C^{reg}_{iyx}\, \beta + C^{off}_{iyx} + \delta\right\} \\ &\quad - S^{degr}_{ix}\mathbb{S}_{ix} \\ &\quad + S^{synth}_{ix}\gamma \\ &\quad + \pi \end{aligned} \tag{3.5}$$

Where:

$$\beta = \begin{cases} \Sigma_z \mathbb{S}_{jz} R^{bind}_{jikz} R^{bindreg}_{jikz} & \text{if } I^{type}_{jik} = \text{'B'} \text{ where: } \{j,k\} = C^{bind}_{ixy} \\ (\Sigma_z \mathbb{S}_{jz} R^{enz}_{jikz} R^{enzreg}_{jikz})/(\Sigma_z \mathbb{S}_{jz} R^{enz}_{jikz}) & \text{if } I^{type}_{jik} = \text{'E'} \text{ where: } \{j,k\} = C^{enz}_{ixy} \end{cases} \tag{3.6}$$

$$\delta = \begin{cases} (\Sigma_z \mathbb{S}_{jz} R^{degr}_{jikz} S^{degr}_{jz}/(\Sigma_z \mathbb{S}_{jz} R^{degr}_{jikz}) & \text{if } k^{on}_{yx} \neq 0 \text{ where: } \{j,k\} = C^{degr}_{ixy} \\ 0 & \text{if } k^{on}_{yx} = 0 \end{cases} \tag{3.7}$$

$$\gamma = \begin{cases} (\Sigma_z \mathbb{S}_{jz} S^{synth}_{jz})/(\mathbb{S}_{jz}) & \text{if } a^{type}_j =' gene' \text{ where: } j = a^{ref}_i \\ 1 & \text{if } a^{type}_j \neq' gene' \end{cases} \tag{3.8}$$

$$\pi = \begin{cases} \Sigma_q \mathbb{P}_q P^{phen2prot}_{ixq} & \text{if } P^{phen2prot}_{ixq} \geq 0 \\ \Sigma_q \mathbb{P}_q P^{phen2prot}_{ixq} \mathbb{S}_{ix} & \text{if } P^{phen2prot}_{ixq} < 0 \end{cases} \tag{3.9}$$

Equations for phenomenological terms:

$$\mathbb{P}_q = \Sigma\Sigma_{ix} P^{prot2phen}_{ixq} \mathbb{S}_{ix} \tag{3.10}$$

$$jobP_i(\mathbb{P}, \mathbb{B}) \tag{3.11}$$

Conservation relationships:

$$\mathbb{T}_i = \Sigma_x \mathbb{S}_{ix} \tag{3.12}$$

In the kinase-phosphatase example, a following set of equations is automatically composed by the simulation framework:

$$\begin{aligned}
d[B_{000}]/dt &= -k_1[B_{000}][A1_0] + k_{-1}[B_{100}] + k_{22}[B_{011}] \\
d[B_{100}]/dt &= k_1[B_{000}][A1_0] - k_{-1}[B_{100}] - k_{11}[B_{100}] \\
d[B_{010}]/dt &= -k_2[B_{010}][A2_0] + k_{-2}[B_{011}] + k_{11}[B_{100}] \\
d[B_{011}]/dt &= k_2[B_{010}][A2_0] - k_{-2}[B_{011}] - k_{22}[B_{011}] \\
d[A1_0]/dt &= -k_1[B_{000}][A1_0] + (k_{-1} + k_{11})[A1_1] \\
d[A1_1]/dt &= +k_1[B_{000}][A1_0] - (k_{-1} + k_{11})[A1_1] \\
d[A2_0]/dt &= -k_2[B_{010}][A2_0] + (k_{-2} + k_{22})[A2_1] \\
d[A2_1]/dt &= +k_2[B_{010}][A2_0] - (k_{-2} + k_{22})[A2_1]
\end{aligned} \tag{3.13}$$

Note, that the job file for this system presented in Fig. 3.1 contains only 2 rows, compared to 8 ODEs that are automatically derived from this input (Equation System 3.13), which demonstrates the advantage of automation already for relatively simple systems.

3.5.3 Numerical integration

The automatically derived Eq. (3.5) - (3.12) are subsequently integrated numerically using a standard Runge-Kutta fourth order algorithm [66] with dynamical time step size.

For any first-order ODE system in a general form: $d\mathbb{S}/dt = f(\mathbb{S}, t)$ with initial conditions $\mathbb{S}(t_0) = \mathbb{S}_0$, the fourth-order Runge-Kutta algorithm allows stepwise calculation of $\mathbb{S}(t)$ using following general form:

$$\mathbb{S}_{t+dt} = RK_4(\mathbb{S}_t) \tag{3.14}$$

where:

$$RK_4(\mathbb{S}_t) = \mathbb{S}_t + \frac{dt}{6}(Q_1 + 2Q_2 + 2Q_3 + Q_4) \tag{3.15}$$

$$\begin{aligned}
Q_1 &= f(\mathbb{S}, & t) \\
Q_2 &= f(\mathbb{S} + \tfrac{dt}{2} Q_1, & t + \tfrac{dt}{2}) \\
Q_3 &= f(\mathbb{S} + \tfrac{dt}{2} Q_2, & t + \tfrac{dt}{2}) \\
Q_4 &= f(\mathbb{S} + dt\, Q_3, & t + dt)
\end{aligned}$$

As defined in Eq. (3.15), the calculated variable change per time step ($\mathbb{S}_{t+dt} - \mathbb{S}_t$) is a weighted average of four estimates (Q_1 to Q_4) with total accumulated

error $\epsilon_T = dt^4$ and error per step $\epsilon = dt^5$. Calculations presented in this work were conducted with $dt <= 0.001$ which gives error values $\epsilon_T \leq 10^{-12}$ and $\epsilon <= 10^{-15}$.

Additionally, a dynamical time step control based on the conservation equations has been implemented. For this purpose, Eq. (3.5) is split into two parts: conservation-dependent (f_1) and conservation-independent (f_2):

$$\begin{aligned} f_1(\mathbb{S}) &= -\Sigma_y \mathbb{S}_{ix} \left\{ C^{on}_{ixy}\, C^{reg}_{ixy}\, \beta + C^{off}_{ixy} + \delta \right\} \\ &\quad + \Sigma_y \mathbb{S}_{ix} \left\{ C^{on}_{iyx}\, C^{reg}_{iyx}\, \beta + C^{off}_{iyx} + \delta \right\} \end{aligned} \tag{3.16}$$

$$f_2(\mathbb{S}) = -S^{degr}_{ix} \mathbb{S}_{ix} + S^{synth}_{ix} \gamma + \pi \tag{3.17}$$

Where β, δ, γ, π, are defined as for Eq. (3.5).

For every species A_i, The f_1 captures changes in concentration of its sub-species A_{ix} (variable $\mathbb{S}_{ix}$) resulting from binding and enzymatic interactions. Thus, these changes can only affect distribution of sub-species within the total concentration of A_i but not the absolute value of this concentration $\mathbb{T}_i$:

$$\text{IF } (\mathrm{S}_{t+dt)} = f_1(\mathbb{S}_t, \mathbb{T}_t)\,) \;\Rightarrow\; \mathbb{T}_{t+dt} = \mathbb{T}_t. \tag{3.19}$$

On contrary to f_1, the f_2 describes changes of $\mathbb{S}_{ix}$ resulting from synthesis or degradation, which can by definition alter the absolute value $\mathbb{T}_i$.

These properties of f_1 and f_2 are utilized for controlling the time step size when calculating $\mathbb{S}_{(t+dt)}$, from $\mathbb{S}_t$, with use of following algorithm:

1. Set $dt = dt_0$.
2. Calculate $f_1(\mathbb{S}_t)$ using the RK_4 procedure and the conservation Eq. (3.12).
3. Check condition (3.18) and eventually redo previous steps with $dt_n < dt_{(n-1)}$ until this condition is fulfilled.
4. Calculate $f_2(\mathbb{S}_t)$ and Eq. (3.10) using the RK_4 procedure.
5. Calculate the algebraic equations Eq. (3.11) - (3.12).

As evident from the above algorithm, dynamical time step size control was applied only to the conservation-sensitive f_1 part of Eq. (3.5). However, we assume this extent of control to be sufficient since f_2 contains only rate constants that refer to phenomenological description of processes happening on slower time scales (synthesis and degradation). Thus, it seems reasonable to expect the rate constants in f_2 to be of one or more orders of magnitude lower than the rate constants in f_1 and thus not carry risk of driving the system into explosion.

Another important aspect of the time-course integration reliability is the open question of updating strategy while executing the user-defined algebraic relationships between phenomena defined in Eq. (3.11). Currently, Synchronous updating is implemented.

3.5.4 Steady state analysis

Time courses of species and sub-species concentrations ($\mathbb{T}(t)$ and $\mathbb{S}_(t)$) calculated by means of numerical integration as described above can eventually stabilize at a set of steady-state values that corresponds to a specific physiological state of the cell.

Possible steady states of an ODE system can be inferred by examination of eingenvalues of the Jacobian matrix representing this system [22]. It remains to be determined, if an automated eigenvalue derivation is possible for equation systems in the form of Eq. (3.5) - (3.12). So far, the analytical capabilities of aceSim are limited to identification of stable steady states by running the simulation for a sufficiently long time. More complicated forms of system's behavior, such as limit cycles, can be only examined by visual inspection of time course plots. Thus, the functionalities related to steady-state examination should be subject of future development.

3.5.5 Parameter exploration and bifurcation diagrams

One of typical approaches is to test system's behavior for various values of parameters, such as rate constants or initial concentration values. Automation of this functionality is implemented by extending the job file with two lists:

- $jobSig$ (string vector) - list of labels that can be entered into the job file instead of numeric values (eg: $jobI_a^{k_n}$ = 'A to B binding rate' instead $jobI_a^{k_n} = 0.001$).

- *jobSigRange* (number array) - a mapping of the *jobSig* list to value ranges, e.g. 'A to B binding rate' $= 0.001, 0.002, 0.003$ etc.

By this means, the core program can be wrapped in loops corresponding to the dimensions of *jobSigRange* and return an array of steady-state values corresponding to different parameter ranges. currently, the parameter space can be scanned with a superposition of two parameter sets in order to obtain two-dimensional parameter sensitivity surfaces or bifurcation diagrams (cf. Sec. 4.2.3).

3.5.6 Comparison with experimental data

A further extension was implemented to allow comparison of the simulation results with two data types:

- Downloaded data array. This feature can be applied to compare simulation results with experimental data.
- Data array automatically generated by integration of a manually-entered ODE system. This requires at least some of variable names and parameter labels of the entered system to be identical with names of the agents and rate constant labels in the job file describing the simulated system. This feature allows comparison of simulation results of two alternative descriptions of the same system, e.g. using EMA vs. AMA formalisms. This approach was applied to obtain the results presented in Chpt. 4.

The comparison refers to two aspects of dynamics of a given variable: time course and value at steady state.

3.5.7 Mutation and knock-out analysis

The presented framework enables testing of many real-life relevant hypotheses such as:

- gene knock-out over/under expression (by altering initial concentrations)
- effects of mutations on changed enzymatic activity (by altering rate constants).
- effects of pharmacological interventions on changed enzymatic activity (by altering both rate constants and initial concentrations).

Chapter 4

Simulation of selected biological systems

The kinase-phosphatase motif described in Sec. 2.3.2 is perhaps the most common building block of enzyme signaling cascades and other biological systems [181], [183], [208]. In this chapter, we will analyze the behavior of a basic kinase-phosphatase motif (Sec. 4.1.1) and increasingly complex systems composed of this motif: a linear cascade (Sec. 4.1.2) and several branched cascades, including a core mechanism of the entry into mitosis in the eukaryotic cell cycle (G2/M transition, Sec. 4.1.3).

We will investigate the behavior of the above systems by means of parameter sensitivity and bifurcation analysis. We will also examine the influence of Michaelis-Menten approximation and combinatorial complexity on the kinetic modeling of these systems by comparing outcomes of EMA and AMA.

4.1 Description of analyzed systems

The Sec. 4.1.1 presents a standard Michaelis-Menten form for the kinase-phosphatase motif (Eq. 4.1), from which AMA descriptions of other analyzed systems can be easily derived as presented in Eq. (4.2)and Eq. (4.3) - (4.4).

The EMA descriptions are automatically generated by the simulation software aceSim described in this work (Chpt. 2 and 3). A manual approach would require writing of tens of equations (e.g. 36 for he branched cascade).

Thus, this investigation demonstrates well the modeling capabilities of the aceSim.

4.1.1 Kinase-phosphatase motif

A basic kinase-phosphatase motif consists of one kinase $A1$ and one phosphatase $A2$ acting on the same substrate protein $B1$ as depicted in Fig. 4.1 A.

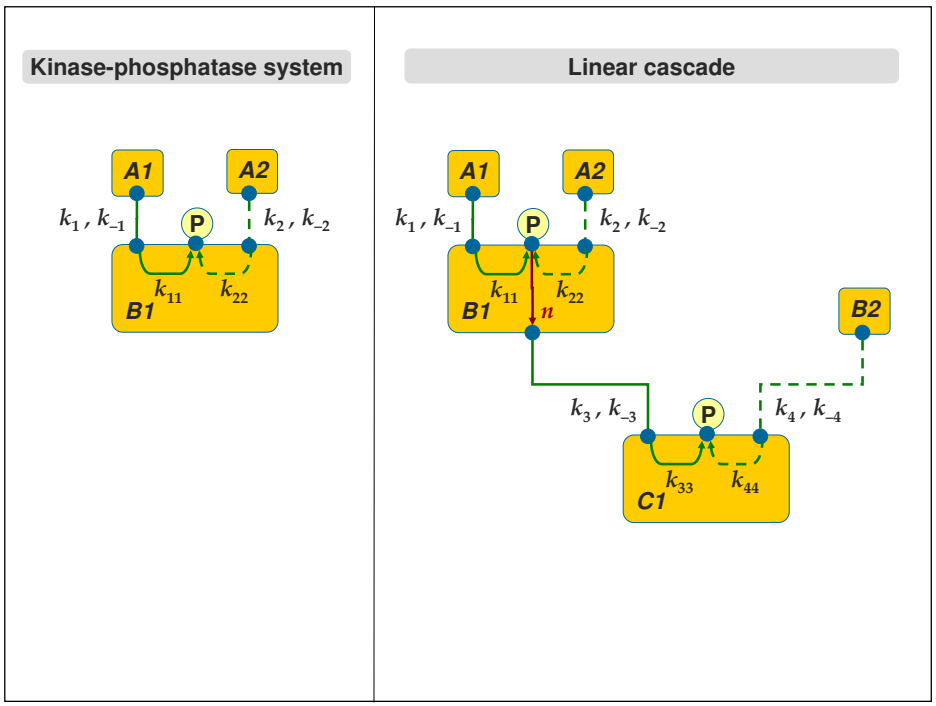


Figure 4.1: Graphical representation of a basic kinase-phosphatase motif (A) and an enzymatic cascade composed of two such motifs (B). Rate constants and arrows conform the formalism introduced in Sec. 2.2. In particular, the arrow with symbol 'n' indicates, that phosphorylation by $A1$ is necessary for $B1$ to carry out its enzymatic function on $C1$.

Elementary mass action description of the kinase-phosphatase motif is defined in the Equation System (3.13). For the same variables and parameters, an AMA description can be derived using a simplifying assumption that concentrations of enzyme-substrate complexes $[B1_{100}]$ and $[B1_{011}]$ are negligible in respect to the total substrate concentration $[B1_T]$, thus

$[B1_T] \approx [B1_{000}] + [B1_{010}]$. This results in following equation based on Michaelis kinetics [94]:

$$\frac{d[B1_{010}]}{dt} = \frac{k_{11}[A1_T][B1_{000}]}{[B1_{000}] + \frac{k_{-1}+k_{11}}{k_1}} - \frac{k_{22}[A2_T][B1_{010}]}{[B1_{010}] + \frac{k_{-2}+k_{22}}{k_2}} \tag{4.1}$$

Analytical steady-state solutions exist for both the EMA description (3.13) and AMA description (4.1). The EMA steady-state solution is a third - order polynomial, thus up to three steady-state solutions of the system are possible. This will be discussed in sec. 4.3.1 in a more detail. The AMA steady-state solution is a quadratic polynomial with only one non-negative root, known as the Goldbetter-Koshland function [94] and it has been applied, together with Eq. 4.1, in modeling of several MIN underlying such phenomena as cell division [55] or embryonic development [153]. However, it is not always clear if the assumptions underlying the derivation of these AMA steady-state forms are realistic for biological systems [41], [155]. As shown in Sec. 4.3, for some parameter ranges substantial discrepancies can occur between an EMA and AMA description of analyzed systems.

4.1.2 Linear cascade

Fig. 4.1 B presents a simple, linear enzymatic cascade, where the phosphorylated substrate $B1$ from previous example can itself act as a kinase on a downstream substrate $C1$. The $C1$ can also be dephosphorylated by a phosphatase $B2$. We further assume, that only the phosphorylated form of $B1$ is able to perform kinase activity.

An AMA description of such cascade can be easily derived from Eq. 4.1 upon analogue assumptions $[B1_T] \approx [B1_{000}]+[B1_{010}]$ and $[C1_T] \approx [C1_{000}]+[C1_{010}]$:

$$\begin{aligned}
\frac{d[B1_{010}]}{dt} &= \frac{k_{11}[A1_T][B1_{000}]}{[B1_{000}] + \frac{k_{-1}+k_{11}}{k_1}} - \frac{k_2[A2_T][B1_{010}]}{[B1_{010}] + \frac{k_{-2}+k_{22}}{k_2 2}} \\
\frac{d[C1_{010}]}{dt} &= \frac{k_{33}[B1_{010}][C1_{000}]}{[C1_{000}] + \frac{k_{-3}+k_{33}}{k_3}} - \frac{k_{44}[B2_T][C1_{010}]}{[C1_{010}] + \frac{k_{-4}+k_{44}}{k_4}}
\end{aligned} \tag{4.2}$$

Where: $B1_{000}$, $C1_{000}$ are unphosphorylated and $B1_{010}$, $C1_{010}$ are phosphorylated forms of $B1$ and $C1$, respectively. Rate constants k_n as in Fig. 4.1 B.

An EMA description of the above cascade contains 18 equations and it was generated automatically by aceSim.

4.1.3 Branched cascade and G2/M transition pathway.

Fig. 4.2 A presents a branched cascade, where both the kinase $B1$ and phosphatase $B2$ acting on substrate $C1$ are themselves regulated by a kinase/phosphatase pair, $A1/A2$ and $D1/D2$, respectively. We further assume, that only the phosphorylated forms of $B1$ and $B2$ are able to perform enzymatic activity on $C1$.

An AMA description of such branched cascade can be easily derived from Eq. 4.1 upon following assumptions: $[B1_T] \approx [B1_{000}] + [B1_{010}]$, $[B2_T] \approx [B2_{000}] + [B2_{010}]$ and $[C1_T] \approx [C1_{000}] + [C1_{010}]$:

$$\frac{d[B1_{010}]}{dt} = \frac{k_{11}[A1_T][B1_{000}]}{[B1_{000}] + \frac{k_{-1}+k_{11}}{k_1}} - \frac{k_{22}[A2_T][B1_{010}]}{[B1_{010}] + \frac{k_{-2}+k_{22}}{k_2}} \tag{4.3}$$

$$\frac{d[C1_{010}]}{dt} = \frac{k_{33}[B1_{010}][C1_{000}]}{[C1_{000}] + \frac{k_{-3}+k_{33}}{k_3}} - \frac{k_{44}[B2_{010}][C1_{010}]}{[C1_{010}] + \frac{k_{-4}+k_{44}}{k_4}} \tag{4.4}$$

$$\frac{d[B2_{010}]}{dt} = \frac{k_{55}[D1_T][B2_{000}]}{[B2_{000}] + \frac{k_{-5}+k_{55}}{k_5}} - \frac{k_{66}[D2_T][B2_{010}]}{[B2_{010}] + \frac{k_{-6}+k_{66}}{k_6}} \tag{4.5}$$

Where: $B1_{000}$, $B2_{000}$, $C1_{000}$ are unphosphorylated and $B1_{010}$, $B2_{010}$, $C1_{010}$ are phosphorylated forms of $B1$, $B2$ and $C1$, respectively. Rate constants k_n as in Fig. 4.2 A.

An EMA description of the branched cascade contains 28 equations and was generated automatically by aceSim.

Such branched cascade can be further modified by adding feedback loops, leading to a structure presented in Fig. 4.3 B. This system constitutes a core mechanism of entry into mitosis in eukaryotic cells, where the cyclin-dependent kinase cdk1 ($C1$) is inactivated by the kinase Wee1 ($B1$) and activated by the phosphatase CDC25 ($B2$) [24], [145]. The activated cdk1 in complex with a regulatory protein Cyclin B (omitted here for simplicity) is called the Mitosis Promoting Factor (MPF) and can itself inhibit Wee1 and activate CDC25, which results in a double negative and a double positive feedback loop, respectively [112], [120] [144].

The whole MIN responsible for cell cycle progression of course far more complicated [44] (see also [195] for a recent review). However, the above idealized mechanism explains well the peak of activated MPF triggering entry into mitosis and has been applied at the heart of several cell cycle models [63], [150], [161].

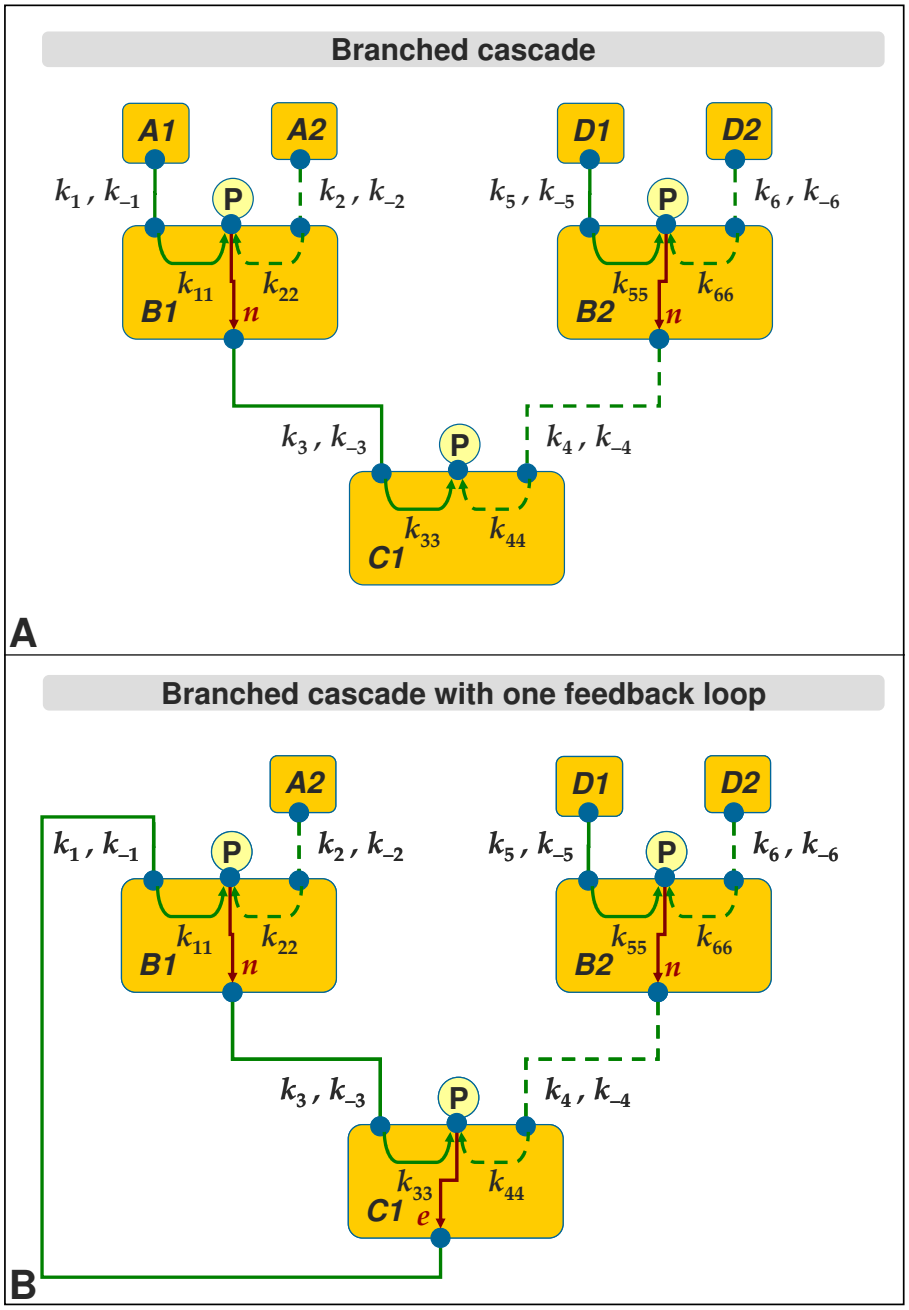


Figure 4.2: Graphical representation of a branched enzymatic cascade composed of 3 kinase-phosphatase motifs, without feedback loops (A) or with one positive feedback loop (B). Rate constants and arrows conform the formalism introduced in Sec. 2.2. In particular, an arrow with symbol 'n' indicates, that phosphorylation is necessary for a given protein to carry out its enzymatic function, whereas an arrow with symbol 'e' indicates that phosphorylation disables this function.

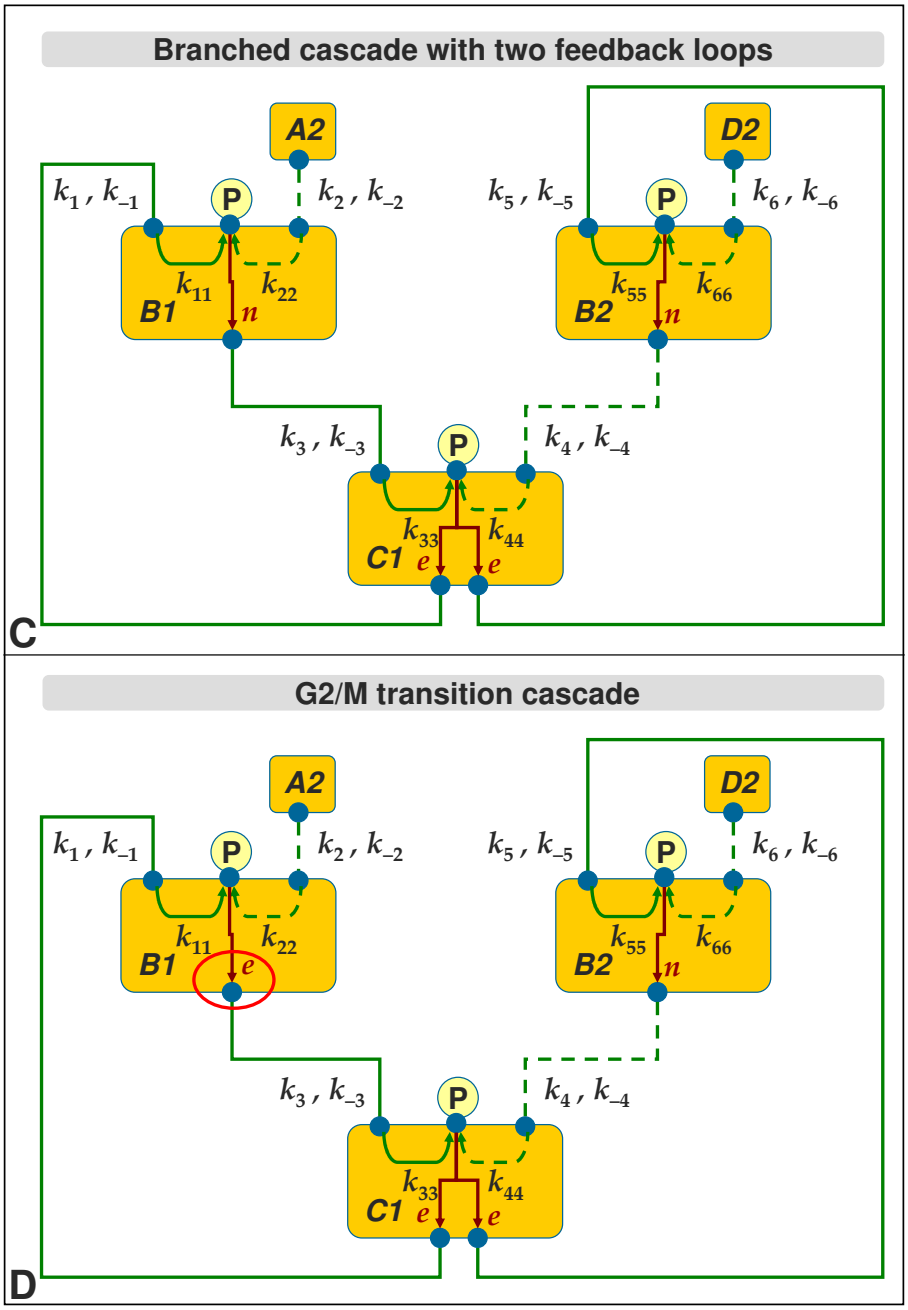


Figure 4.3: Graphical representation of a branched enzymatic cascade composed of 3 kinase-phosphatase motifs, with one positive and one negative feedback loop (A) or with a positive and a double-negative feedback loop (B). The structure depicted in (B) represents a basic mechanism of the G2/M transition in the eukaryotic cell cycle. Rate constants and arrows conform the formalism introduced in Sec. 2.2. In particular, an arrow with symbol 'n' indicates, that phosphorylation is necessary for a given protein to carry out its enzymatic function, whereas an arrow with symbol 'e' indicates that phosphorylation disables this function.

In order to investigate the role of each structural component in the G2/M transition pathway behavior, we will modify and analyze the basic branched pathway in a step-by step manner:

- First, a feedback loop can be added by replacing the upstream kinase $A1$ by the downstream kinase $C1$ (Fig. 4.3 A). To conform the aimed G2/M transition pathway structure, we assume that only unphosphorylated $C1$ can perform kinase activity on $B1$. This means, that the feedback has a negative character: $C1$ activates $B1$, which in turn inactivates $C1$. In the AMA description, an introduction of such negative feedback requires replacing the term $[A1_T]$ in Eq. 4.3 with $[C1_{000}]$.

- Further, a second feedback loop can be added by replacing the upstream kinase $D1$ by the downstream kinase $C1$ (Fig. 4.2 B). Again, to conform the aimed G2/M transition pathway structure, we assume that only unphosphorylated $C1$ can perform kinase activity on $B2$. Since $B2$ is a phosphatase, such feedback has a double positive character: $C1$ activates $B2$, which in turn activates $C1$ by dephosphorylation. In the AMA description, an introduction of such positive feedback requires replacing the term $[D1_T]$ in Eq. 4.5 with $[C1_{000}]$.

- Finally, we modify the first negative feedback into a double negative feedback by assuming that only unphosphorylated $B1$ can perform kinase activity on $C1$ (Fig. 4.3 B): $C1$ and $B1$ can inactivate each other by phosphorylation. This requires replacing the term $[B1_{010}]$ in Eq. 4.4 with $[B1_{000}]$. Such antagonism can intuitively lead to two states, one with activated $B1$ / inactivated $C1$ (i.e. low MPF concentration) and one with a reverse situation (i.e. high MPF concentration). These states correspond to the G2 and M phases of cell cycle and the described pathway structure enables a rapid G2/M transition; we will investigate this behavior in a more detail in Sec. 4.3.3.

Addition of interactions between $C1$ and $B1$ and $B2$ increases the number of possible combinatorial states of $C1$, thus an EMA description of such system generated by the aceSim contains 36 equations.

4.2 Methods.

4.2.1 Definition of response

In general, we will define here the system's response π as a fractional concentration of all phosphorylated forms $[A^P]$ of a given protein A to its total concentration $[A_T]$ at the steady state, i.e. $\pi = [A^P]/[A_T]$. For the kinase-phosphatase motif, the response vector contains only π_1, i.e. the phosphorylated $B1$. For other systems, it contains additionally π_2 and π_3, i.e. phosphorylated $B2$ and $C1$, respectively. Such fractional definition implies that π is dimensionless and can only take values between 0 and 1.

The difference between EMA and AMA description has also impact on the response. For the kinase-phosphatase motif, the elementary response $\pi_1^E = ([B1_{010}]+[B1_{011}])/[B1_T]$. However, since the term $[B1_{011}]$ is neglected in AMA description, the approximated response $\pi_1^A = [B1_{010}]/[B1_T]$. similarly for other systems, $\pi_2^A = [B2_{010}]/[B2_T]$ and $\pi_3^A = [C1_{010}]/[C1_T]$. Due to a large number of combinatorial forms in the EMA description, the elementary π_1^E, π_2^E and π_3^E are defined automatically by aceSim for each system by subsuming concentrations of all phosphorylated sub-species of $B1$, $B2$ and $C1$, respectively.

4.2.2 Definition of parameters

The investigated parameters were defined individually for each system in a sequential manner as described below.

Parameter set for the kinase-phosphatase motif

For the kinase-phosphatase motif, we have set following parameters: a) relative enzyme-substrate binding rate k_1/k_2(dimensionless), b) relative enzyme-substrate unbinding rate k_{-1}/k_{-2}(dimensionless), c) relative transformation rate k_{11}/k_{22} (dimensionless), d) relative kinase-phosphatase concentration $[A1]/[A2]$ (dimensionless), e) absolute kinase-substrate binding rate k_1 ($\mu M^{-1}s^{-1}$), f) absolute kinase-substrate unbinding rate k_{-1} (s^{-1}), g) absolute phosphorylation rate k_{11} ($\mu M^{-1}s^{-1}$) and h) absolute kinase concentration $[A1]$ (μM). The relative rates a/b were investigated in the range of $\log a/b \in \langle -2,\ 2\rangle$ and the absolute rates a in the range $a \in \langle 0,\ 1\rangle$ with the corresponding dephosphorylation reaction parameter b (k_2, k_{-2}, k_{22} and $[A2]$, respectively) fixed at $a/b = 2$.

Parameter set for the linear cascade

Based on the parameter sensitivity results for the kinase-phosphatase motif (Sec. 4.3.1), only the relative (thus dimensionless) rates k_1/k_2, k_{-1}/k_{-2} and k_{11}/k_{22} were chosen for further investigation in the linear cascade.

The relative concentrations $[A1]/[A2]$ were excluded, because they showed similar sigmoidal sensitivity as the relative rates k_1/k_2 and k_{11}/k_{22}. Moreover, these rates determine concentrations of active forms of downstream enzymes; investigation of influence of the active, not total enzyme concentrations seems to be biologically more meaningful. The total concentrations of all enzymes were fixed at 1.

The absolute rates and concentrations were not further investigated, because they showed in general relatively flat sensitivity curves and in some cases big discrepancies between π^E and π^A (Sec. 4.3.1). These discrepancies would probably dominate the behavior of the cascade, interfering with investigation of other features.

Parameter set for the branched cascade

Based on the parameter sensitivity results for the linear cascade (Sec. 4.3.2), only the relative (thus dimensionless) transformation rates $\kappa_{12} = k_{11}/k_{22}$, $\kappa_{34} = k_{33}/k_{44}$ and $\kappa_{56} = k_{55}/k_{66}$ were chosen for further investigation in the branched cascade. This is because on both levels of the linear cascade, i.e. π_1 and π_2, the parameters k_1/k_2 and k_{11}/k_{22} showed almost identical sigmoidal response and the parameter k_{-1}/k_{-2} a relatively flat response (Sec. 4.3.2). Reaction rates not included in a given κ parameter were fixed at $k_n = 1$, $k_{-n} = 0$, $k_{nn} = 1$, where $n = 1, \ldots, 6$.

4.2.3 Calculations.

Based on the response and parameter sets defined above, we have calculated response π levels at steady state for two-dimensional combinations of parameter values. Initial levels of π, ranging from 0 to 1, were treated as one of the parameters, which results in bifurcation plots for the second parameter. This method allows only investigation of steady states stable in respect to a single dimension of the parameter space. More fundamental analytical approaches , e.g. based on Lyapunov exponents, do not seem to be feasible for large EMA systems.

The approximated response π^A was measured for the kinase-phosphatase motif, linear cascade and branched cascade by integrating the Eq. Sys-

tems (4.1), (4.2) and (4.3 - 4.4), respectively, using a standard Runge-Kutta 4th order algorithm. Modifications of the branched cascade were considered in the Eq.(4.3 - 4.4) as described in Sec. 4.1.3. The elementary response π^E was calculated using the aceSim framework.

In the obtained bifurcation plots (Fig. 4.4, 4.6, 4.8, 4.10, and 4.12), the bifurcation parameter is indicated on the abscissa and the measured π on the ordinate. The π^E is marked with blue diamonds and the π^A is marked with red circles.

For selected parameter values, time course plots of the π are presented with time marked on the abscissa and the π on the ordinate. The π^E is marked with light and dark blue lines and the π^A is marked with magenta and red lines.

For validation, we have recalculated the steady state levels for the kinase-phosphatase motif from existing analytical solutions of both EMA and AMA description ([94]) for the same bifurcation parameter values using $Mathematica^{TM}$ and obtained results of relative difference $< 10^{-5}$ per data point to both π_1^E and π_1^A.

4.3 Results

4.3.1 Simulation of a kinase-phosphatase motif

Fig. 4.4 presents response sensitivity of the basic kinase-phosphatase motif described in Sec. 4.1.1 for the parameter set defined in Sec. 4.2.2 and Fig. 4.5 depicts time courses of the phosphorylated substrate concentration for selected parameter values.

The relative rates k_1/k_2 (Fig. 4.4 A), k_{11}/k_{22} (Fig. 4.4 C) and concentrations $[A1]/[A2]$ (Fig. 4.4 D) show a similar sigmoidal sensitivity pattern. The transition from low ($\pi_1 \approx 0$) to high ()$\pi_1 \approx 1$) response values occurs around the point of equal rates (concentrations), i.e. $\log a/b = 0$, with π_1^A having slightly steeper sensitivity curves than π_1^E. Still, steepness of this transition does not have a jump character and thus does not indicate ultrasensitivity [94]. The relative rate k_{-1}/k_{-2} shows also a sigmoidal, but relatively flattened sensitivity pattern (Fig. 4.4 B). The relaxation time of π is similar for different relative parameter values and also between π_1^A and π_1^E for a given parameter value (Fig. 4.5 A-D).

The absolute parameters are set in such a way, that for a changing value of a rate (concentration) a related to the phosphorylation reaction (k_1, k_{-1}, k_{11} and $[A1]$), a corresponding rate (concentration) b of the dephosphory-

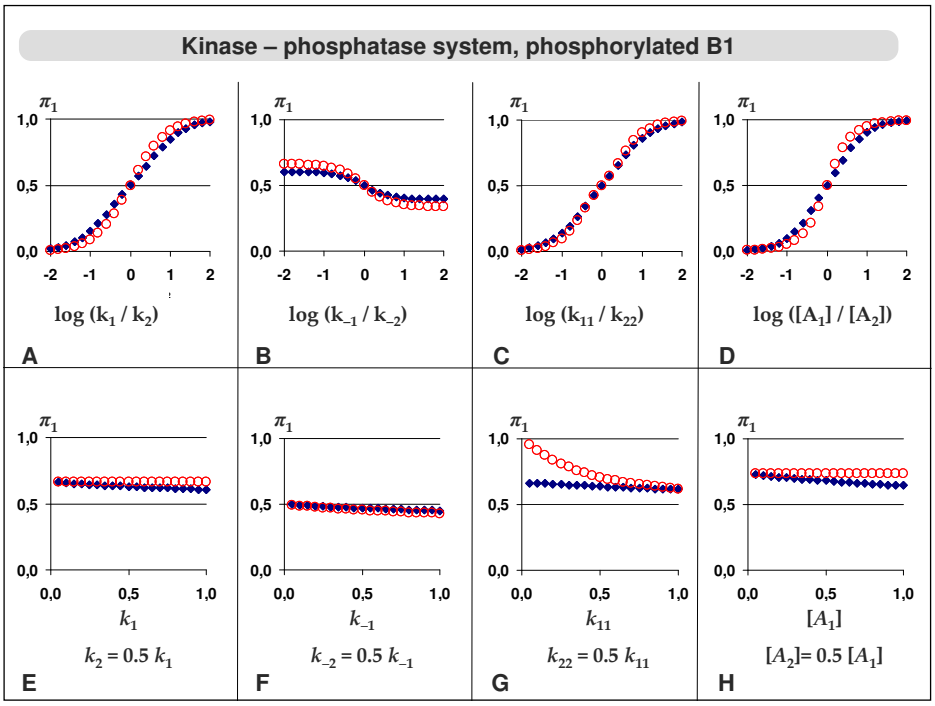


Figure 4.4: Bifurcation diagrams for different kinetic parameters of the kinase-phosphatase motif. For definition of the parameters k_n and $[A_n]$ and the response π_1, see Sec. 4.2. Blue diamonds - steady state π_1 levels calculated with EMA, red circles - steady state π_1 levels calculated with AMA.

lation reaction (k_2, k_{-2}, k_{22} and $[A2]$, respectively) is set to $a/b = 2$. The π relaxes substantially quicker for higher absolute parameter values, except for the kinase unbinding rate k_{-1}, where this effect is not significant (Fig. 4.5 E - H).

The π_1^E stabilizes at a constant level of roughly 2/3 (Fig. 4.4 E - H). However, π_1^A shows some substantial discrepancies here, especially a much higher sensitivity to the transformation rate k_{11} (Fig. 4.4 G). Moreover, π^E slightly diminishes when the absolute value of parameters increases. This can be attributed to a simultaneously decreasing amount of phosphatase-product complex. This complex is neglected in the AMA description, thus π_1^A is completely insensitive to kinase-substrate binding rate k_1 (Fig. 4.4 E) and kinase concentration $[A1]$ (Fig. 4.4 H).

Thus, the AMA description can produce overestimated response levels

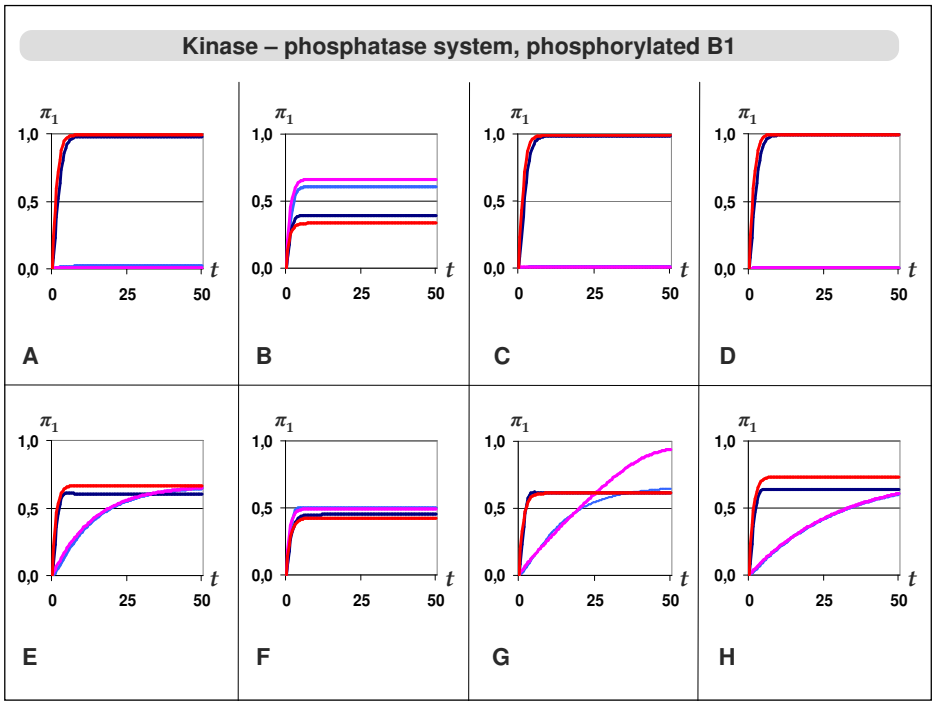


Figure 4.5: Time courses of phosphorylated protein $B1$ in the kinase-phosphatase motif. Tested kinetic parameters in **A** - **H** correspond to the parameters presented in Fig. 4.4 A - H, respectively. For definition of the parameters and the response π_1, see Sec. 4.2. Time courses are color-coded depending on parameter value p and response type, i.e. π^A vs. π^E. **A** - **D**: **magenta** - π_1^A, $p=-2$; **red** - π_1^A, $p=2$; **light blue** - π_1^E, $p=-2$; **dark blue** - π_1^E, $p=2$. **E** - **H**: **magenta** - π_1^A, $p=0.05$; **red** - π_1^A, $p=1.05$; **light blue** - π_1^E, $p=0.05$; **dark blue** - π_1^E, $p=1.05$.

for either very small transformation rates k_{11} and k_{22} (and thus possibly low Michaelis constants) or when enzyme-substrate concentrations are of the same order of magnitude (in the presented case: $[A1] = [A2] = [B1] = 1$), which is an realistic assumption for MIN based on kinase-phosphatase motifs ([41]).

As shown in Fig. 4.4, the response π_1, both π_1^A and π_1^E, is monostable towards all tested parameter ranges. This is consistent with π_1^E recalculation using the analytical steady state solution in form of a 3rd order polynomial as described in Sec. 4.2.3, where for each tested bifurcation parameter value, only one of the obtained roots was biologically realistic, i.e. $0 \leq \pi_1^E \leq 1$.

4.3.2 Simulation of a linear cascade

Fig. 4.6 shows response sensitivity of the linear cascade described in Sec. 4.1.2 to the parameter set defined in Sec. 4.2.2 and Fig. 4.7 depicts concentration time courses of phosphorylated proteins in the cascade for selected parameter values.

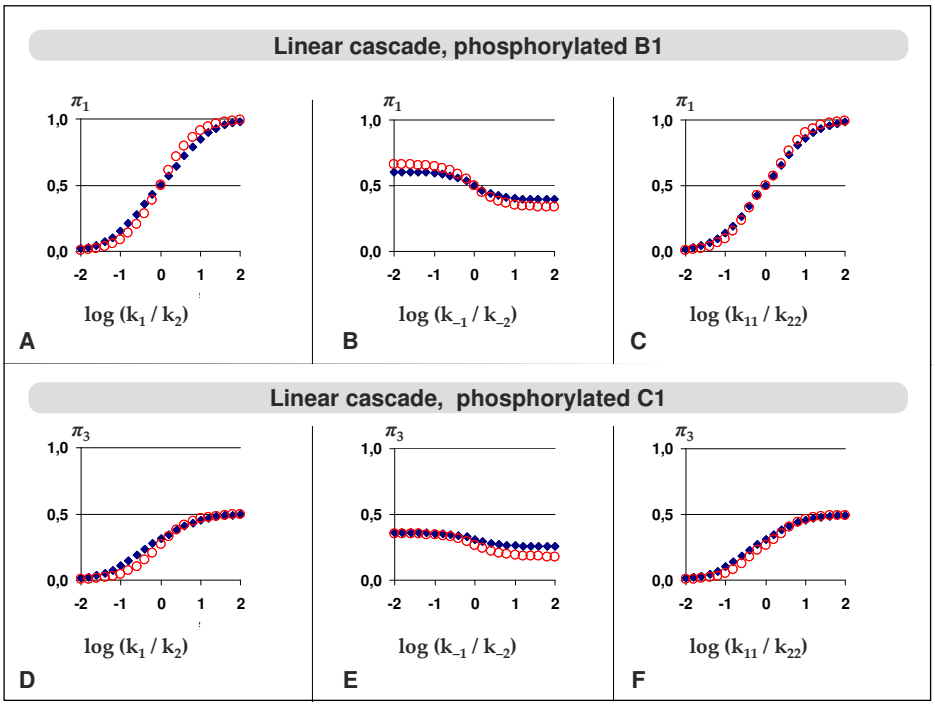


Figure 4.6: Bifurcation diagrams for different kinetic parameters of the linear cascade. For definition of the parameters k_n and the response π_1 and π_2, see Sec. 4.2.2. Blue diamonds - steady state π_1 and π_2 levels calculated with EMA, red circles - steady state π_1 and π_2 levels calculated with AMA

At the upstream cascade step, π_1, the sensitivity and dynamics of the response are identical as for the basic kinase-phosphatase motif (Fig. 4.6 A-C and Fig. 4.7 A-C).

At the downstream cascade step, π_3, the sigmoidal response sensitivity for k_1/k_2 and k_{11}/k_{22} is substantially damped: π_3 reaches only the maximal level of 0.5 compared to 1 achieved by π_1 for the same parameter ranges (Fig. 4.6 D-F). This is because the level of π_3 is additionally determined by the concentration of active $B1$ and $B2$. While the level of active $B2$

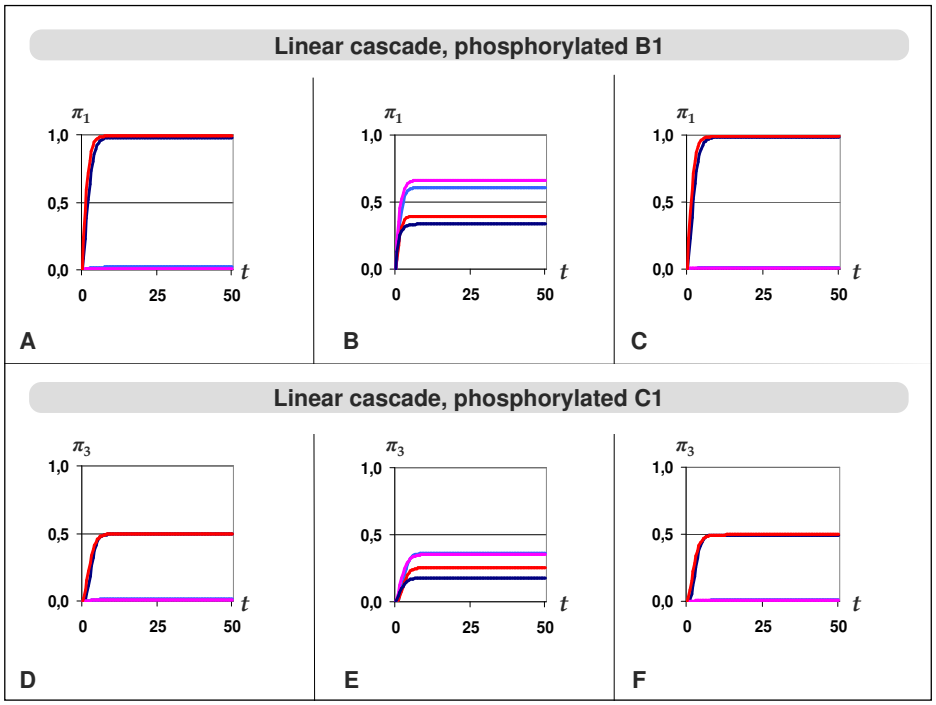


Figure 4.7: Time courses of phosphorylated proteins $B1$ (**A** - **C**) and $C1$ (**D** - **F**) in the linear cascade. Tested kinetic parameters in **A** - **F** correspond to the parameters presented in Fig. 4.6 A - F, respectively. For definition of the parameters and the response π_1 and π_3, see Sec. 4.2. Time courses are color-coded depending on parameter value p and response type, i.e. π^A vs. π^E: **magenta** - π^A, $p=-2$, **red** - π^A, $p=2$, **light blue** - π^E, $p=-2$, **dark blue** - π^E, $p=2$.

is equal to total concentration $B2_T = 1$, the level of active $B1$ depends on its phosphorylation, i.e. is equal to $\pi_1 \leq 1$. This demonstrates how the superposition of basic kinase-phosphatase motif sensitivity leads to the dampening of the response level in a linear cascade. In Sec 4.3.3, we will investigate such superposition in the case when the activity of both $B1$ and $B2$ depends on their phosphorylation.

Additionally, the relaxation time of π_3 is slightly longer when compared with the time courses of π_1 for the same values of respective parameter (Fig. 4.7 D-F).

Again, both π_3^A and π_3^E are monostable and have similar sensitivity patterns for a given parameter, with π_3^A having slightly steeper curves than

π^E (Fig. 4.4 A-D). Both π_3^E and π_3^A stabilize after similar relaxation time (Fig. 4.5 A-D).

4.3.3 Simulation of a branched cascade and the G2/M transition pathway.

Response sensitivity of different branched cascade variations described in Sec. 4.1.3 to the parameter set defined in Sec. 4.2.2 is presented in Fig. 4.8, 4.10 and 4.12. Each figure presents the steady state phosphorylation level of one of the three main pathway components: upstream kinase $B1$ (π_1), upstream phosphatase $B2$ (π_2) and downstream kinase $C1$ (π_3), respectively. Time courses of π_1, π_2 and π_3 for selected parameter values are shown in Fig. 4.8, 4.10 and 4.12, respectively.

Branched cascade

In a basic branched cascade without feedback loops (Fig. 4.2 A), the phosphorylation level of $B1$, $B2$ and $C1$ is sensitive only to the parameters of reactions in which the given protein is directly involved; i.e. κ_{12} for $B1$ (Fig. 4.8 A - C and Fig. 4.9 A - C), κ_{56} for $B2$ (Fig. 4.10 A - C and Fig. 4.11 A - C) and κ_{34} for $C1$ (Fig. 4.10 A - C and Fig. 4.13 A - C). The sigmoidal sensitivity pattern and dynamics are identical here as in the basic kinase-phosphatase motif (cf. Fig. 4.4 C and Fig. 4.5 C).

However, π_3 is also sensitive to κ_{12} and κ_{56}. These parameters determine concentrations of active (i.e. phosphorylated) enzymes $B1$ and $B2$, which in turn directly influence the phosphorylation level of $C1$. The sensitivity pattern for κ_{12} and κ_{56} is a slightly damped sigmoid, but not as much as for the linear cascade, where the response reaches maximally $\pi_3 = 0.5$ (Sec. 4.3.2). This shows, that the response damping in a linear cascade can be reduced in a branched cascade, where both kinase and phosphatase determining the response are regulated. However, the relaxation time of π_3 in the branched cascade remains longer than in the linear cascade (Fig. 4.12 B).

Branched cascade with negative feedback

Addition of a negative feedback loop between $C1$ and $B1$ ($C1$ activates $B1$, which in turn inactivates $C1$; Fig. 4.2 B) causes a superimposed sensitivity pattern for response and parameters related directly to these two proteins, i.e. π_1, π_3 and κ_{12}, κ_{34}. This pattern has a form of a bisigmoidal curve , i.e. with one sigmoid leveling off around $\log \kappa = 0$ and other sigmoid starting

on the top of it (Fig. 4.8 D, E, Fig. 4.12 D, E) and it is also related to extended relaxation times, especially for $\kappa_{12} = 2$ and $\kappa_{34} = 2$ (Fig. 4.9 D, E, Fig. 4.13 D, E).

Furthermore, the feedback transmits the sensitivity of π_3 to κ_{56} onto π_1 with a dumped sigmoid pattern similar to the π_3 in linear cascade but with different dynamics (Fig. 4.8 F and Fig. 4.9 F). Due to lack of such a feedback link for π_2, its sensitivity pattern and dynamics remain unchanged from the case without feedback (Fig. 4.10 D - F and Fig. 4.11 D - F).

Branched cascade with positive and negative feedback

A further positive feedback loop can be added between $C1$ and $B2$ ($C1$ activates $B2$, which in turn activates $C1$ by dephosphorylation; Fig. 4.3 A). The positive feedback makes sigmoidal responses to κ_{34} and κ_{56} more extreme (Fig. 4.8 H, J, Fig. 4.10 H, J and Fig. 4.12 H, J). However, the relaxation times of π_3 get slightly longer when compared to the single feedback case (Fig. 4.13 H, J).

The sensitivity to κ_{12} exhibits a very interesting pattern with substantial discrepancies between π^A and π^E (Fig. 4.8 G, Fig. 4.10 G and Fig. 4.12 G). Investigation of time courses in the long run reveals also substantial differences in the dynamics of π^A vs. π^E for selected values of κ_{12} (Fig. 4.14). Especially, if $\log \kappa_1 = 2$, the π^E levels off continuously, while the π^A shows damped oscillations, which last at least 5 times longer than the relaxation time of π_1^E (Fig. 4.14 A). These oscillations are transmitted also to π_2^A and π_3^A (Fig. 4.14 B, C).

G2/M transition cascade

Reversal of the feedback between $C1$ and $B1$ in the G2/M transition cascade ($C1$ and $B1$ can inactivate each other by phosphorylation; Fig. 4.3 B) leads to ultrasensitive behavior of π_1, π_2 and π_3 for all tested parameters (Fig. 4.8 K - M, Fig. 4.10 K - M and Fig. 4.12 K - M). Interestingly, the ultrasensitve transition point is shifted from $\log \kappa = 0$ to $\log \kappa \approx 1$ or -1.

Through all tested cascade variations, the sensitivity curves for π^A are slightly steeper than for π^E; this holds even for the ultrasensitive curves of the G2/M transition cascade.

The antagonsim between $C1$ and $B1$ leads to bistability, which occurs for π_3 in the range of $\log \kappa_{12} > 1$ (Fig. 4.12 K). Strikingly, π_3^A exhibits bistability also for some ranges of κ_{34} and κ_{56}, which is not the case for π_3^E (Fig. 4.12 L - M). This demonstrates, that G2/M transition models based

on AMA can produce bistable behavior which does not exist according to the EMA description.

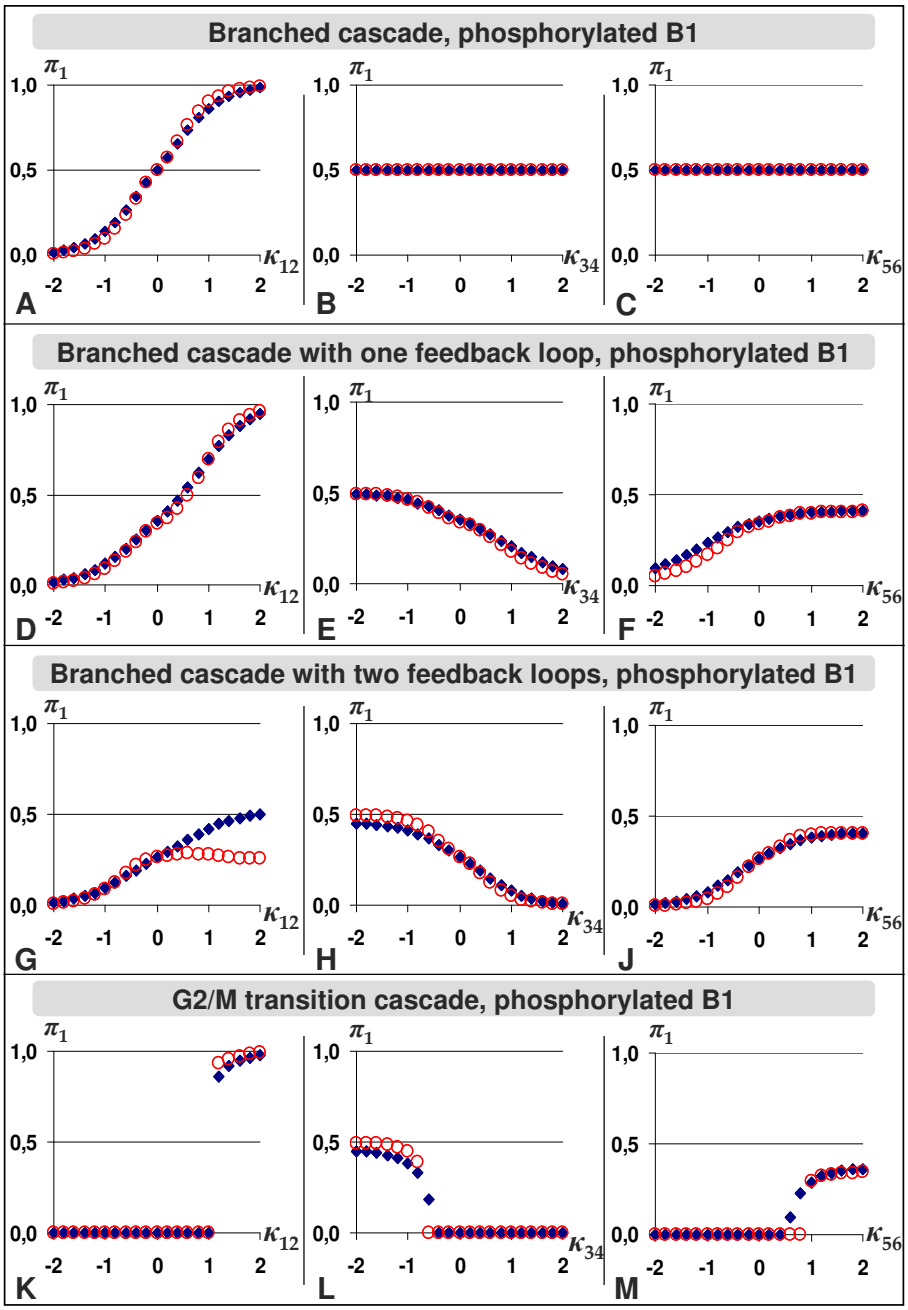


Figure 4.8: Bifurcation diagrams for different kinetic parameters of branched cascade variations (Sec. 4.1.3) on the level of phosphorylated protein $B1$. For definition of the parameters κ_n and the response π_1, see Sec. 4.2. Blue diamonds - steady state π_1 levels calculated with EMA, red circles - steady state π_1 levels calculated with AMA.

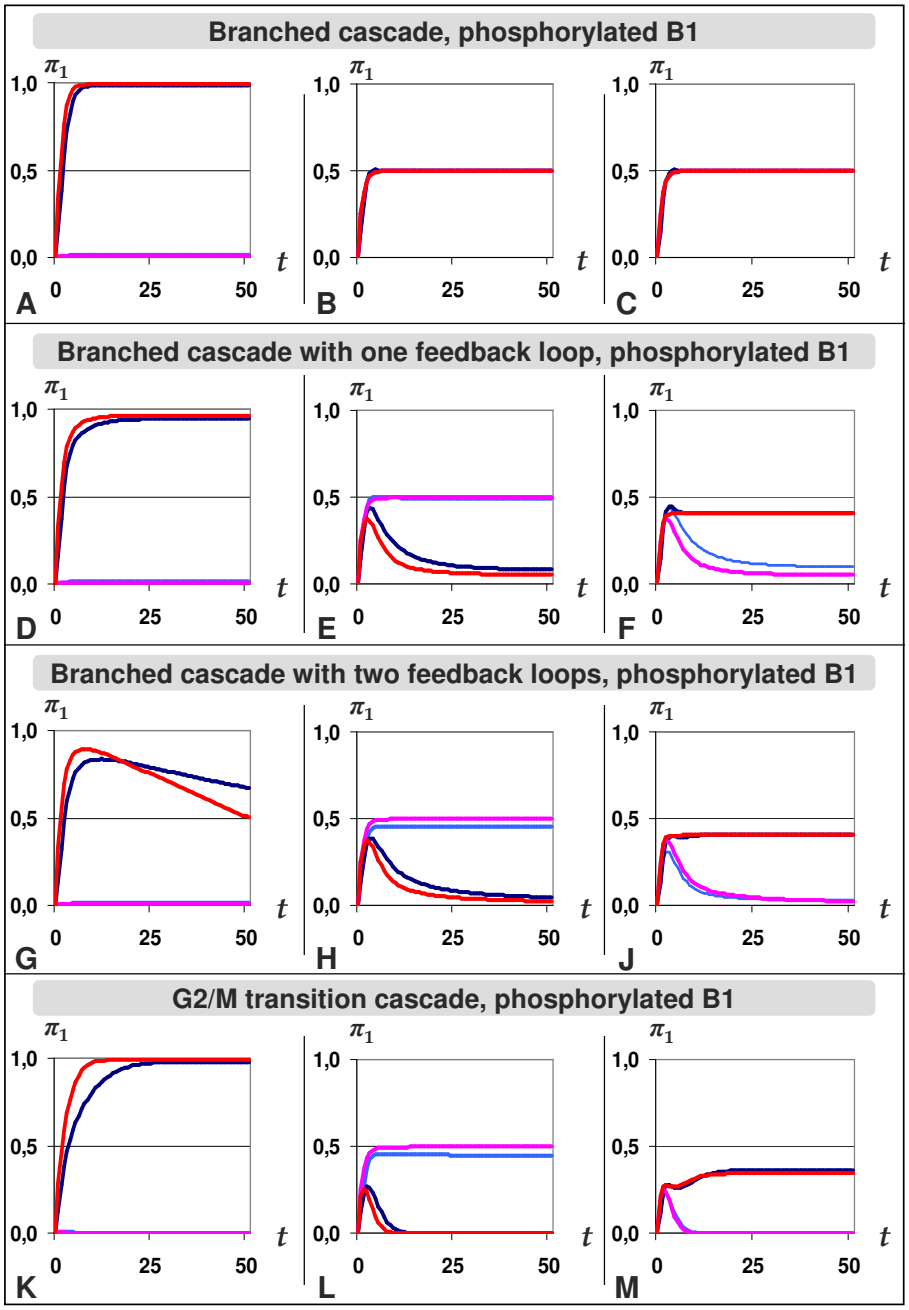


Figure 4.9: Time courses of phosphorylated protein $B1$ in the branched cascade variations (Sec. 4.1.3). Tested kinetic parameters in panels **A** - **M** correspond to the parameters depicted in Fig. 4.8 A - M, respectively. For definition of the parameters and the response π_1, see Sec. 4.2. Time courses are color-coded depending on parameter value p and response type π^A vs. π^E: **magenta** - π_1^A, $p=-2$, **red** - π_1^A, $p=2$, **light blue** - π_1^E, $p=-2$, **dark blue** - π_1^E, $p=2$.

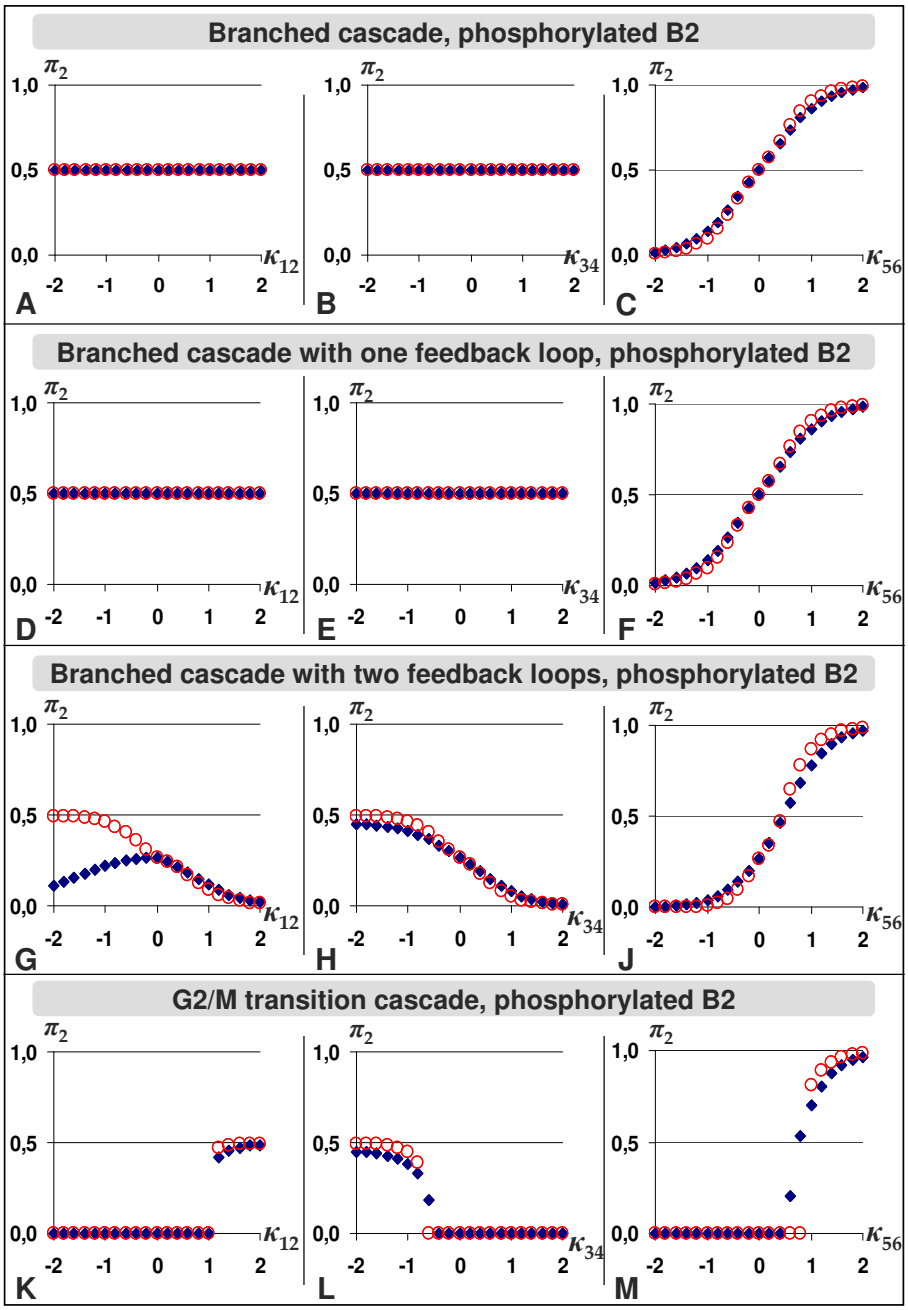


Figure 4.10: Bifurcation diagrams for different kinetic parameters of branched cascade variations (Sec. 4.1.3) on the level of phosphorylated protein $B2$. For definition of the parameters κ_n and the response π_2, see Sec. 4.2. Blue diamonds - steady state π_2 levels calculated with EMA, red circles - steady state π_2 levels calculated with AMA.

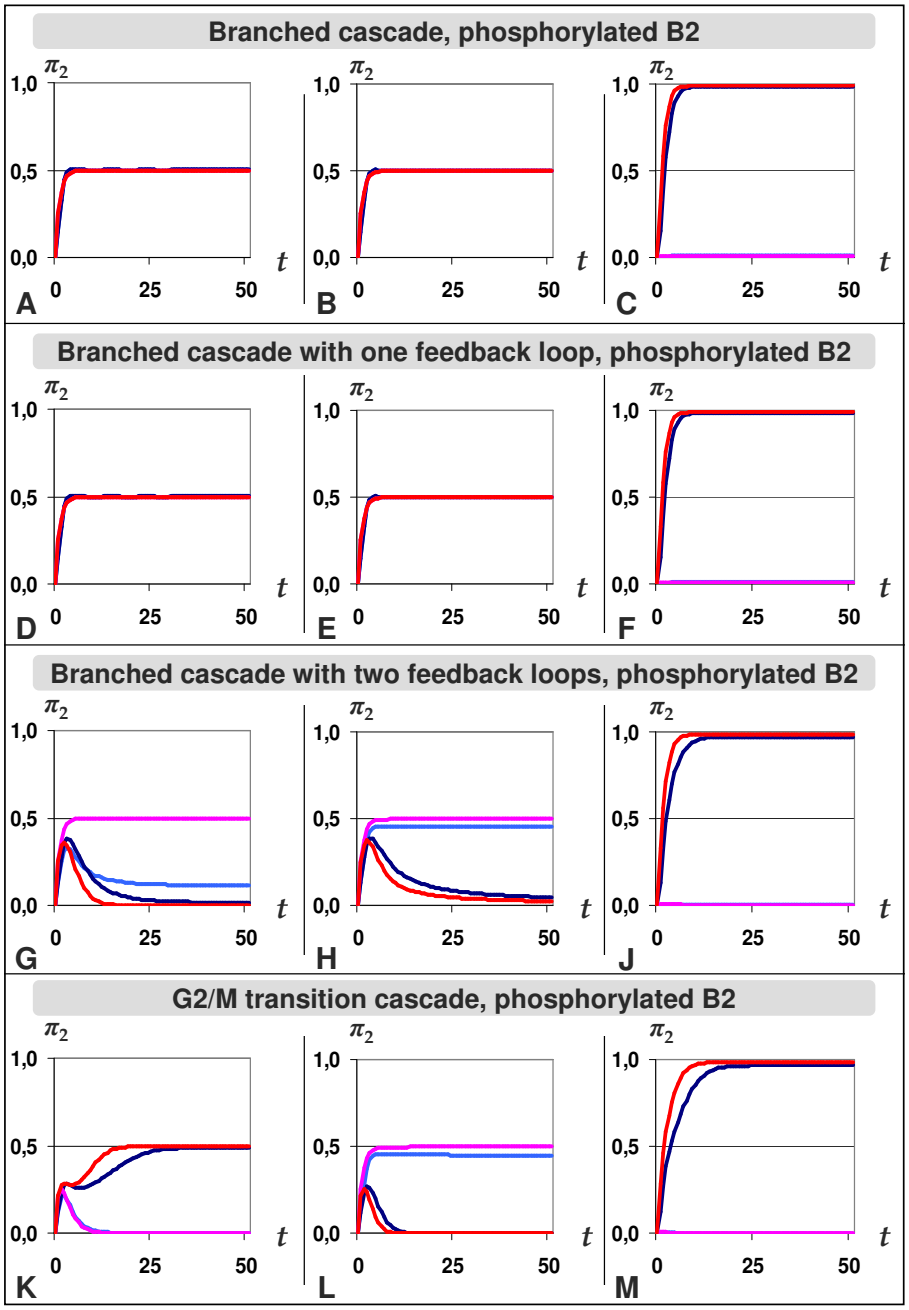


Figure 4.11: Time courses of phosphorylated protein $B2$ in the branched cascade variations (Sec. 4.1.3). Tested kinetic parameters in panels **A** - **M** correspond to the parameters depicted in Fig. 4.10 A - M, respectively. For definition of the parameters and the response π_2, see Sec. 4.2. Time courses are color-coded depending on parameter value p and response type π^A vs. π^E: **magenta** - π_2^A, $p=-2$, **red** - π_2^A, $p=2$, **light blue** - π_2^E, $p=-2$, **dark blue** - π_2^E, $p=2$.

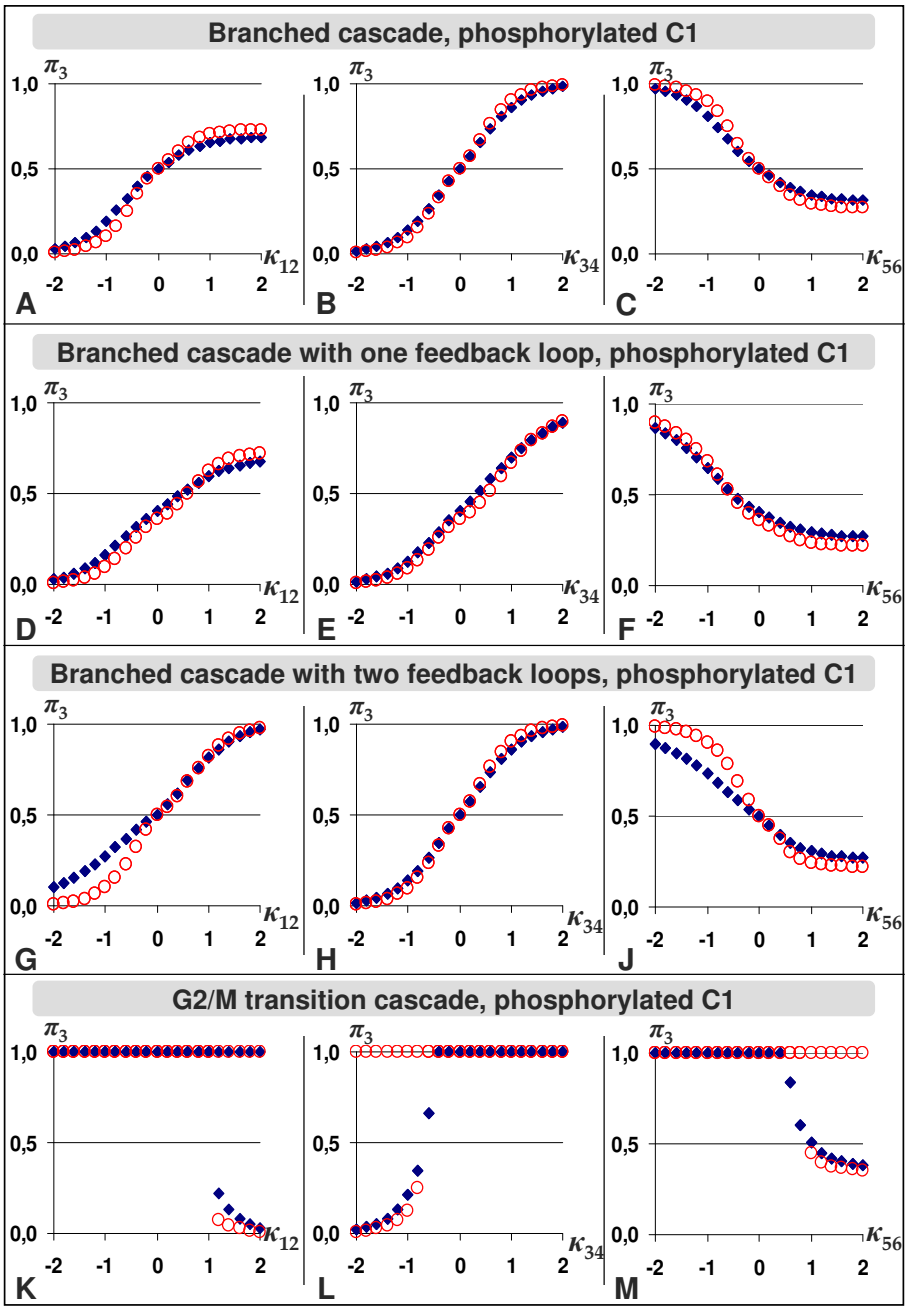


Figure 4.12: Bifurcation diagrams for different kinetic parameters of branched cascade variations (Sec. 4.1.3) on the level of phosphorylated protein $C1$. For definition of the parameters κ_n and the response π_3, see Sec. 4.2. Blue diamonds - steady state π_3 levels calculated with EMA, red circles - steady state π_3 levels calculated with AMA.

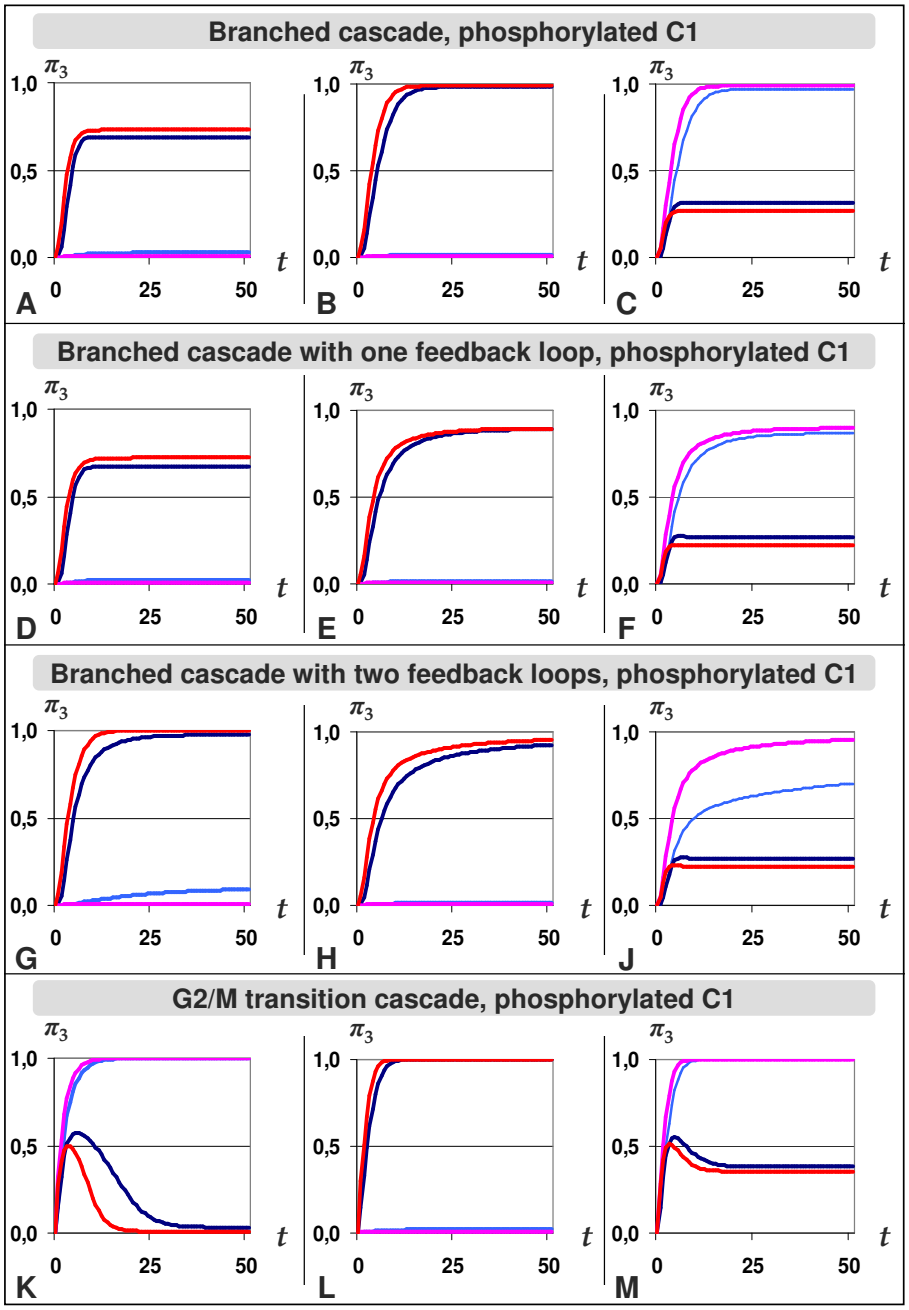


Figure 4.13: Time courses of phosphorylated protein $C1$ in the branched cascade variations (Sec. 4.1.3). Tested kinetic parameters in panels **A** - **M** correspond to the parameters depicted in Fig. 4.12 A - M, respectively. For definition of the parameters and the response π_3, see Sec. 4.2. Time courses are color-coded depending on parameter value p and response type π^A vs. π^E: **magenta** - π_3^A, $p=-2$, **red** - π_3^A, $p=2$, **light blue** - π_3^E, $p=-2$, **dark blue** - π_3^E, $p=2$.

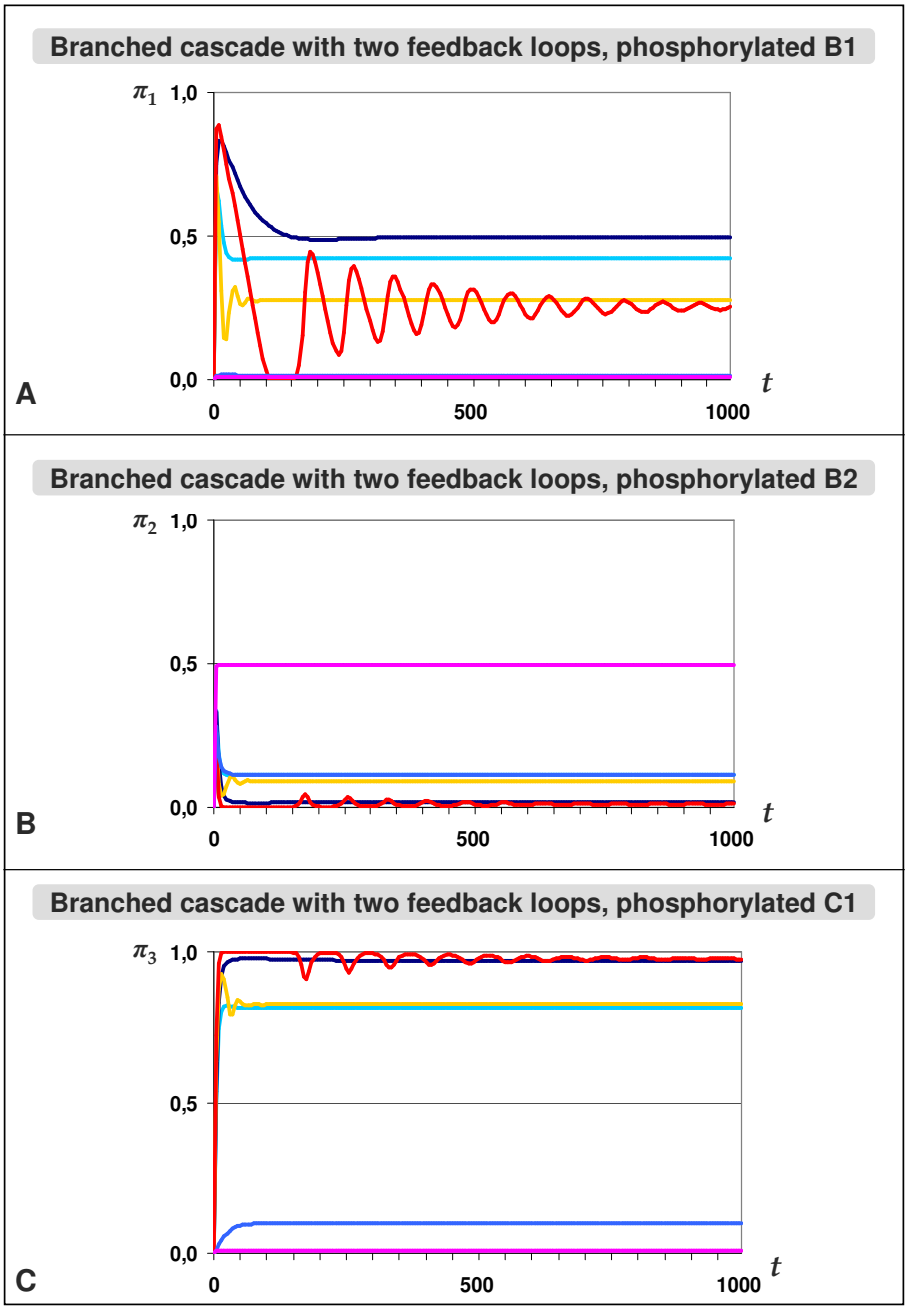


Figure 4.14: Time courses of phosphorylated proteins $B1$ (**A**), $B2$ (**B**) and $C1$ (**C**) in the branched cascade with two feedback loops for selected values of the κ_{12}. For definition of the parameter κ_{12} and the response π_1, π_2 and π_3, see Sec. 4.2. Time courses are color-coded depending on κ_{12} value and response type π^A vs. π^E: **magenta** - π^A, $log\kappa_{12} = -2$, **yellow** - π^A, $log\kappa_{12} = 1$, **red** - π^A, $log\kappa_{12} = 2$, **light blue** - π^E, $log\kappa_{12} = -2$, **cyan** - π^E, $log\kappa_{12} = 1$, **dark blue** - π^E, $log\kappa_{12} = 2$.

Chapter 5

Discussion

This work presents a combinatorial, EMA-based formalism for MIN description together with a formalized visual representation of such networks (Chpt. 2), a software implementation of this formalism, called aceSim (Chpt. 3) and simulation results of various enzymatic cascades conducted with aceSim (Chpt. 4).

In this chapter, we discuss related work in the domain of graphical and mathematical representation of MIN (Sec. 5.1), existing MIN simulation software packages (Sec. 5.2) and behavior of enzymatic cascades simulated with various approaches (Sec. 5.3).

5.1 Formal description

In this section we discuss various approaches to graphical and mathematical representation of MIN in context of the formalism introduced in Chpt. 2. We discuss the advantages of structure-oriented formalisms over the process-oriented approaches in graphical representing of combinatorial complexity. We further discuss EMA, AMA-based and other mathematical approaches to the ODE description of MIN.

5.1.1 Graphical description of MIN

Graphical representation of various kind of systems, such as electric circuits, buildings, social organizations etc., can focus on several aspects, including: structure (system's components and relations between them), information

flow through the system (e.g. signal processing algorithms), processes (sequence of events taking place while the system works) etc. [124].

In the domain of MIN, the numerous existing approaches to graphical representation can be divided into two main groups according to their descriptive focus:

- **structure-oriented** - these approaches depict simultaneously all possible interactions between molecular species. The most established example are Molecular Interaction Maps (MIM) [131], [133]. Other examples include the extended MIM (xMIM) [124], [132] and the universal visual language for systems biology (BioD) [61]. Formalism presented in this work (aceSim) is to a large extent a combination of these approaches and falls also in the structure-oriented category.

- **process-oriented** - these approaches depict sequentially a subset of interactions in the system that corresponds to a specific biological process, e.g. a signaling pathway. The most established example is the Process Diagram (PD) [124], [125], [126] with many related forms commonly used for depicting signaling and metabolic pathways. Another example is the rule-based signal transduction modeling language (BioNetGen) [38], [73], [74].

As presented in Fig. 5.1, the aceSim formalism can be compared with other approaches according to following criteria: managing combinatorial complexity, temporal sequence and information flow, descriptive scope, availability of mathematical translation and computer implementation.

Combinatorial complexity

Typically, combinatorial complexity is handled by explicit assignment of network nodes to any possible combinatorial sub-species [40]. The sub-species not allowed by regulation are simply not depicted. However, this plethora of resulting network nodes can be reduced in the structure-oriented approaches as described below.

The MIM formalism pictures always only one labeled instance of the species and reduces all combinatorial sub-species to nodes in form of singular dots. These dots can be connected with each other to depict higher-level complexes. However, The same dot can be duplicated along an interaction line which may be confusing and it is not always easy to track higher-level complexes back to their basic parts.

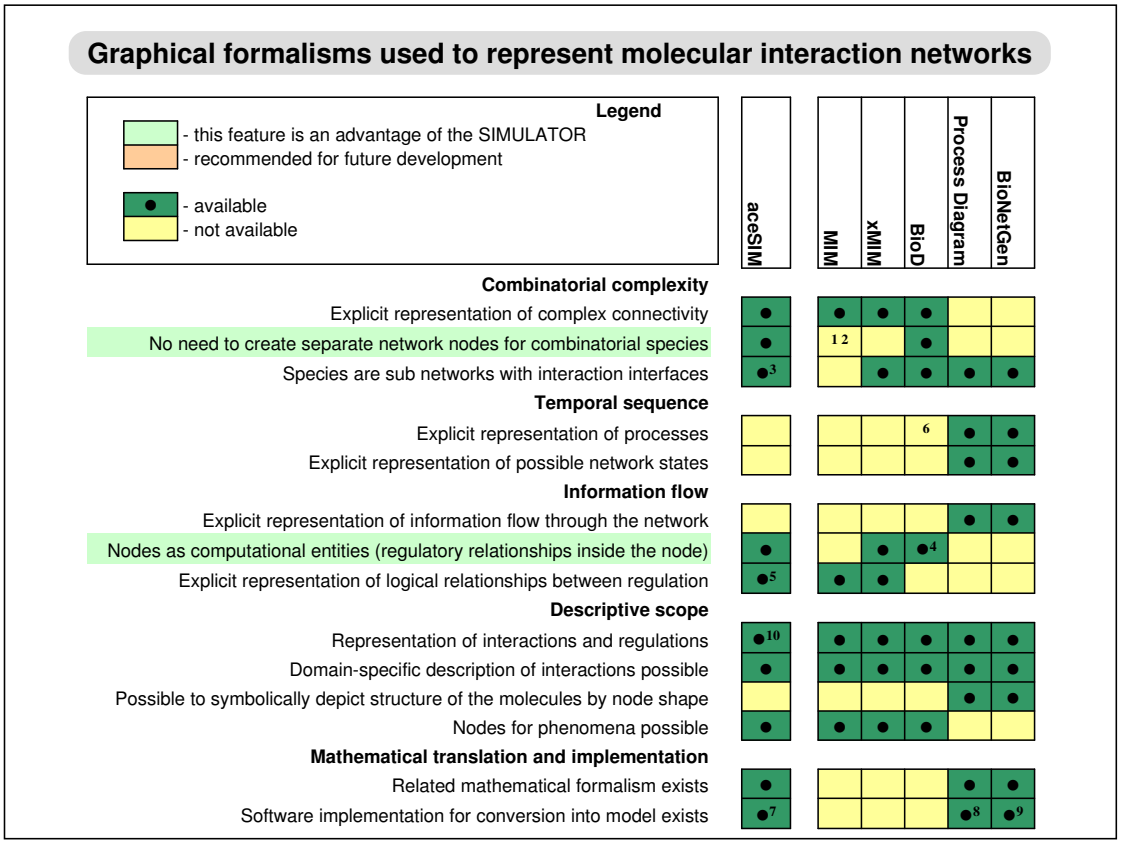


Figure 5.1: Graphical formalisms for representation of MIN. Comments: **1** - Only basic node labeled, all combinatorial nodes are generic nodes depicted as a dot. These dots can be connected with each other to depict higher-level complexes. Isolated nodes for homodimers. **2** - The same node can be duplicated along interaction line which may be confusing. **3** - Interfaces are abstract and do not correspond literally to physical sites of the molecule. **4** - Both inside and outside possible. **5** - AND and NOR via 'exclusion' and 'necessity', respectively. **6** - Authors postulate an extension to depict event sequences derived from the network structure. **7** - 'aceSim' (Java). **8** - 'CellDesigner' (Java). Most simulation software packages use this or similar graphic formalism. **9** - BioNetGen' (Perl). **10** - Without transport

.

BioD offers a visually simpler approach by expanding species nodes to a second level networks consisting of interaction interfaces. The combinatorial sub-species can be read out from such a diagram only implicitly, by tracking interaction links between species and the eventual regulatory influences. However, we believe such description has a sufficient level of detail, since the functional roles of multimolecular complexes can be tracked back to functionalities of single species modulated by regulation from interacting partners in the complex (see Sec. 2.1 for details).

The two-level network approach of BioD was adapted in other presented formalisms, also the process-oriented ones, to express combinatorial complexity. Especially, xMIM and aceSim employ the interface concept in the same visually simplistic way as originally BioD.

The xMIM uses interfaces for a more concise representation of regulation, placing it inside the basic species node, but still displays dot-nodes for combinatorial sub-species. Actually BioD does not exclude an option for extra nodes for combinatorial sub-species. As described in Sec. 2.1, aceSim consequently restricts the network representation to the two-level network approach without extra nodes for combinatorial sub-species, also for the description of gene regulation (Sec. 2.4.2) and phenomena (Sec. 2.5).

aceSim is unique in defining interfaces as abstract variables not literally corresponding to physical sites of the molecule, which is necessary to represent interface states with Boolean values. This is in contrast to other formalisms, where the interfaces are related to physical interaction sites and are allowed to have multiple states. We believe the Boolean notation facilitates to a large extent automation of mathematical description.

aceSim uses also a slightly different set of interface types compared to BioD and xMIM. These two formalisms introduce a special interface type 'active site' for enzymes. aceSim reduces this case to a 'binding site' interface, since we argue that an enzymatic modification reaction on the side of enzyme can be seen as a simple binding-unbinding reaction. Furthermore, on the substrate side, aceSim splits the 'modification site' into two interfaces (binding to enzyme and modification). This allows assigning to each interface the maximum of two reaction rates (binding and unbinding or modification and demodification) and Boolean description of the interface state.

Temporal sequence

The process-oriented formalisms by definition depict sequences of distinct network states corresponding to subsequent events in a given process.

Structure-oriented formalisms, also by definition, do not have such capability and would have to be extended with special parsing algorithms to derive possible process pathways from the network structure. Importantly, it would be feasible to derive possible process pathways from a given network structure, as postulated in the case of BioD [61]. However, the reverse is not necessary true.

Information flow

Process-oriented formalisms allow also for explicit representation of information flows through the system (e.g. signal transduction). Structure-oriented approaches, on the other hand, depict the whole information-processing circuits, since species can be regarded as computational units wired with one another via the interaction and regulation links.

The information processing infrastructure can be conveniently visualized using formalisms combining the two-level network approach with depiction of regulations inside the node, which is the case in BioD, xMIM and aceSim. This explicitly shows species as computational units containing logical circuits.

The logical operations can be depicted either with logical gates between regulatory links (xMIM) or using Boolean regulatory categories such as 'necessity' and 'exclusion' (aceSim). MIM employs the same logical-gate approach as xMIM, however, it depicts regulations outside the node which we do not consider optimal for visual tracking of the information processing. BioD depicts regulations inside node, but it does not offer explicit logical relationships between them.

Descriptive scope

A very basic requirement for MIN description is the ability to represent various types of interactions (binding, enzymatic reactions, synthesis, degradation, transport etc.) and regulations (activation, inhibition) etc. This requirement is fulfilled by all discussed formalisms, with the exception that aceSim is incapable of representing transport. This could be achieved by adapting one of the existing interaction formats, e.g. enzymatic modification, but is not yet present in the software implementation.

The interface approach allows domain-specific description of interaction and was adapted by all formalisms. However, only the process-oriented ones are capable of depicting case-specific structural features of combinatorial complexes by symbolic shapes, e.g., imitating ligands, receptors, antibodies etc.

Another aspect of description is an explicit incorporation of nodes representing phenomena into the network. This is allowed by all discussed structure-oriented formalisms. aceSim allows additionally linking of phenomena with the rest of the network as described in Sec. 2.5.

Mathematical translation and computer implementation

Among the discussed formalisms, only BioNetGen and aceSim have a corresponding ODE-based mathematical formalism and a software implementation. BioNetGen is implemented with a program of the same name written in Perl. aceSim is implemented in $Java^{TM}$, however uses a relatively simple spreadsheet input mask that should be further developed. The details of both mathematical representation and software implementation of aceSim and BioNetGen will be discussed in detail in the Sec. 5.2.

Process Diagram is implemented with Java as software called CellDesigner with feature-rich GUI for drawing MIN diagrams. It is not related to any explicit mathematical formalisms but it allows model export in the SBML format for further simulation via the multi-software platform Systems Biology Workbench [182]. Moreover, process-oriented formalisms similar to the PD are in fact a standard for the most software packages for MIN simulation and are typically related to the AMA description of reaction systems as described in Sec. 5.2.

5.1.2 ODE-based description of MIN

As outlined in the Sec. 1.2, MIN can be described mathematically in various ways, including systems of coupled ODE describing the dynamics of species concentration. These equations can take various forms, as outlined below.

Elementary mass action (EMA)

This type of equations is commonly used to describe binding reactions and as we argue in the results section, it is also most appropriate for describing enzymatic reactions in a stepwise manner. Description of synthesis, degradation or transport at this level is rather rare, however one can imagine

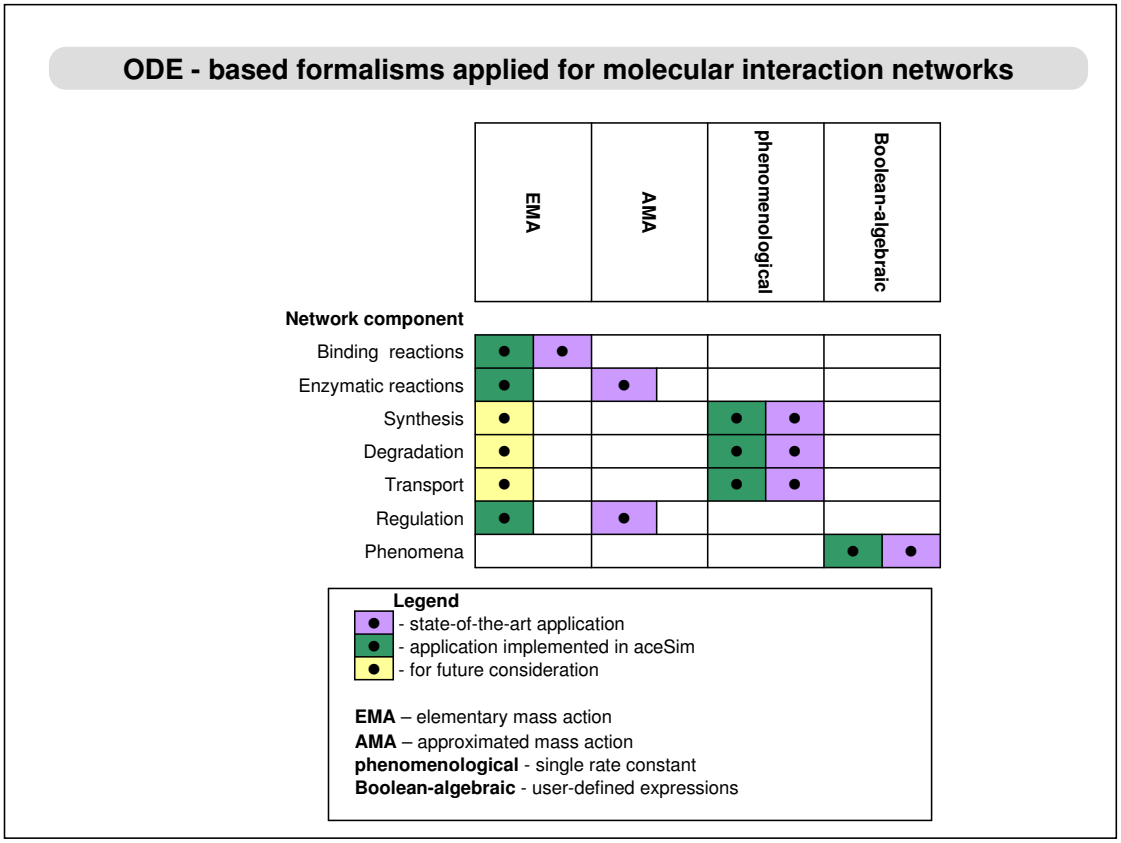


Figure 5.2: Mathematical formalisms for MIN representation. 'EMA' - elementary mass action (see Sec. 1.2.3 for definition), 'AMA' - approximated mass action (see Sec. 1.2.3 for definition), 'phenomenological' - a single rate constant approximation, 'Boolean-algebraic' - user-defined expressions with logical operators.

detailed kinetic rate laws for protein synthesis taking into account concentrations of nucleotides, aminoacids, polymerases or ribosomes etc. or for degradation considering proteasomes and ubiquitine or transport considering transporter proteins.

Approximated mass action (AMA)

These equations are commonly used for description of enzymatic reactions and their regulations, sometimes also for description of regulation of protein synthesis. We argue all these schemes should be replaced by EMA for more realistic calculations and for enabling modular composition of more complex reaction and regulation schemes.

Phenomenological reaction rate.

These are even more simplified forms containing all mechanisms lumped into one reaction rate. Such mechanism is common for describing transport, synthesis and degradation and was also implemented in aceSim.

Boolean-algebraic equations.

Algebraic functions are frequently used for including conservation relationships into the equation system. This application is irrelevant for aceSim since the conservation laws are automatically fulfilled when using EMA 2.2.1. However, algebraic equations combined with logical conditions are suitable for describing phenomena like thresholds, discrete events etc.

In general, we argue that EMA formulations are most suitable for describing binding and enzymatic reactions as well as regulation. We also consider Phenomenological rate constants to be a suitable alternative for synthesis, degradation and transport, however, replacing those with EMA formalism taking concentrations of RNA polymerases and ribosomes, proteasomes or transporter proteins into account should be also considered. Based on the results presented in chpt. 4 and discussed in Sec. 5.3, we discourage the use of AMA in any case, especially for enzymatic reactions and regulation, which is still very popular as discussed in Sec. 5.2.3.

5.2 MIN simulation software

In this section we discuss various features of existing software solutions for ODE-based MIN modeling. A comparison with the aceSim software

presented in Chpt. 3 reveals, that the unique character of aceSim is the ability to automatically construct EMA-based models of systems taking account of combinatorial complexity.

5.2.1 Existing MIN simulation software

There are several established simulation software solutions for systems biology (see [27], [33], [127], [160], [169] for reviews). Moreover, many of the tools are integrated into larger, multifunctional platforms, such as Systems Biology Workbench and BioSPICE [182].

Below we focus on 12 applications that primarily and explicitly deal with MIN simulation with using deterministic (ODE-based) algorithms and thus, as we believe, allow the most informative feature comparison with the software presented in this work. The short descriptions are based on published papers, technical documentation and internet presence as referenced below. Enumerations in text follow the alphabetical order.

- **BioNetGen** - Allows automatic generation of mathematical models of biological systems from user-specified rules for biomolecular interactions. Uses own language for explicit indication of the parts of proteins involved in an interaction, the conditions upon which an interaction depends, the connectivity of proteins in a complex, and other aspects of protein-protein interactions. Graphical front end for construction of BioNetGen rules using graphical icons exists. Version: 2.0, ref: [38], link: [9].

- **CADLIVE** - Computer-Aided Design of LIVing systEms, includes GUI editors, Simulator and Grid Layout Program. GUI editors enable to construct large-scale biochemical network maps. Simulator converts biochemical network maps into dynamic models and simulates their dynamics. Grid Layout Program places biochemical networks on 2-dimensional squared grid. Version: 2.14, ref: [136], link: [16].

- **Cellerator/kMech/Sigmoid** - Mathematica package for generating, translating, and numerically solving a potentially unlimited number of biochemical interactions on the level of signal transduction networks, single cells and multi-cellular tissues. Version: 1.0, ref: [191], [218] , link: [14].

- **Cellware** - Integrated environment not only for modeling and simulation of gene regulatory and metabolic pathways but also other di-

verse mathematical representations, parameter estimation and optimization. First grid based modeling and simulation tool in the field of Systems Biology. Version: 3.0, ref: [5], [65], link: [15].

- **COPASI / GEPASI** - COmplex PAthway SImulator for Simulation and analysis of biochemical networks. Spreadsheet GUI for model editing and broad analytical functions including: steady state stability, metabolic control, sensitivity, elementary mode, mass conservation analysis, calculation of Lyapunov exponents, Parameter scans and estimation, optimization of arbitrary objective functions. Version: 4.0, ref: [7], [154], link: [17].

- **Dizzy** - Model definition environment and kinetic simulation of chemical reaction systems with various stochastic (Gillespie, Gibson-Bruck, Tau-Leap) and deterministic algorithms. Version: 1.11.4, ref: [172], [173], link: [11].

- **Dynetica** - Biologist-oriented modeling tool for constructing, visualizing, and analyzing kinetic models of biological systems. Intuitive interface for easy model construction. Ver 1.2, ref: [219], link: [18].

- **E-cell** - Software platform for modeling, simulation and analysis of complex, heterogeneous and multi-scale systems like the cell. Capable of running various different algorithms simultaneously in a single simulation. Version: 3.0, ref: [202], [203], [204], link: [19].

- **Jarnac / Jdesigner** - JDesigner allows drawing a biochemical network and exporting it in the form of SBML or to Jarnac as a simulation server (via SBW). Jarnac is a language for describing and manipulating any physical system in terms of a network and associated flows, especially metabolic, signal transduction and gene networks. Version: 2.5, ref: [1], [180], link: [12], [13].

- **JigCell** - A set of tools for model creation, simulation and analysis. A spreadsheet interface allows definition of chemical species, equations, relationships and events that occur when a user-defined condition is met. Analysis tools include comparison of simulated and experimental data. Bifurcation analysis tools are under development. Version: 6.1.4, ref:[213], link: [10].

- **MATLAB Systems Biology Toolbox** - Open and user extensible environment, in which to explore ideas, prototype and share new al-

gorithms, and build applications for the analysis and simulation of biological systems. Version: 1.6, ref: [185], [186], link: [20].

- **VirtualCell** - Associates biochemical and electrophysiological data describing individual reactions with experimental microscopic image data describing their subcellular locations. Cell physiological events can then be simulated within the empirically derived geometries, thus facilitating the direct comparison of model predictions with experiment. Version: 4.2, ref: [2], [146], link: [4].

The detailed features of these applications along with the software presented in this work will be compared below with focus on the automation extent of the modeling process and the descriptive scope.

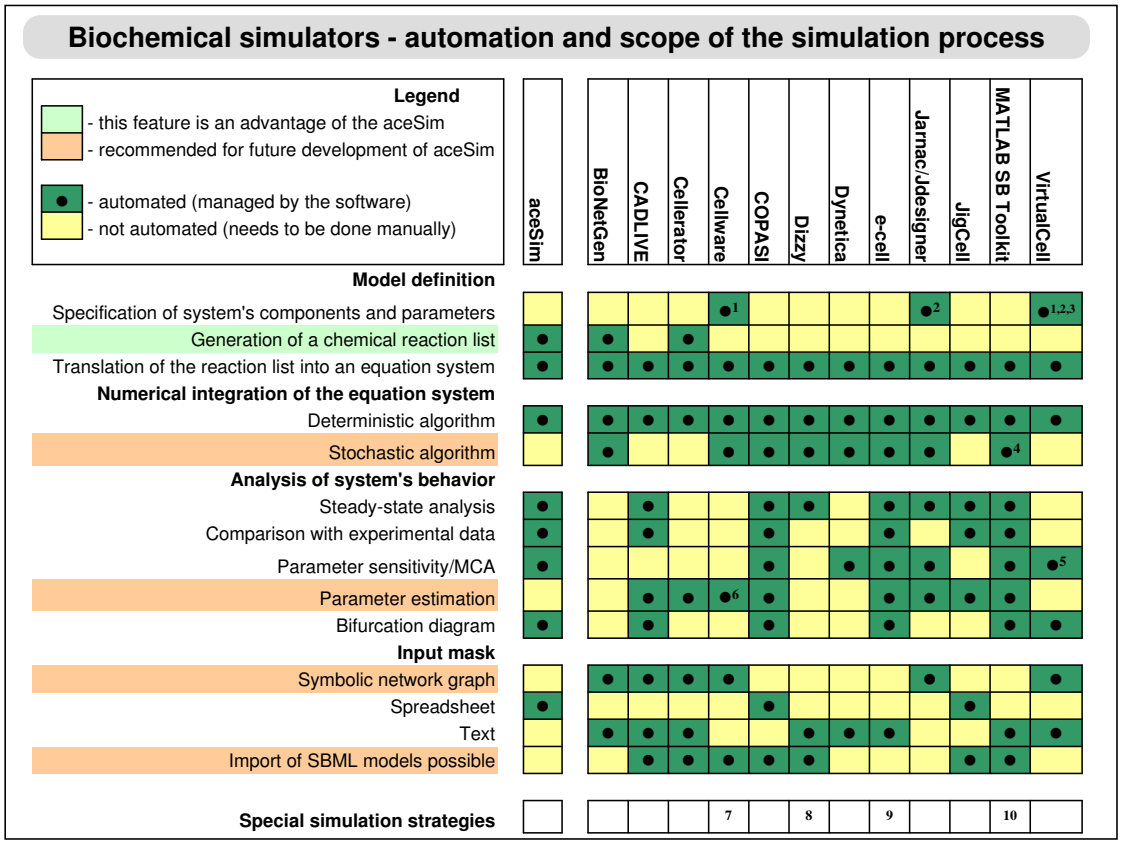


Figure 5.3: Simulation and analysis capabilities of existing software solutions to MIN simulation. 'MCA' - metabolic control analysis. Comments: **1** - Possible to derive data from KEGG database. **2** - Possible to derive data from SigPath database **3** - Possible to derive data from VirtualCell database. **4** - Available only as a stand-alone method. **5** - Sensitivity only estimated from the nominal value without rerunning simulation. **6** - Methods for estimating elementary rates from experimental KM and Kcat. **7** - Simulation using grid computing. **8** - Possibility to define reusable model segments ('templates'). **9** - Multi-algorithmic simulation possible (e.g. deterministic and stochastic). **10** - Localization of functions leading to complex behaviors.

5.2.2 Automation and scope of the simulation process

The process of MIN modeling can be divided into 3 generic phases: 1) definition of the model, 2) numerical simulation and 3) analysis of the results. At each stage, different features are offered by different software solutions to facilitate the modeling process, as summarized in Fig. 5.3 and discussed below in a more detail.

Model definition

This first stage of simulation process can be further divided into following steps:

1. Specification of system's components and parameters.
2. Generation of a chemical reaction list.
3. Translation of the reaction list into an equation system.

Most of the available software packages offer only automation of the last step, so that the user needs to specify an explicit reaction list himself. This means that the combinatorial complexity is not really being managed by the software, since the user needs to enumerate all possible combinatorial sub-species and reactions between them and control manually eventual modifications of rate constants due to regulation.

Only aceSim, BioNetGen and Cellerator are able to create a combinatorial reaction list automatically from a limited, user-defined set of rules specifying system's components (species) and interactions between them in a generalized way (as described for aceSim in Chpt. 3). Furthermore, Cellware, Jarnac/Jdesigner and Virtual Cell enable automatic derivation of system specifications from external biochemical databases, such as KEGG [114] or SighPath [53], however, the user still has to specify resulting reaction lists himself. It would be highly desired to combine the above feed-from-database feature with automated reaction list composition in order to fully automate the process of model definition. Such combination of features is to our knowledge not available so far.

In respect to automation of the model definition, we consider aceSim to be unique in the extent that combinatorial complexity can be automatically managed thanks to the Boolean representation of combinatorial species and extrapolation of reaction rates and regulatory coefficients as described in Sec. 2.1. This allows an even more limited set of initial rules necessary to generate the reaction list compared to the BioNetGen approach, where

interfaces can have multiple values and thus some combinatorial cases, especially of regulation, need to still be handled individually.

Moreover, as discussed in detail in Sec.5.2.3, both BioNetGen and Cellerator translate the reaction lists into equations partially using mathematical forms that we discourage (AMA and single phenomenological rates for enzymatic reactions and regulations).

Numerical integration of the equation system

After the model has been defined as an ODE system, it can be numerically integrated to obtain the time-courses of species concentration changes in the modeled system. All discussed software packages offer some standard algorithms for numerical integration including Euler's and Runge-Kutta methods [66] (see Sec. 3.5.3 for an outline of the Runge-Kutta algorithm).

Additionally, most of the packages offer stochastic algorithms such as the Gillespie or Tau-leap method [89], [90] . This simulation procedure more suitable for systems, where small molecular quantities occur, such as DNA strands (esp. genes) or small cellular compartments (e.g. synaptic regions) and thus the mass action approximation necessary for applying ODE is no longer valid. Thus, extending aceSim with stochastic simulation capabilities would be highly desirable.

Analysis of system's behavior

As outlined in the Sec. 3.5, the system's behavior can be analyzed in many ways, including: determination of the number and stability of steady states, comparison of the concentration time-courses with experimental data, testing system's behavior for various values of parameters, such as rate constants or initial concentration values, estimation of parameter values to fit experimental data, plotting bifurcation diagrams etc.

aceSim, CADLIVE, COPASI, e-cell, Jarnac/Jdesigner and MATLAB SB Toolkit offer most or all of the mentioned analytic features, whereas other packages have rather limited capabilities at this point. aceSim does not support parameter estimation, this feature should be considered for future development.

Special simulation strategies

Some of the packages offer sophisticated special features, including:

- Possibility to define reusable model segments called 'templates' (Dizzy).

- Simulation using grid computing (Cellware)
- Partitioning the model into separate simulation units, that can be run simultaneously using different algorithms or parameters, e.g. time step size (e-cell).
- Methods for estimating elementary rates from experimental Michaelis constant and catalytic constant (Cellware).
- Methods for localization of functions leading to complex behaviors within the network structure, e.g. feedback loops (MATLAB SB Toolkit).

Input mask and model importing

Apart for the ability of the software to automate the modeling process, the model definition can be also facilitated by a graphical user interface (GUI). Such interface can take various forms [3], including:

- **Symbolic network graph** - this interface type allows composing a graphical representation of the system using pre-defined icons and drag-and-drop operations. As outlined in Sec. 5.1.1, the graphical representation can have a process-oriented or a structure-oriented form. The process-oriented approach corresponds to a reaction list, whereas a structure-oriented approach allows defining a more limited set of components and rules and thus a large extent of automation of the model definition. BioNetGen, CADLIVE, Cellerator, Cellware, Jarnac/Jdesigner and VirtualCell implement a network-graph-based GUI, all using the process-oriented graphical formalism.
- **Spreadsheet** - this interface type allows specification of both reaction lists and components/rules lists in a tabular form. Spreadsheet interfaces offers a compact system's representation that for larger systems might be easier to track comparing to the network graph representations. COPASI and JigCell employ spreadsheet interfaces for entering reaction lists.
- **Text** - this is the most basic interface type and entering model specifications in this form requires learning some specific syntax. Some packages offer this input type as an alternative to graphical interface. Dizzy, Dynetica, e-cell and MATLAB SB Toolkit offer a text interface only.

aceSim has an worksheet input mask which is however still in development and should be ideally replaced/complemented with a network graph-based interface, necessarily using a structure-oriented graphic formalism in order to support management of combinatorial complexity.

The model definition phase of the simulation process can be also completely omitted by using ready-made models from a model repository such as the BioModels Database [140]. This requires capability of importing models in some standardized, structured format. The most established standard for coding biochemical models is the Systems Biology Markup Language level 2.0 (SBML) [83], [109].

Most of the discussed software packages support importing of models coded in SBML. aceSim does not offer SBML support so far, since this format does not yet support the components/rules logic for management of combinatorial complexity. However, such extension has been proposed for the upcoming SBML level 3.0 [82], [39].

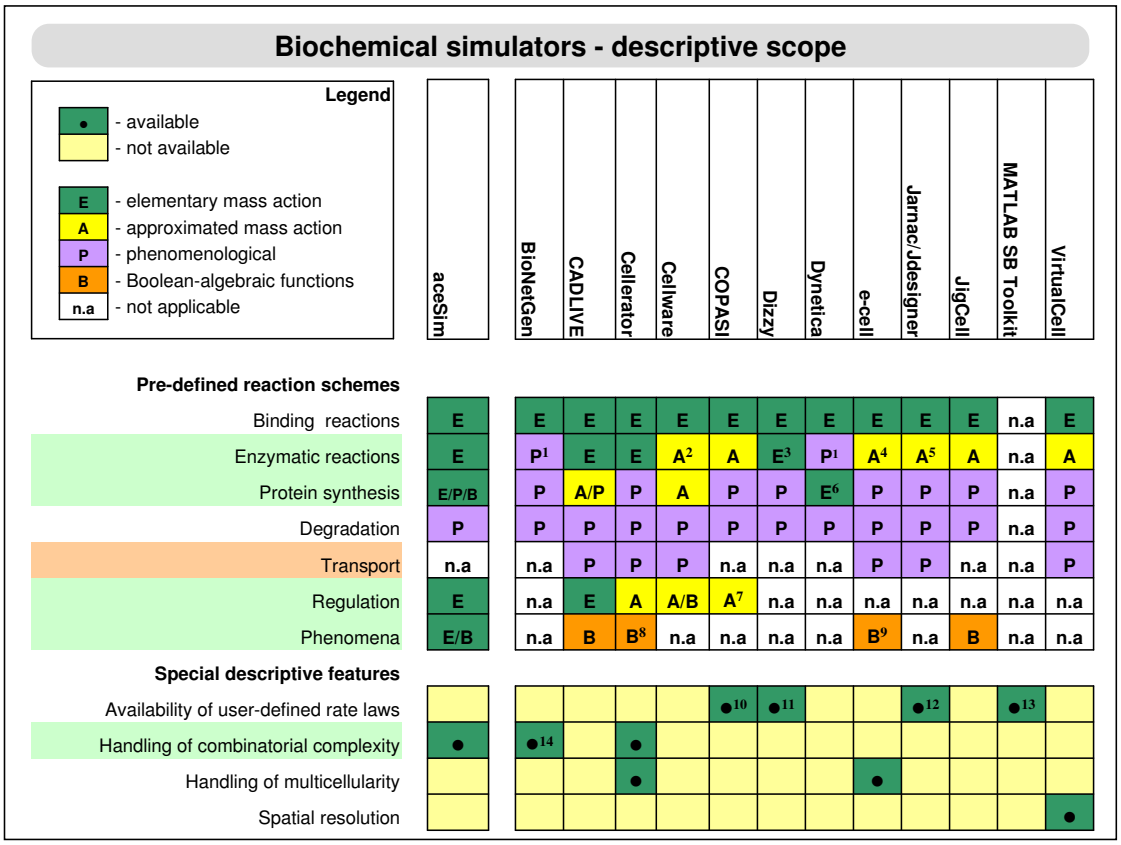


Figure 5.4: Descriptive scope of existing software solutions to MIN simulation. Comments: **1** - Enzyme concentrations parameterized as single flux rate. **2** - Possible to use EMA but two separate reactions per one enzymatic modification have to be specified manually (substrate binding and transformation). **3** - Multistep enzymatic reactions can be approximated with an integro-differential equation. **4** - Process class library with Michaellis-related enzymatic mechanisms and simple fluxes. **5** - Enzyme concentration parameterized as Vmax. **6** - Synthesis takes into account concentration of nucleotides, aminoacids, polymerases, ribosomes etc. **7** - Extensive library of approximations for enzymatic reactions, n.a. for regulation of other interactions. **8** - Thresholds for enzymatic activities using Heaviside function. **9** - Discrete events implemented with Python commands. **10** - User-defined rate laws possible for any reaction with various operators and functions. **11** - Only EMA or user-defined possible. **12** - External XML file for user-defined rate laws. **13** - All rates user-defined in string representation or directly as ODE. **14** - The validity of assigning the same rate constant(s) to a set of reactions is the responsibility of the modeler. Case specific rate modifications due to regulation need to be done manually.

5.2.3 Descriptive scope

Almost all analyzed software packages offer predefined reaction schemes for the basic reactions such as binding, enzymatic reactions, degradation and synthesis. Fewer packages support also transport, regulation and description of phenomena. An exception is MATLAB SB Tool, where only user-defined rate laws or other equations can be entered. Copasi, Dizzy and Jarnac/Jdesigner allow also user-defined rate laws together with pre-defined reaction schemes. The descriptive capabilities of different software solutions are summarized in Fig. 5.4 and discussed below in a more detail.

Binding and enzymatic reactions

Binding reactions are consistently implemented using EMA. However, enzymatic reactions are implemented using AMA schemes or even single phenomenological rate schemes (BioNetGen, Dynetica). For the reasons outlined in the Sec. 5.1.2, we argue that enzymatic reactions and their regulation should be described using EMA schemes as presented in Chpt. 2. This requirement is fulfilled for only by aceSim, CADLIVE, Cellerator and Dizzy. However, Cellerator uses AMA forms for enzymatic regulation and Dizzy does not support enzymatic regulation at all. CADLIVE supports regulation with EMA schemes, but since it uses the process-oriented formalism, where regulated rate constants have to be specified individually for every possible reaction, it does not really support combinatorial complexity. Thus we believe that the combination elementary mass action description of regulation with the capability to automate combinatorial complexity is an unique and perhaps the most outstanding advantage of aceSim.

Synthesis, degradation and transport

All these interaction types are most often described in a phenomenological way using single rate schemes. For transcriptional regulation, CADLIVE and Cellware offer also AMA schemes based on the Hill equation and Dynetica implements an even more detailed scheme taking account of the components of protein synthesis machinery, such as polymerases or ribosomes.

aceSim describes protein synthesis using the single phenomenological rate scheme and it offers two further approaches for transcriptional regulation. The first is based on EMA description of the interactions between a given gene A and other molecular species that can alter the phenomenological synthesis rate of protein A (Sec. 2.4.2). The second allows linking

the synthesis rate to values of phenomenological objects that can be derived from concentrations of user-defined ensembles of combinatorial species (Sec. 2.5.3). Both approaches are implemented using the same basic set of rules as for other interaction types and thus can also be processed in a fully automated way.

Such detailed schemes as the ones employed by aceSim or Dynetica for protein synthesis could also be considered for describing degradation and transport. Especially, aceSim should in any form support transport which is not the case yet.

Regulation

Pre-defined regulation schemes almost exclusively rely on AMA formalism, as it is the case for Cellerator, Cellware and COPASI. Especially COPASI offers an extensive library of such regulation schemes for enzymatic reactions but not for other interaction types. Only aceSim and CADLIVE implement regulatory schemes with EMA, with the exclusive advantage of aceSim of being fully automated as explained in Sec. 5.1.1 (regulated complexes and rate constants do not need to be specified manually).

Phenomena

Only aceSim, CADLIVE, e-cell and JigCell offer explicit support for phenomenological variables, such as discrete events, that are implemented using user-defined, Boolean-algebraic functions. Cellerator offers only limited support for phenomena allowing thresholds for enzymatic activities using Heaviside function.

aceSim supports phenomena as both real-valued and Boolean variables and allows multiple ways of linking them with the molecular species and with each other in a formalized and thus possible to automate way (Sec. 2.5).

Special descriptive features

Some of the software packages have additional specific descriptive features, such as handling of multicellularity (Cellerator and e-cell), simulation with spatial resolution (VirtualCell) and, last but not least, management of combinatorial complexity (aceSim, BioNetGen and Cellerator).

5.3 Simulation results

In this section, we discuss the behavior of various enzymatic cascades simulated using aceSim as presented in Chpt. 4. This analysis demonstrates, that application of our combinatorial and EMA-based formalism brings new insights into the behavior of enzymatic cascades, in some cases contrary to the results of the AMA-based modeling. Moreover, this investigation is meant to demonstrate analytical capabilities of aceSim.

5.3.1 Behavior of different systems

A biological system can be described in terms of signal-response analysis, where any of system's parameters/variables or combination thereof can be treated as a signal or a response [183], [208]. We have investigated the behavior of a basic kinase-phosphatase system and various cascades composed of this motif, defining the response as the fractional concentration of the phosphorylated substrate and various types of signal as relative and absolute kinetic rate constants of this system (Sec. 4.2). Altering of these rate constants can e.g. result from enzymatic regulation of interacting proteins [183].

The elementary response pattern, i.e. the change in the numerical value of the response relative to the change in the value of the signal, depends on the **sensitivity** of the system. A sigmoidal response pattern is in fact common in nature; it actually resembles the response of collector current on the base current in a transistor [183].

In extreme cases, the response changes from values close to 0 to values close to 1 (or reverse) within a very small range of signal values. This phenomenon has been described as **ultrasensitivity** [58], [94]. As opposite to ultrasensitivity, systems with a flat signal-response curve are called **subsensitive** [138].

Ultrasensitivity has been observed experimentally in regulation of several enzymes, such as Mitogen-Activated Protein Kinase (MAPK) in maturating *Xenopus* oocytes [77], isocitrate dehydrogenase [137], glycogen phosphorylase [152], glutamine synthetase cascade in *E. coli* [159], AMPactivated protein kinase (AMPK) [100]. The biological functions of ultrasensitivity include binarization of response by means of a threshold [79], signal amplification [80] or filtering [93]. There are several sources of ultrasensitivity in biological systems, including multisite phosphorylation [95], feedback loops [78] or enzymes operating under saturation [94]. The last case was

termed **zero-order sensitivity**, since the enzyme saturated with substrate operates in zero-order regime [94]. However, in many cases the enzymes included in MIN are at concentrations of the same order of magnitude as their substrates and thus operate in the first-order regime, which reduces ultrasensitivity to regular sigmoidal sensitivity and even subsensitivity [41]. Similar effect was observed when the enzymes are assumed to be product-sensitive [163], when the phosphorylation and dephosphorylation are catalyzed by the same ambiguous enzyme [163] or for systems with low numbers of molecules simulated using stochastic approach [32]. However, our results demonstrate, that **first-order ultrasensitivity** is also possible in branched cascades containing a specific configuration of feedback loops, as will be discussed below.

Behavior of the kinase-phosphatase system

The kinase-phosphatase system has the typical sigmoidal sensitivity pattern for the relative enzyme substrate binding rate k_1/k_2 and for the relative transformation rate k_{11}/k_{22} and that it is much less sensitive to the enzyme-substrate complex dissociation rate k_{-1}/k_{-2}.

This finding suggests, that assuming the rates $k_{-1} = k_{-2} = 0$ would not have a major impact on the simulated behavior of systems based on the kinase-phosphatase motif. This has a major practical consequence, since it reduces the number of parameter required to describe such a system and allows deriving the rates k_n and k_{nn} from the values of $KM = (k_{-n} + k_{nn})/k_n$ and $V_{max} = E_T k_{nn}$ that can often be obtained experimentally (E_T is the total concentration of an enzyme catalyzing the given reaction during measurement) [62]. This also motivates our assumption about linking regulatory relationships to the association rates k_n instead of the dissociation rates k_{-n}.

Behavior of linear and branched cascades

The sensitivity on consecutive levels of a cascade has been shown to increase, leading eventually to ultrasensitivity, if concentration gradients between these levels exist, allowing them to operate in zero-order regions [80], [108]. The reverse has been shown for cascades where the concentrations are of the same order of magnitude [41]. We have shown a similar drop in sensitivity downstream a linear cascade described with EMA, together with an extended relaxation time. However, we have also shown, that response damping in a linear cascade can be reduced in a branched cascade, where

both the kinase and the phosphatase determining the response are regulated, though this involves even longer relaxation times.

Introduction of feedback can also increase sensitivity of enzymatic cascades [184]. As we have shown for the branched cascade, the phosphorylation level of $B1$, $B2$ and $C1$ is sensitive only to the parameters of reactions in which the given protein is directly involved; obviously, introduction of feedback invokes in the insensitive proteins a dumped sigmoidal response pattern to for the parameters indirectly related via feedback. Interestingly, in a couple of proteins involved in feedback each has a specific response pattern superimposed from two sigmoids.

Furthermore, negative feedback can bring about oscillations [93], [117] [209] and a combination of positive and negative feedback can lead to bistability and hysteresis [29], [36]. In a first-order cascade described with EMA, after introducing negative feedback and a negative-positive feedback, we have only observed sharpened sensitivity patterns , but still not ultrasensitivity or bistability.

Strikingly, oscillations were only observed in the branched cascade with a positive and a negative feedback loop described with AMA. These oscillations were damped, but for some parameter ranges, they sustained for a longer time (Fig. 4.12 L - M).

Behavior of the G2/M transition cascade

The G2M transition cascade has a specific combination of a positive and a double-negative feedback loop, (Sec. 4.1.3), which can invoke ultrasensitivity even in a first-order regime. The ultrasensitivity of the G2M transition cascade was combined with bistability and hysteresis. However, in the AMA description these phenomena have been again observed outside the parameter ranges where they have been exhibited by the EMA description. This demonstrates, that G2/M transition models based on AMA description may produce bistable behavior which does not exist according to the EMA description.

Interestingly, the zero-to-one jump point of the ultrasensitive response was shifted away from the typical value of the signal parameter $\log \kappa = 0$; a similar effect, together with the bisiogmoidal sensitivity pattern mentioned above, can be obtained in a cascade where the phosphorylation has a multisite character [95]. The multisite phosphorylation has also been shown to produce bistable behavior [56], [149], though it might depend on the exact phosphorylation mechanism [98], such as order of phosphate process-

ing (random, sequential or distributive) and the characteristics of protein-protein interactions [179]. Multisite phosphorylation does in fact occur in the basic G2/transition mechanism on several points [120] [144] and has an impact on the sensitivity of this system [119]. Thus, future research should include multiphosphorylation into the model of the G2/M transition pathway.

5.3.2 Effect of combinatorial complexity and EMA

Combinatorial complexity in first-order enzymatic cascades

The concentrations of enzymes and substrates in MIN, on contrary to purely metabolic networks, can be presumed to be of the same order of magnitude [23]. We have taken this into account in our calculations, setting equal concentrations of all proteins in a given modeled MIN (Sec. 4.2). The similarity of concentrations can lead to substrate sequestration by enzyme, which has been shown by Blüthgen *et al.* to reduce the zero-order ultrasensitivity in the MAP kinase even to subsensitivity [41]. The authors conclude from this, that the enzyme saturation shifting the reaction regime into first order cannot be a primary mechanism for generating ultrasensitivity in cell and multisite phosphorylation could be a convenient alternative here, especially because it can also generate bistability and hysteresis [41]. Our results suggest, that ultrasensitivity, bistability and hysteresis can also be achieved in sequestered systems, i.e. where enzymes act in the first-order regime, by means of feedback.

Moreover, we have observed that some part of the sigmoidal sensitivity pattern, though still not ultrasensitivity, can be recovered by two means in sequestered systems. First, due to branching the pathway's structure i.e. regulating both kinase and phosphatase. Second, because of a combinatorial definition of the response as a total of all phosphorylated forms of the substrate, i.e. including intermediate complexes with phosphatase. A similar response definition has been recently proposed by [59].

Finally, we believe that sequestration of enzyme with substrate can in some cases elevate sensitivity - if large amounts of enzyme are occupied by a high affine substrate, then only little amounts of free active enzyme are available for competing substrates, which may shift the enzymatic action back to the zero-order regime. Moreover, the reaction order can be decreased effectively by scaffold proteins which hold an enzyme and its substrate together, providing it from diffusion in the bulk aqueous phase [117].

Approximated versus elementary mass action description

We have shown, that application of AMA in first-order enzymatic cascades can lead to some substantial discrepancies to EMA, especially in systems with low Michaelis constant. First, the sensitivity pattern for AMA is slightly steeper than for EMA, which leads to response under/overestimation for extreme parameter values, which can impact threshold behavior. Similar effect may result from the fact, that the jumping points of ultrasensitive curves resulting from AMA and EMA do not overlap, so the response value for some limited parameter ranges differs by almost 100%.

Moreover, the superposition of even slight deviations in more complex systems can lead to substantial discrepancies between AMA and EMA, including differences in steady state levels or producing oscillatory and bistable behavior by AMA where it is not exhibited by EMA.

For all these reasons, we strongly discourage the still commonly practiced (Sec. 5.2) application of AMA for MIN modeling and suggest EMA instead.

Recently a new method has been proposed for MIN modeling in order to omit the imperfections of AMA, called total quasi state approximation (tQSSA), which partitions the timescales of a system described with EMA and thus allows algebraization of relatively very slow and very fast reactions in the system [59]. However, this method still requires manual composition of an EMA description and its further transformation into a tQSSA form. Thus, we consider the automated approach using aceSIM to be a convenient alternative.

5.4 Conclusions

We propose a new, combinatorial approach to modeling of molecular interaction networks (MIN) with ordinary differential equations (ODE). A MIN is described with a set of simple, user-defined rules containing names of interacting species, reaction rate constants and optional regulation coefficients. These parameters are automatically transformed into equation modules, which are automatically combined into an equation system describing all combinatorial reaction pathways possible in the modeled network. Thus, no further parameters need to be entered nor manual modification of the existing parameters is required to obtain the system description.

The ODE modules have the form of elementary mass action (EMA) kinetics, which ensures compliance with mass conservation laws even for large and complicated systems and thus greater mathematical precision compared

to approximated kinetic formalisms (AMA), such as Michaelis-Menten kinetics. The entirely automated parameter extrapolation and module combination is facilitated by an agent-like, Boolean representation of combinatorial molecular species.

We have implemented the above formalism with $Java^{TM}$. We have called the resulting software 'aceSim'. The acronym 'ace' refers to '**a**utomated, **c**ombinatorial, **e**lementary (mass action)', which we believe to be the key characteristics of the presented simulator. AceSim offers following features:

- Automatic MIN model construction, simulation and analysis based on a limited input set of interaction rules.
- Consistent treatment of various biological processes like signaling cascades, transcriptional regulation or protein degradation
- Incorporation into the model abstract terms relating to physiological phenomena.

The unique characteristics of SIMULATOR is that it complements a structure-oriented approach to MIN description with a consistent mathematical formalism using EMA kinetics, which by definition incorporates conservation relationships. This combination allows for a far-reaching automation of the modeling process, successfully coping with combinatorial complexity without compromising mathematical precision or expanding the parameter space.

AceSim can be applied to investigation of behavior of various molecular systems in the cell, like cell division, cell death or intracellular signaling and aberrations of those systems related to disease mechanisms. Various analysis tools, such as steady-state analysis or parameter sensitivity analysis allow tracking the steady states of the system that correspond to different physiological cell states and investigating cell responses to parameter changes resulting from disease, genetic mutations or pharmacological interventions.

We have investigated several signaling cascades, including the G2/M transition network responsible for cell division, using aceSim. This analysis demonstrates that application of our combinatorial and EMA-based formalism brings new insights into the behavior of enzymatic cascades, in some cases contrary to the results of the AMA-based modeling.

Especially, we have shown that ultrasensitivity is possible in sequestered enzymatic cascades by means of feedback and that combinatorial complexity can reduce oscillatory behavior of such cascades. Our results also suggest,

that the G2/M transition models based on AMA description may produce bistable behavior which does not exist according to the EMA description.

Following features should be especially considered for future development of the aceSim:

- Description of cell compartments and transport.
- Simulation with stochastic algorithms.
- Analysis of the steady state using eigenvalues of the Jacobian matrix.
- Import and export of SBML models, as soon as SBML support for rule-based network description becomes available.
- Diagrammatic GUI for model definition

Chapter 6

Zusammenfassung

6.1 Einführung

Molekulare Interaktionsnetzwerke (MIN) zeichnen sich durch gleichzeitige, multivalente Interaktionen aus [130]. Daraus ergibt sich eine Vielzahl von möglichen Kombinationen der wechselwirkenden Moleküle [73]; diese Anzahl wächst exponentiell mit der Größe des Systems [147]. Diese Eigenschaft von biologischen Netzwerken wurde als **kombinatorische Komplexität** (engl. combinatorial complexity) bezeichnet [106] und in zahlreichen biologischen Systemen beobachtet, wie z.B. zellulären Signalwegen [48] oder Stoffwechselwegen [101]. Die kombinatorische Komplexität hat dort eine hoch verzweigte Wegstruktur zur Folge, im Gegensatz zu der traditionellen, linearen Darstellung dieser Prozesse [37], [96]. Deshalb muss eine Simulationssoftware eine große Anzahl von Variabeln verarbeiten und vielschichtig aufgebaute biologische Prozesse repräsentieren können, um die kombinatorische Komplexität erfolgreich zu bewältigen

Die MIN können auf unterschiedliche Weise mathematisch beschrieben werden [198], [110], u.a. mit Systemen von gekoppelten, Gewöhnlichen Differenzialgleichungen (engl. ordinary differential equations; ODE) [71], [129], [160]. Diese Modellklasse erlaubt eine detaillierte, auf molekularen Mechanismen basierte Beschreibung von Interaktionen [208].

Dennoch basieren die klassischen, biochemischen ODE-Formulierungen, mit dem Standardbeispiel der Michaelis-Menten-Kinetik, auf Näherungen, z.B. über den stationären Charakter von Enzym-Substrat -Komplexen oder der Enzymsättigung [62]. Diese Näherungen sind für die Modellierung von

MIN nicht zwingend gültig [41], [59], weil dort Proteine sowohl Enzym, als auch Substrat sein können und dadurch sehr wohl in Konzentrationen der gleichen Größenordnung vorhanden sein können [23]. Letzteres wurde im Gegensatz zur Enzymsättigung als **Sequestrierung** bezeichnet [41]. Deshalb werden folglich zwei Subklassen von kinetischen ODE-Formalismen definiert: **Elementare Massenwirkung** (engl. elementary mass action; EMA) und **Angenäherte Massenwirkung** (engl. approximated mass action; AMA).

EMA beschreibt die molekularen Wechselwirkungen als elementare Assoziation-Dissoziation-Reaktion in Form des Massenwirkungsgesetzes, d.h. als ein Produkt von Ratenkonstanten und Reaktantkonzentrationen [134]. Die Darstellung eines bestimmten Systems mit AMA basiert auf der EMA-Beschreibung mit zusätzlichen Annahmen. Diese Annahmen vereinfachen die Massenerhaltungsrelationen (z.B. durch Vernachlässigung des Enzym-Substrat-Komplexes) oder die Dynamik von gewissen Systemvariablen (z.B. durch die quasi-stationäre Näherung des Enzym-Substrat-Komplexes). Die Beschreibung von MIN mit EMA wird für sachgerechter gehalten als mit AMA [59], [155]. Dennoch erhöht sich die Anzahl und Länge der Gleichungen für ein bestimmtes System, weshalb es schwierig ist, die Beschreibungsgenauigkeit von EMA mit der kombinatorischen Komplexität von biologischen Systemen zu vereinbaren.

Um diesem Problem entgegenzuwirken, schlagen wir einen neuartigen Ansatz vor, der auf einer automatisierten Kombination der EMA-ODE-Module beruht. Diese Module werden aus einem Satz simpler, benutzerdefinierter Regeln abgeleitet, welche die Ratenkonstanten und andere Reaktionsparameter beinhalten. Dabei werden die Ratenkonstanten extrapoliert. Der Parameterraum wird also dadurch nicht vergrößert im Vergleich zu einer AMA-Beschreibung des gleichen Systems. Wir wendeten diesen Formalismus für die Untersuchung von verschiedenen biologischen Systemen an und verglichen die Ergebnisse mit den Resultaten einer entsprechenden, klassischen AMA-Beschreibung.

6.2 Methode

In unserem Ansatz wurde ein MIN mit einem Satz von einfachen, benutzerdefinierten Regeln beschrieben, welche die Namen wechselwirkender molekularer Spezies, die Ratenkostanten und optionale Regulationsparameter beinhalteten. Es erfolgt eine automatische Ableitung dieser Parameter in

Gleichungsmodule, welche wiederum automatisch zu einem Gleichungssystem kombiniert werden. Dieses Gleichungssystem beschreibt alle möglichen kombinatorischen Reaktionswege in dem modellierten Netzwerk. Um derartige Systembeschreibung zu erzeugen ist es nicht nötig, die vorgegebenen Parameter per Hand zu modifizieren, oder neue Parameter einzugeben.

Die Gleichungsmodule haben eine Form von EMA, welche die Massenerhaltung auch für große und komplizierte Systeme sichert. Dadurch kann, im Gegensatz zu AMA-Formalismen, wie der Michaelis-Menten-Kinetik, eine größere mathematische Präzision erzielt werden.

Die komplett automatisierte Parameterextrapolation und Kombination der Gleichungsmodule wird ermöglicht durch eine Boolesche Repräsentation kombinatorischer molekularer Spezies. Wir definieren Spezies als eine Molekülart, insbesondere ein Protein (z.B. Insulin oder Cyclin B), welches in einem MIN verfügbar ist. Die Spezies können miteinander in verschiedenen Kombinationen wechselwirken. Deshalb beschreiben wir ein Molekül als Sammlung von Interaktionsschnittstellen zu anderen Spezies. Diese Schnittstellen können z.B. Bindungs- und Enzymmodifikationsstellen entsprechen. Jeder Schnittstelle wird ein Boolescher Wert zugeordnet, von dem abgelesen werden kann, ob die dazugehörige Wechselwirkung einer bestimmten Kombination von Molekülen gerade stattfindet oder nicht. Dadurch wird das gestammte MIN in zwei Ebenen unterteilt: (1) den Spezies entsprechende Nezwerkcluster, (2) Schnittstellen der einzelnen Netzwerkknoten. Die Wechselwirkungen zwischen Spezies werden als Netzwerkkanten zwischen diesen Knoten dargestellt.

Der oben beschriebene Ansatz führt dazu, dass die einzelnen Spezies als individuelle Agenten betrachtet werden können, die jeweils auf Grund einer vorgegebenen Liste von Interaktionsregeln ihre Subspezies, d.h. die Booleschen Kombinationen von Schnittstellenwerten, bestimmen können. Dabei ist hier der Begriff Agent als eine abstrakte, zur Wahrnehmung und Handlung fähige Einheit zu verstehen und nicht mit dem Fachterminus Software-Agent zu verwechseln. Unser Agent-ähnlicher Ansatz ermöglicht eine regelbasierte Bewältigung der Kombinatorischen Komplexität. Dazu beugt es auch eine künstliche Erzeugung von Polymermolekülarten vor, was als methodisches Problem für andere regelbasierte Ansätze zur MIN-Modellierung beschrieben wurde [147].

Darüber hinaus erlaubt dieser Vermittler-ähnlicher Ansatz eine beträchtliche Minmierung der kombinatorischen Variablen. In einem System mit n Spezies, die jeweils k Interaktionen zueinander haben, beträgt die Anzahl von möglichen kombinatorischen Molekülarten für jede Spezies $2k$ und im gestammten Net-

zwerk bis zu 2^{nk}. In unserem Ansatz wird aber diese Zahl auf $n2^k$ reduziert.

6.3 Implementierung

Wir haben den oben beschriebenen Formalismus mit der Programmiersprache $Java^{TM}$ implementiert. Die daraus folgende MIN-Simulationssoftware wurde als aceSim bezeichnet, wobei die Abkürzung ace für **a**utomated, **c**ombinatorial, **e**lementary (automatisiert, kombinatorisch, elementar) steht, welche die Schlüsselmerkmale unseres Modellierungsansatzes waren. Die aceSim bietet dazu folgende Anwendungsmerkmale an:

- Automatisierte Erstellung eines MIN-Modells ,einschließlich Simulation und Analyse, basierend auf einem begrenzten Eingabesatz von Interaktionsregeln.
- Einheitliche Repräsentation von verschiedener biologischer Systeme, wie
- Einbeziehung abstrakten Begriffen in das Modell, die physiologischen Phänomenen entsprechen.

Ein Vergleich zu anderen vorhandenen Softwarelösungen für MIN-Modellierung erlaubt die Formulierung verschiedener Vorteile von aceSim. Programme in diesem Bereich basieren in der Regel auf einer prozessorientierten Netzwerkbeschreibungsmethode (d.h. eine sequentielle Reaktionsliste; [124], [125], [38], [73], [74]) und können dadurch die sich aus der Vielfalt gleichzeitiger Reaktionen ergebende kombinatorische Komplexität nicht bewältigen. Für diesen Zweck wäre ein in der Theorie vorhandener strukturorientierter Ansatz besser geeignet [131], [133], [124], [132], [61]. Dennoch wurde, soweit uns bekannt ist, bisher kein mathematischer ODE-Formalismus für strukturorientierte Netzwerkbeschreibungsmethoden entwickelt. Darüber hinaus basieren die meisten Programme auf einem AMA-Formalismus, wie z.B. die Michaelis-Menten-Kinetik, welcher für die Beschreibung von MIN unter Umständen nicht geeignet ist, wie bereits in der Einführung erklärt wurde.

Deshalb besteht die Einzigartigkeit von aceSim darin, dass es eine strukturbasierte Netzwerkbeschreibung mit einem ODE-Formalismus verbindet. Darüber hinaus beruht dieser Formalismus auf der EMA-Kinetik. Diese Eigenschaften erlauben eine weitgehende Automatisierung der Modellierungsprozess und eine erfolgreiche Bewältigung der kombinatorischen Komplexität, auch bei komplizierten MIN, ohne die mathematische Präzision zu beeinträchtigen.

Die aceSim kann zur Untersuchung verschiedener zellulärer Prozesse angewandt werden, wie z.B. die Zellteilung, der Zelltod oder die Signalübertragung, oder auch zur Simulation von Störungen solcher Prozesse. Dadurch können bestimmte Krankheitsmechanismen untersucht werden. Die innerhalb von aceSim implementierten Anwendungen erlauben verschiedene Aspekte solcher Analysen zu betrachten. Beispielsweise ermöglicht die Dauerzustandsanalyse eine Verfolgung von stationären Zuständen des Systems, die unterschiedlichen physiologischen Zuständen der Zelle entsprechen. Die Parametersensitivitätsanalyse gestattet eine Untersuchung zellulärer Reaktion auf Parameterveränderungen, die z.B. aus Krankheit, genetischen Mutationen oder pharmakologischen Eingriffen resultieren.

6.4 Simulation biologischer Systeme

Wir haben verschiedene enzymatische Signalkaskaden mithilfe von aceSim simuliert, u.a. das für die Zellteilung zuständige G2/M Netzwerk. Die Anwendung von unseren kombinatorischen, EMA-basierten Ansatz hat neue Erkenntnisse über das Verhalten von Enzymkaskaden gebracht, die sich zum Teil widersprüchlich zu Ergebnissen einer AMA-basierten Simulation der gleichen Systeme erwiesen haben.

Ein biologisches System kann im Rahmen einer Signal-Reaktion-Analyse untersucht werden, wobei jede der Systemvariabeln oder Parameter als Signal oder Reaktion bezeichnet werden kann [183], [208]. In unseren Simulationen haben wir die Reaktion als Bruchkonzentration des phosphorylierten Substrats und die Signale als verschiedene Kombinationen von Reaktionsratenkonstanten definiert. Die Veränderungen dieser Parameter *in vivo* können z.B. aus Krankheit, genetischen Mutationen oder pharmakologischen Eingriffen erfolgen. Eine Veränderung des Reaktionsniveaus relativ zu einer Veränderung des Signalwerts wird als **Sensitivität** bezeichnet.

Die Sensitivitätskurven biologischer Systeme sind oftmals sigmoidal [183], eine besonders steile sigmoidale Kurve wird als **Ultrasensitivität** bezeichnet [94]. Es wurde kürzlich nachgewiesen, dass eine realistische Annahme der Sequestrierung in den enzymatischen Kaskaden die bisher angenommene Sensitivität, insbesondere die Ultrasensitivität am Ende der Kaskade, wesentlich dämpfen kann [41].

Insbesondere haben wir nachgewiesen, dass die Ultrasensitivität auch in sequestrierten Kaskaden dank Feedback möglich ist und dass die kombinatorische Komplexität in solchen Kaskaden Oszillationen reduzieren kann.

Unsere Ergebnisse weisen auch darauf hin, dass die AMA-basierten G2/M-Modelle ein bistabiles Verhalten andeuten können, das in bestimmten Parameterwertbereichen laut EMA nicht vorkommt.

Die durch Sequestrierung gedämpfte Sensitivität einer enzymatischen Kaskade kann laut unseren Befunden auch durch zwei weitere Mechanismen zumindest teilweise wiederherstellt werden. Erstens, durch eine verzweigte Kaskadenstruktur, wo sowohl die Kinase, als auch die entsprechende Phosphatase der Regulation durch Phosphorylierung unterliegt. Zweitens, durch eine kombinatorische Definition der Systemreaktion als summarische Konzentration aller phosphorylierten Substratformen, inklusive der Zwischenkomplexe durch die Phosphatase. Eine ähnliche Definition wurde vor kurzem im Bezug auf EMA von [59] vorgeschlagen. Diese Ergebnisse weisen darauf hin, dass eine effiziente Modulierung der Signalwege durch kinetische Parameterveränderungen auch unter den verschärften Annahmen über kombinatorische Komplexität und Sequestrierung erfolgen kann.

Für unterschiedliche Kaskadenstrukturen haben wir teilweise erhebliche Diskrepanzen zwischen den Ergebnissen einer AMA- und einer EMA-basierten Simulation nachgewiesen, dies gilt insbesondere für Enzyme mit einer niedrigen Michaelis-Konstante. Die aus AMA stammenden Sensitivitätskurven sind meistens ein wenig steiler im Vergleich zur EMA, was zur Unter- bzw. Überschätzung des Reaktionsniveaus für extreme Parameterwerte führt und ein Schwellenverhalten beeinflussen kann. Ähnliche Folgen kann die Diskrepanz zwischen Schwingungspunkten von ultrasensitiven Kurven haben, wodurch sich für gewisse Parameterwertbereiche Reaktionsniveauunterschiede von 100% zwischen AMA und EMA ergeben. Die Überlagerung selbst von geringen Unterschieden zwischen AMA und EMA in komplexeren Systemen kann zu wesentlichen Abweichungen führen, wie z.B. Differenzen im Reaktionsniveau, Bistabilität oder Oszillationen die laut AMA vorhanden sind und laut EMA für die gleichen Parameterwerte nicht vorkommen.

Aus diesen Gründen halten wir eine AMA-basierte Beschreibung von MIN für nicht empfehlenswert und schlagen einen EMA-basierten Ansatz vor, z.B. wie die in dieser Arbeit vorgestellte Modellierungsmethode.

Bibliography

[1] An Introduction to Biochemical Modeling using JDesigner. Reference and Tutorial Manual. Available via the World Wide Web at http://sbw.kgi.edu/software/jdesigner.htm. [cited at p. 112]

[2] The Virtual Cell 4.2 user guide. Available via the World Wide Web at http://www.nrcam.uchc.edu/login. [cited at p. 113]

[3] Interface Paradigms for Pathway Model Builders. Available via the World Wide Web at http://jigcell.biol.vt.edu/MBparadigm.html, 2004. [cited at p. 117]

[4] http://www.nrcam.uchc.edu/login/login.html, 2006. [cited at p. 113]

[5] Cellware manual Ver 3.0. Available via the World Wide Web at http://www.bii.a-star.edu.sg/docs/sbg/cellware, 2006. [cited at p. 112]

[6] Combinatorial explosion. Available via the World Wide Web at http://en.wikipedia.org/wiki/Combinatorial explosion, 2006. [cited at p. 7]

[7] COPASI Documentation. Version 4.0 (Build 18). Available via the World Wide Web at http://www.copasi.org/tiki-index.php, 2006. [cited at p. 112]

[8] Double Precision. Available via the World Wide Web at http://en.wikipedia.org/wiki/Double precision, 2006. [cited at p. 57]

[9] http://cellsignaling.lanl.gov/bionetgen/, 2006. [cited at p. 111]

[10] http://jigcell.biol.vt.edu/, 2006. [cited at p. 112]

[11] http://magnet.systemsbiology.net/software/Dizzy/, 2006. [cited at p. 112]

[12] http://sbw.kgi.edu/software/jarnac.htm, 2006. [cited at p. 112]

[13] http://sbw.kgi.edu/software/jdesigner.htm, 2006. [cited at p. 112]

[14] http://www-aig.jpl.nasa.gov/public/mls/cellerator/, 2006. [cited at p. 111]

[15] http://www.bii.a-star.edu.sg/achievements/applications/cellware/index.asp, 2006. [cited at p. 112]

[16] http://www.cadlive.jp/, 2006. [cited at p. 111]

[17] http://www.copasi.org/tiki-index.php, 2006. [cited at p. 112]

[18] http://www.duke.edu/ you, 2006. [cited at p. 112]

[19] http://www.e-cell.org/, 2006. [cited at p. 112]

[20] http://www.sbtoolbox.org/, 2006. [cited at p. 113]

[21] Intelligent agent. Available via the World Wide Web at http://en.wikipedia.org/wiki/Intelligent agent, 2006. [cited at p. 19]

[22] Jacobian. Available via the World Wide Web at http://en.wikipedia.org/wiki/Jacobian, 2006. [cited at p. 77]

[23] K.R. Albe, M.H. Butler, and B.E. Wright. Cellular concentrations of enzymes and their substrates. *J Theor Biol*, 143:163–95, 1990. [cited at p. 13, 125, 130]

[24] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J.D. Watson. *Molecular Biology of the Cell*. Garland Science New York, 2002. [cited at p. 3, 82]

[25] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman Hall/CRC Mathematical and Computational Biology Series, 2006. [cited at p. 3]

[26] P. Aloy and R.B. Russell. Ten thousand interactions for the molecular biologist. *Nat Biotechnol*, 22:1317–21, 2004. [cited at p. 4]

[27] R. Alves, F. Antunes, and A. Salvador. Tools for kinetic modeling of biochemical networks. *Nature Biotechnology*, 24:667–672, 2006. [cited at p. 111]

[28] G.D. Bader, E. Brauner, M.P. Cary, K. Dahlquist, E. Demir, P. D'Eustachio, K. Fukuda, F. Gibbons, M. Gillespie, R. Goldberg, et al. BioPAX–Biological Pathways Exchange Language Level 2, Version 1.0 Documentation. Available via the World Wide Web at http://www.biopax.org, 2006. [cited at p. 4]

[29] C.P. Bagowski, J. Besser, C.R. Frey, and J.E. Ferrell. The JNK Cascade as a Biochemical Switch in Mammalian Cells Ultrasensitive and All-or-None Responses. *Current Biology*, 13:315–320, 2003. [cited at p. 124]

[30] A.L. Barabasi and Z.N. Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5:101–113, 2004. [cited at p. 5]

[31] N. N. Batada, T. Reguly, A. Breitkreutz, L. Boucher, B. J. Breitkreutz, L. D. Hurst, and M. Tyers. Stratus not altocumulus: a new view of the yeast protein interaction network. *PLoS Biol*, 4:e317, 2006. [cited at p. 6]

[32] O. G. Berg, J. Paulsson, and M. Ehrenberg. Fluctuations and quality of control in biological cells: zero-order ultrasensitivity reinvestigated. *Biophys J*, 79:1228–1236, 2000. [cited at p. 123]

[33] F.T. Bergmann, R.R. Vallabhajosyula, and H.M. Sauro. Computational Tools for Modeling Protein Networks. *Current Proteomics*, 3:181–197, 2006. [cited at p. 111]

[34] A. Beyer and T. Wilhelm. Dynamic simulation of protein complex formation on a genomic scale. *Bioinformatics*, 21:1610, 2005. [cited at p. 7]

[35] U.S. Bhalla. The chemical organization of signaling interactions. *Bioinformatics*, 18:855–863, 2002. [cited at p. 8]

[36] U.S. Bhalla, P.T. Ram, and R. Iyengar. MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network. *Science*, 297:1018–23, 2002. [cited at p. 124]

[37] M.L. Blinov, J.R. Faeder, B. Goldstein, and W.S. Hlavacek. A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *Biosystems*, 83:136–151, 2006. [cited at p. 7, 129]

[38] M.L. Blinov, J.R. Faeder, J.R. Goldstein, and W.S. Hlavacek. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*, 20:3289–3291, 2004. [cited at p. 104, 111, 132]

[39] M.L. Blinov, Faeder. J.R., B. Goldstein, A. Finney, and W.S. Hlavacek. Rule-based modeling of multi-component species. Proposal for SBML level 3. Available via the World Wide Web at http://www.cds.caltech.edu/ afinney/LA-UR-04-5445-Oct6.pdf, 2004. [cited at p. 118]

[40] M.L. Blinov, J. Yang, J.R. Faeder, and W.S. Hlavacek. Depicting signaling cascades. *Nat. Biotechnol*, 24:137–138, 2006. [cited at p. 104]

[41] N. Bluthgen, F.J. Bruggeman, S. Legewie, H. Herzel, H.V. Westerhoff, and B.N. Kholodenko. Effects of sequestration on signal transduction cascades. *FEBS J*, 273:895–906, 2006. [cited at p. 11, 13, 81, 90, 123, 125, 130, 133]

[42] N.M. Borisov, N.I. Markevich, J.B. Hoek, and B.N. Kholodenko. Signaling through Receptors and Scaffolds: Independent Interactions Reduce Combinatorial Complexity. *Biophysical Journal*, 89:951–966, 2005. [cited at p. 8]

[43] N.M. Borisov, N.I. Markevich, J.B. Hoek, and B.N. Kholodenko. Trading the micro-world of combinatorial complexity for the macro-world of protein interaction domains. *Biosystems*, 83:152–166, 2006. [cited at p. 8]

[44] D. Botstein and K.W. Kohn. Molecular Interaction Map of the Mammalian Cell Cycle Control and DNA Repair Systems. *Molecular Biology of the Cell*, 10:2703–2734, 1999. [cited at p. 82]

[45] J.M. Bower and H. Bolouri. *Computational Modeling of Genetic and Biochemical Networks*. MIT Press, 2001. [cited at p. 3]

[46] D. Bray. Genomics. Molecular prodigality. *Science*, 299:1189–90, 2003. [cited at p. 4]

[47] D. Bray and S. Lay. Computer-based analysis of the binding steps in protein complex formation. *Proc Natl Acad Sci US A*, 94:13493–8, 1997. [cited at p. 7]

[48] R. Breitling and D. Hoeller. Current challenges in quantitative modeling of epidermal growth factor signaling. *FEBS Lett*, 579:6289–94, 2005. [cited at p. 7, 129]

[49] N.F. Britton. *Essential Mathematical Biology*. Springer, 2004. [cited at p. 3]

[50] K.R. Brown and I. Jurisica. Online Predicted Human Interaction Database. *Bioinformatics*, 21:2076–2082, 2005. [cited at p. 5]

[51] S.C. Bunnell, D.I. Hong, J.R. Kardon, T. Yamazaki, C.J. McGlade, V.A. Barr, and L.E. Samelson. T cell receptor ligation induces the formation of dynamically regulated signaling assemblies. *The Journal of Cell Biology*, 158:1263–1275, 2002. [cited at p. 7]

[52] S.C. Bunnell, A.L. Singer, D.I. Hong, B.H. Jacque, M.S. Jordan, M.C. Seminario, V.A. Barr, G.A. Koretzky, and L.E. Samelson. Persistence of Cooperatively Stabilized Signaling Clusters Drives T-Cell Activation. *Molecular and Cellular Biology*, 26:7155–7166, 2006. [cited at p. 7]

[53] F. Campagne, S. Neves, C.W. Chang, L. Skrabanek, P.T. Ram, R. Iyengar, and H. Weinstein. Quantitative information management for the biochemical computation of cellular networks. *Sci STKE*, 2004:248, 2004. [cited at p. 115]

[54] G. Chaurasia, Y. Iqbal, C. Hanig, H. Herzel, E. E. Wanker, and M. E. Futschik. UniHI: an entry gate to the human protein interactome. *Nucleic Acids Res*, 35:D590–D594, 2007. [cited at p. 5]

[55] K. C. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak, and J. J. Tyson. Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol Biol Cell*, 11:369–391, 2000. [cited at p. 81]

[56] J. L. Cherry and F. R. Adler. How to make a biological switch. *J Theor Biol*, 203:117–133, 2000. [cited at p. 124]

[57] C. Chien, P.L. Bartel, R. Sternglanz, and S. Fields. The Two-Hybrid System: A Method to Identify and Clone Genes for Proteins that Interact with a Protein of Interest. *Proceedings of the National Academy of Sciences*, 88:9578–9582, 1991. [cited at p. 4]

[58] P.B. Chock and E.R. Stadtman. Superiority of Interconvertible Enzyme Cascades in Metabolic Regulation: Analysis of Multicyclic Systems. *Proceedings of the National Academy of Sciences*, 74:2766–2770, 1977. [cited at p. 122]

[59] A. Ciliberto, F. Capuani, and J.J. Tyson. Modeling Networks of Coupled Enzymatic Reactions Using the Total Quasi-Steady State Approximation. *PLOS Comp. Biol*, 3:0463–72, 2007. [cited at p. 11, 13, 125, 126, 130, 134]

[60] H. Conzelmann, J. Saez-Rodriguez, T. Sauter, B.N. Kholodenko, and E.D. Gilles. A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. *BMC Bioinformatics*, 7:34, 2006. [cited at p. 7]

[61] D.L. Cook, J.F. Farley, and S.J. Tapscott. A basis for a visual language for describing, archiving and analyzing functional models of complex biological systems. *Genome Biology*, 2:1–3, 2001. [cited at p. 104, 107, 132]

[62] A. Cornish-Bowden. *Fundamentals of enzyme kinetics*. Portland Press, 2004. [cited at p. 9, 10, 11, 12, 13, 22, 23, 123, 129]

[63] A. Csikasz-Nagy, D. Battogtokh, K.C. Chen, B. Novak, and J.J. Tyson. Analysis of a Generic Model of Eukaryotic Cell-Cycle Regulation. *Biophysical Journal*, 90:4361–4379, 2006. [cited at p. 82]

[64] M. Dejori and M. Stetter. Bayesian Inference of Genetic Networks from Gene-Expression Data: Convergence and Reliability. *Proceedings of the 2003 International Conference on Artificial Intelligence (IC-AI03)*, pages 321–327, 2003. [cited at p. 8]

[65] P.K. Dhar, T.C. Meng, S. Somani, L. Ye, K. Sakharkar, A. Krishnan, A.B.M. Ridwan, S.H.K. Wah, M. Chitre, and Z. Hao. Grid Cellware: the first grid-enabled tool for modelling and simulating cellular processes. *Bioinformatics*, 21:1284–1287, 2005. [cited at p. 112]

[66] J. R. Dormand and P. J. Prince. A family of embedded runge-kutta formulae. *J Comp Appl Mat*, 6:19–26, 1980. [cited at p. 75, 116]

[67] A.K. Dunker, M.S. Cortese, P. Romero, L.M. Iakoucheva, and V.N. Uversky. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS JOURNAL*, 272:5129, 2005. [cited at p. 6]

[68] L. Edelstein-Keshet. *Mathematical Models in Biology (Classics in Applied Mathematics)*. SIAM, 2005. [cited at p. 3]

[69] D. Ekman, S. Light, Å.K. Björklund, and A. Elofsson. What properties characterize the hub proteins of the protein-protein interaction network of Saccharomyces cerevisiae? *Genome Biology*, 7:R45, 2006. [cited at p. 6]

[70] D. Endy and R. Brent. Modelling cellular behaviour. *Nature*, 409:391–395, 2001. [cited at p. 8]

[71] N.J. Eungdamrong and R. Iyengar. Modeling cell signaling networks. *Biology of the Cell*, 96:355–362, 2004. [cited at p. 9, 11, 129]

[72] H. Eyring. The Activated Complex in Chemical Reactions (reprint). *Journal of Chemical Physics*, 3:107–115, 2004. [cited at p. 9]

[73] J.R. Faeder, M.L. Blinov, B. Goldstein, and W.S. Hlavacek. Rule-based modeling of biochemical networks. *Complexity*, 10:22–41, 2005. [cited at p. 7, 104, 129, 132]

[74] J.R. Faeder, M.L. Blinov, and W.S. Hlavacek. Graphical rule-based representation of signal-transduction networks. *Proceedings of the 2005 ACM symposium on Applied computing*, pages 133–140, 2005. [cited at p. 7, 104, 132]

[75] J.R. Faeder, M.L. Blinov, W.S. Hlavacek, and B. Goldstein. Networks That Govern Complex Formation during Signal Transduction Exhibit Narrow Flows. Available via the World Wide Web at http://www.t10.lanl.gov/faeder/Reprints/cc abstract.pdf. [cited at p. 8]

[76] C.P. Fall, E.S. Marland, J.M. Wagner, and J.J. Tyson. *Computational Cell Biology*. Springer, 2005. [cited at p. 3]

[77] J. E. Jr Ferrell and E. M. Machleder. The biochemical basis of an all-or-none cell fate switch in xenopus oocytes. *Science*, 280:895–898, 1998. [cited at p. 122]

[78] J.E. Ferrell. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Chem. Biol*, 6:140–148, 2002. [cited at p. 122]

[79] J.E. Ferrell Jr. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends Biochem Sci*, 21:460–6, 1996. [cited at p. 122]

[80] J.E. Ferrell Jr. How responses get more switch-like as you move down a protein kinase cascade. *Trends Biochem Sci*, 22:288–9, 1997. [cited at p. 122, 123]

[81] S. Fields and O. Song. A novel genetic system to detect protein-protein interactions. *Nature*, 340:245–246, 1989. [cited at p. 4]

[82] A. Finney. Systems Biology Markup Language (SBML) Level 3 Proposal: Multi-component Species Features. Proposal manuscript. Available via the World Wide Web at http://www.cds.caltech.edu/ afinney/multi-component-species.pdf, 2004. [cited at p. 118]

[83] A. Finney and M. Hucka. Systems biology markup language: Level 2 and beyond. *Biochem. Soc. Trans*, 31:1472–1473, 2003. [cited at p. 4, 118]

[84] E. Formstecher, S. Aresta, V. Collura, A. Hamburger, A. Meil, A. Trehin, C. Reverdy, V. Betin, S. Maire, C. Brun, et al. Protein interaction mapping: A Drosophila case study. *Genome Res*, 15:376–384, 2005. [cited at p. 5]

[85] M.Y. Galperin. The Molecular Biology Database Collection: 2006 update. *Nucleic Acids Res*, 34:D3–5, 2006. [cited at p. 4]

[86] T.K. Gandhi, J. Zhong, S. Mathivanan, L. Karthick, K.N. Chandrika, S.S. Mohan, S. Sharma, S. Pinkert, S. Nagaraju, B. Periaswamy, et al. Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat Genet*, 38:285–93, 2006. [cited at p. 5]

[87] A.C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L.J. Jensen, S. Bastuck, B. Dümpelfeld, et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 440:631–636, 2006. [cited at p. 6]

[88] A.C. Gavin, M. Bösche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J.M. Rick, A.M. Michon, C.M. Cruciat, et al. Functional

organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415:141–147, 2002. [cited at p. 5]

[89] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81:2340–2361, 1977. [cited at p. 10, 116]

[90] D.T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115:1716–1733, 2001. [cited at p. 116]

[91] L. Giot, J.S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y.L. Hao, C.E. Ooi, B. Godwin, E. Vitols, et al. A Protein Interaction Map of Drosophila melanogaster. *Science*, 302:1727–1736, 2003. [cited at p. 5]

[92] R.N. Goldberg. Thermodynamics of enzyme-catalyzed reactions–a database for quantitative biochemistry. *Bioinformatics*, 20:2874–2877, 2004. [cited at p. 10]

[93] A. Goldbeter. A Minimal Cascade Model for the Mitotic Oscillator Involving Cyclin and cdc2 Kinase. *Proceedings of the National Academy of Sciences*, 88:9107–9111, 1991. [cited at p. 122, 124]

[94] A. Goldbeter and D. E. Jr Koshland. An amplified sensitivity arising from covalent modification in biological systems. *Proc Natl Acad Sci U S A*, 78:6840–6844, 1981. [cited at p. 81, 88, 122, 123, 133]

[95] A. Goldbeter and D. E. Jr Koshland. Ultrasensitivity in biochemical systems controlled by covalent modification. interplay between zero-order and multistep effects. *J Biol Chem*, 259:14441–14447, 1984. [cited at p. 122, 124]

[96] B. Goldstein, J.R. Faeder, and W.S. Hlavacek. Mathematical and computational models of immune-receptor signalling. *Nature Reviews Immunology*, 4:445–456, 2004. [cited at p. 7, 129]

[97] B.D. Gomperts, I.M. Kramer, and P.E.R. Tatham. *Signal transduction*. Elsevier Academic Press, 2003. [cited at p. 3]

[98] J. Gunawardena. Multisite protein phosphorylation makes a good threshold but can be a poor switch. *Proc Natl Acad Sci U S A*, 102:14617–14622, 2005. [cited at p. 124]

[99] J.D. Han, N. Bertin, T. Hao, D.S. Goldberg, G.F. Berriz, L.V. Zhang, D. Dupuy, A.J. Walhout, M.E. Cusick, F.P. Roth, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, 430:88–93, 2004. [cited at p. 5, 6]

[100] D. G. Hardie, I. P. Salt, S. A. Hawley, and S. P. Davies. Amp-activated protein kinase: an ultrasensitive system for monitoring cellular energy charge. *Biochem J*, 338 (Pt 3):717–722, 1999. [cited at p. 122]

[101] V. Hatzimanikatis, C. Li, J.A. Ionita, and L.J. Broadbelt. Metabolic networks: enzyme function and metabolite structure. *Curr Opinion Struct Biol*, 14:300–306, 2004. [cited at p. 7, 129]

[102] C. Haynes, C. J. Oldfield, F. Ji, N. Klitgord, M. E. Cusick, P. Radivojac, V. N. Uversky, M. Vidal, and L. M. Iakoucheva. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol*, 2:e100, 2006. [cited at p. 6]

[103] R. Heinrich and S. Schuster. *The regulation of cellular systems*. ITP International Thomson Publishing, 1996. [cited at p. 11]

[104] H. Hermjakob, L. Montecchi-Palazzi, G. Bader, J. Wojcik, L. Salwinski, A. Ceol, S. Moore, S. Orchard, U. Sarkans, C. von Mering, et al. The HUPO PSI's Molecular Interaction formata community standard for the representation of protein interaction data. *Nature Biotechnology*, 22:177–183, 2004. [cited at p. 4]

[105] H. Hermjakob, L. Montecchi-Palazzi, C. Lewington, S. Mudali, S. Kerrien, S. Orchard, M. Vingron, B. Roechert, P. Roepstorff, A. Valencia, et al. IntAct: an open source molecular interaction database. *Nucleic acids research*, 32:452–455, 2004. [cited at p. 4]

[106] W.S. Hlavacek, J.R. Faeder, M.L. Blinov, A.S. Perelson, and B. Goldstein. The complexity of complexes in signal transduction. *Biotechnology and Bioengineering*, 84:783–794, 2003. [cited at p. 7, 129]

[107] Y. Ho, A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, et al. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature*, 415:180–183, 2002. [cited at p. 5]

[108] CY Huang and J.E. Ferrell Jr. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc Natl Acad Sci US A*, 93:10078–83, 1996. [cited at p. 123]

[109] M. Hucka, A. Finney, H.M. Sauro, H. Bolouri, J.C. Doyle, H. Kitano, A. Cornish-Bowden, A.A. Cuellar, S. Dronov, E.D. Gilles, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19:524–531, 2003. [cited at p. 4, 118]

[110] T. Ideker and D. Lauffenburger. Building with a scaffold: emerging strategies for high-to low-level cellular modeling. *Trends Biotechnol*, 21:255–262, 2003. [cited at p. 8, 10, 129]

[111] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences*, 98:4569–4574, 2001. [cited at p. 5]

[112] M.R. Jackman and J.N. Pines. Cyclins and the G2/M transition. *Cancer Surv*, 29:47–73, 1997. [cited at p. 82]

[113] H. Jeong, S.P. Mason, A.L. Barabasi, and Z.N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411:41–2, 2001. [cited at p. 6]

[114] M. Kanehisa and S. Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28:27–30, 2000. [cited at p. 115]

[115] M. Kaufman, F. Andris, and O. Leo. A logical analysis of T cell activation and anergy. *Proc Natl Acad Sci US A*, 96:3894–9, 1999. [cited at p. 9]

[116] D.B. Kell. Metabolomics and systems biology: making sense of the soup. *Current Opinion in Microbiology*, 7:296–307, 2004. [cited at p. 10]

[117] B.N. Kholodenko. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur J Biochem*, 267:1583–8, 2000. [cited at p. 124, 125]

[118] A.M. Kierzek. STOCKS: STOChastic Kinetic Simulations of biochemical systems with Gillespie algorithm. *Bioinformatics*, 18:470–81, 2002. [cited at p. 10]

[119] S.Y. Kim and J.E. Ferrell Jr. Substrate competition as a source of ultrasensitivity in the inactivation of wee1. *Cell*, 128:1133–45, 2007. [cited at p. 125]

[120] S.Y. Kim, E.J. Song, K.J. Lee, and J.E. Ferrell Jr. Multisite M-Phase Phosphorylation of Xenopus Wee1A. *Molecular and Cellular Biology*, 25:10580–10590, 2005. [cited at p. 82, 125]

[121] H. Kitano. *Foundations of systems biology*. MIT Press Cambridge, 2001. [cited at p. 3]

[122] H. Kitano. Computational systems biology. *Nature*, 420:206–210, 2002. [cited at p. 3]

[123] H. Kitano. Systems Biology: A Brief Overview. *Science*, 295:1662, 2002. [cited at p. 3]

[124] H. Kitano. A graphical notation for biochemical networks. *BIOSILICO*, 1:169–176, 2003. [cited at p. 104, 132]

[125] H. Kitano, A. Funahashi, Y. Matsuoka, and K. Oda. Using process diagrams for the graphical representation of biological networks. *Nature Biotechnology*, 23:961–966, 2005. [cited at p. 104, 132]

[126] H. Kitano, Y. Matsuoka, A. Funahashi, and K. Oda. The Process Diagram: Rationale and Definition. Available via the World Wide Web at http://www.systems-biology.org/cd/documents/ProcessDiagram.html, 2004. [cited at p. 104]

[127] R. Kleppe, E. Kjarland, and F. Selheim. Proteomic and Computational Methods in Systems Modeling of Cellular Signaling. *Current Pharmaceutical Biotechnology*, 7:135–145, 2006. [cited at p. 4, 111]

[128] E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach. *Systems Biology in Practice: Concepts, Implementation and Application*. Wiley-VCH, 2005. [cited at p. 3]

[129] E. Klipp and W. Liebermeister. Mathematical modeling of intracellular signaling pathways. *BMC Neuroscience*, 7:S10, 2006. [cited at p. 9, 11, 129]

[130] K.W. Kohn and M.I. Aladjem. The Mathematics of Networks: Molecular Interaction Maps of Bioregulatory Networks. *SIAM News*, 37, 2004. [cited at p. 7, 129]

[131] K.W. Kohn and M.I. Aladjem. Circuit diagrams for biological networks. *Mol Syst Biol*, 2:Eupub 17Jan2006, 2006. [cited at p. 104, 132]

[132] K.W. Kohn, M.I. Aladjem, S. Kim, J.N. Weinstein, and Y. Pommier. Depicting combinatorial complexity with the molecular interaction map notation. *Mol Syst Biol*, 2:51, 2006. [cited at p. 104, 132]

[133] K.W. Kohn, M.I. Aladjem, J.N. Weinstein, and Y. Pommier. Molecular Interaction Maps of Bioregulatory Networks: A General Rubric for Systems Biology. *Mol Biol Cell*, 17:1–13, 2006. [cited at p. 104, 132]

[134] A.B. Koudriavtsev. *The law of mass action.* Springer, 2001. [cited at p. 9, 130]

[135] A. Kriete and R. Eils. *Computational Systems Biology* . Academic Press, 2005. [cited at p. 3]

[136] H. Kurata, K. Masaki, Y. Sumida, and R. Iwasaki. CADLIVE dynamic simulator: direct link of biochemical networks to dynamic models. *Genome Res*, 15:590–600, 2005. [cited at p. 111]

[137] D. C. LaPorte and D. E. Koshland Jr. Phosphorylation of isocitrate dehydrogenase as a demonstration of enhanced sensitivity in covalent regulation. *Nature*, 305:286–290, 1983. [cited at p. 122]

[138] D. C. LaPorte, K. Walsh, and D. E. Jr Koshland. The branch point effect. ultrasensitivity and subsensitivity to metabolic control. *J Biol Chem*, 259:14068–14075, 1984. [cited at p. 122]

[139] R.S. Larson. *Bioinformatics and Drug Discovery (Methods in Molecular Biology).* Humana Press, 2005. [cited at p. 3]

[140] N. Le Novère, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, et al. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res*, 34:689–691, 2006. [cited at p. 118]

[141] B. Lehner and A.G. Fraser. A first-draft human protein-interaction map. *Genome Biol*, 5:R63, 2004. [cited at p. 5]

[142] S. Li, C.M. Armstrong, N. Bertin, H. Ge, S. Milstein, M. Boxem, P.O. Vidalain, J.D. Han, A. Chesneau, T. Hao, et al. A map of the

interactome network of the metazoan C. elegans. *Science*, 303:540–3, 2004. [cited at p. 5]

[143] W. Liebermeister and E. Klipp. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor Biol Med Model*, 3:41, 2006. [cited at p. 10]

[144] F. Liu, J.J. Stanton, Z. Wu, and H. Piwnica-Worms. The human Myt1 kinase preferentially phosphorylates Cdc2 on threonine 14 and localizes to the endoplasmic reticulum and Golgi complex. *Molecular and Cellular Biology*, 17:571–583, 1997. [cited at p. 82, 125]

[145] H. Lodish, M.P. Scott, P. Matsudaira, J. Darnell, L. Zipursky, C.A. Kaiser, A. Berk, and M. Krieger. *Molecular cell biology*. WH Freeman New York, 2003. [cited at p. 3, 82]

[146] L.M. Loew and J.C. Schaff. The Virtual Cell: a software environment for computational cell biology. *Trends Biotechnol*, 19:401–406, 2001. [cited at p. 11, 113]

[147] L. Lok and R. Brent. Automatic generation of cellular reaction networks with Moleculizer 1. 0. *Nature Biotechnology*, 23:131–136, 2005. [cited at p. 7, 8, 10, 11, 19, 129, 131]

[148] M. Mann, R.C. Hendrickson, and A. Pandey. Analysis of proteins and proteomes by mass spectrometry. *Annual Review of Biochemistry*, 70:437–473, 2001. [cited at p. 4]

[149] N. I. Markevich, J. B. Hoek, and B. N. Kholodenko. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *J Cell Biol*, 164:353–359, 2004. [cited at p. 124]

[150] G. Marlovits, C.J. Tyson, B. Novak, and J.J. Tyson. Modeling M-phase control in Xenopus oocyte extracts: the surveillance mechanism for unreplicated DNA. *Biophys. Chem*, 72:169–184, 1998. [cited at p. 82]

[151] S. Mathivanan, B. Periaswamy, T. Gandhi, K. Kandasamy, S. Suresh, R. Mohmood, Y. Ramachandra, and A. Pandey. An evaluation of human protein-protein interaction data in the public domain. *BMC Bioinformatics*, 7 Suppl 5:S19, 2006. [cited at p. 4]

[152] M. H. Meinke, J. S. Bishop, and R. D. Edstrom. Zero-order ultrasensitivity in the regulation of glycogen phosphorylase. *Proc Natl Acad Sci U S A*, 83:2865–2868, 1986. [cited at p. 122]

[153] G. J. Melen, S. Levy, N. Barkai, and B. Z. Shilo. Threshold responses to morphogen gradients by zero-order ultrasensitivity. *Mol Syst Biol*, 1:28, 2005. [cited at p. 81]

[154] P. Mendes. GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Comput Appl Biosci*, 9:563–571, 1993. [cited at p. 112]

[155] T. Millat, E. Bullinger, J. Rohwer, and O. Wolkenhauer. Approximations and their consequences for dynamic modelling of signal transduction pathways. *Math Biosci*, 207:40–57, 2007. [cited at p. 13, 81, 130]

[156] P.J. Mohr and B.N. Taylor. CODATA recommended values of the fundamental physical constants: 2002. *Reviews of Modern Physics*, 77:1–107, 2005. [cited at p. 9, 10]

[157] C.G. Moles, P. Mendes, and J.R. Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res*, 13:2467–74, 2003. [cited at p. 10]

[158] J.D. Murray. *Mathematical Biology I.* Springer, 2004. [cited at p. 3, 10]

[159] V. K. Mutalik, P. Shah, and K. V. Venkatesh. Allosteric interactions and bifunctionality make the response of glutamine synthetase cascade system of escherichia coli robust and ultrasensitive. *J Biol Chem*, 278:26327–26332, 2003. [cited at p. 122]

[160] S.R. Neves and R. Iyengar. Modeling of signaling networks. *BioEssays*, 24:1110–1117, 2002. [cited at p. 9, 11, 111, 129]

[161] B. Novak, Z. Pataki, A. Ciliberto, and J.J. Tyson. Mathematical model of the cell division cycle of fission yeast. *Chaos*, 11:277, 2001. [cited at p. 82]

[162] S. Orchard, L. Montecchi-Palazzi, H. Hermjakob, and R. Apweiler. The use of common ontologies and controlled vocabularies to enable data exchange and deposition for complex proteomic experiments. *Pac Symp Biocomput*, 186:96, 2005. [cited at p. 4]

[163] F. Ortega, L. Acerenza, H.V. Westerhoff, F. Mas, and M. Cascante. Product dependence and bifunctionality compromise the ultrasensitivity of signal transduction cascades. *Proceedings of the National Academy of Sciences*, 99:1170–1175, 2002. [cited at p. 123]

[164] A. Pandey and M. Mann. Proteomics to study genes and genomes. *Nature*, 405:837–846, 2000. [cited at p. 4]

[165] J.A. Papin, T. Hunter, B.O. Palsson, and S. Subramaniam. Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol*, 6:99–111, 2005. [cited at p. 4]

[166] T. Pawson and P. Nash. Assembly of Cell Regulatory Systems Through Protein Interaction Domains. *Science's STKE*, 300:445–452, 2003. [cited at p. 3]

[167] T. Pawson, M. Raina, and P. Nash. Interaction domains: from simple binding events to complex cellular behavior. *FEBS Lett*, 513:2–10, 2002. [cited at p. 3]

[168] M. Persico, A. Ceol, C. Gavrila, R. Hoffmann, A. Florio, and G. Cesareni. HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms. *BMC Bioinformatics*, 6:S21, 2005. [cited at p. 5]

[169] A. Pettinen, T. Aho, O.P. Smolander, T. Manninen, A. Saarinen, K.L. Taattola, O. Yli-Harja, and M.L. Linne. Simulation tools for biochemical networks: evaluation of performance and usability. *Bioinformatics*, 21:357–363, 2005. [cited at p. 111]

[170] O. Puig, F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, and B. Seraphin. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, 24:218–29, 2001. [cited at p. 4]

[171] A.K. Ramani, R.C. Bunescu, R.J. Mooney, and E.M. Marcotte. Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome. *Genome Biol*, 6:5, 2005. [cited at p. 5]

[172] S. Ramsey. Dizzy User Manual. Available via the World Wide Web at http://magnet.systemsbiology.net/software/Dizzy/docs/UserManual.pdf, 2005. [cited at p. 112]

[173] S. Ramsey, D. Orrell, and H. Bolouri. Dizzy: stochastic simulation of large-scale genetic regulatory networks. *J. Bioinf. Comp. Biol*, 3:415–436, 2005. [cited at p. 112]

[174] T. Reguly, A. Breitkreutz, L. Boucher, B.J. Breitkreutz, G.C. Hon, C.L. Myers, A. Parsons, H. Friesen, R. Oughtred, A. Tong, et al. Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. *J Biol*, 5:11, 2006. [cited at p. 5]

[175] M. Remm, C.E. Storm, and E.L. Sonnhammer. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol*, 314:1041–52, 2001. [cited at p. 4]

[176] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Séraphin. A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnology*, 17:1030–1032, 1999. [cited at p. 4]

[177] P.M. Roberts. Mining literature for systems biology. *Brief Bioinform*, 7:399, 2006. [cited at p. 4]

[178] J.F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G.F. Berriz, F.D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 437:1173–1178, 2005. [cited at p. 5]

[179] C. Salazar and T. Höfer. Versatile regulation of multisite protein phosphorylation by the order of phosphate processing and proteinprotein interactions. *European Journal of Biochemistry*, 274:1046–1061, 2007. [cited at p. 125]

[180] H.M. Sauro. Jarnac: A system for interactive metabolic analysis. In Rohwer J. Hofmeyr, J.-H. and J. Snoep, editors, *Animating the Cellular Map: Proceedings of the 9th International Meeting on Bio-ThermoKinetics.*, volume 33, pages 221–228. Stellenbosch University Press, 2000. [cited at p. 112]

[181] H.M. Sauro. The Computational Versatility of Proteomic Signaling Networks. *Current Proteomics*, 1:67–81, 2004. [cited at p. 79]

[182] H.M. Sauro, M. Hucka, A. Finney, C. Wellock, H. Bolouri, J. Doyle, and H. Kitano. Next Generation Simulation Tools: The Systems Biology Workbench and BioSPICE Integration. *Omics A Journal of Integrative Biology*, 7:355–372, 2003. [cited at p. 108, 111]

[183] H.M. Sauro and B.N. Kholodenko. Quantitative analysis of signaling networks. *Progress in Biophysics & Molecular Biology*, 86:5–43, 2004. [cited at p. 79, 122, 133]

[184] M.A. Savageau. *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology,(1976)*. Reading, MA Addison-Wesley, 1976. [cited at p. 124]

[185] H. Schmidt. Systems Biology Toolbox for MATLAB. Tutorial. Available via the World Wide Web at http://www.sbtoolbox.org/. [cited at p. 113]

[186] H. Schmidt and M. Jirstrand. Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics*, 22:514–515, 2006. [cited at p. 113]

[187] I. Schomburg, A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn, D. Schomburg, I. Schomburg, A. Chang, C. Ebeling, et al. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Research*, 32:431–433, 2004. [cited at p. 10]

[188] I. Schomburg, A. Chang, and D. Schomburg. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res*, 30:47–9, 2002. [cited at p. 10]

[189] B. Schwikowski, P. Uetz, and S. Fields. A network of protein- protein interactions in yeast. *Nature Biotechnology*, 18:1257–1261, 2000. [cited at p. 5]

[190] A.M. Sengupta. *Modeling Biomolecular Networks: An Introduction to Systems Biology*. Oxford University Press, 2006. [cited at p. 3]

[191] B.E. Shapiro, A. Levchenko, E.M. Meyerowitz, B.J. Wold, and E.D. Mjolsness. Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics*, 19:677–678, 2003. [cited at p. 111]

[192] G. P. Singh, M. Ganapathi, and D. Dash. Role of intrinsic disorder in transient interactions of hub proteins. *Proteins*, 66:761–765, 2006. [cited at p. 6]

[193] B.M. Slepchenko, J.C. Schaff, J.H. Carson, and L.M. Loew. Computational Cell Biology: Spatiotemporal Simulation of Cellular Events.

Annual Review of Biophysics and Biomolecular Structure, 31:423–441, 2002. [cited at p. 11]

[194] C.A. Stanyon, G. Liu, B.A. Mangiola, N. Patel, L. Giot, B. Kuang, H. Zhang, J. Zhong, and R.L. Finley Jr. A Drosophila protein-interaction map centered on cell-cycle regulators. *Genome Biol*, 5:R96, 2004. [cited at p. 5]

[195] G.R. Stark and W.R. Taylor. Control of the G2/M transition. *Molecular Biotechnology*, 32:227–248, 2006. [cited at p. 82]

[196] J.I. Steinfeld, J.S. Francisco, and L. William. *Chemical kinetics and dynamics*. Prentice Hall Upper Saddle River, NJ, 1999. [cited at p. 10, 11]

[197] U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F.H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, et al. A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome. *Cell*, 122:957–968, 2005. [cited at p. 5]

[198] M. Stetter, B. Schurmann, and M. Dejori. Systems Level Modeling of Gene Regulatory Networks. *Artificial Intelligence Methods and Tools for Systems Biology*, 5:1–20, 2004. [cited at p. 8, 129]

[199] J.R. Stiles and T.M. Bartol. Monte Carlo Methods for Simulating Realistic Synaptic Microphysiology Using MCell. In E. De Schutter, editor, *Computational Neuroscience: Realistic Modeling for Experimentalists*, pages 87–127. CRC Press, Boca Raton, Florida, 2001. [cited at p. 11]

[200] L. Strömbäck, P. Lambrix, and O. Journals. Representations of molecular pathways: an evaluation of SBML, PSI MI and BioPAX. *Bioinformatics*, 21:4401–4407, 2005. [cited at p. 4]

[201] Z. Szallasi, J. Stelling, and V. Periwal. *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*. MIT Press, 2006. [cited at p. 3]

[202] K. Takahashi, N. Ishikawa, Y. Sadamoto, H. Sasamoto, S. Ohta, A. Shiozawa, F. Miyoshi, Y. Naito, Y. Nakayama, and M. Tomita. E-Cell 2: Multi-platform E-Cell simulation system. *Bioinformatics*, 19:1727–1729, 2003. [cited at p. 112]

[203] K. Takahashi, K. Kaizu, B. Hu, and M. Tomita. A multi-algorithm, multi-timescale method for cell simulation. *Bioinformatics*, 20:538–546, 2004. [cited at p. 112]

[204] M. Tomita, K. Hashimoto, K. Takahashi, T.S. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J.C. Venter, et al. E-CELL: software environment for whole-cell simulation. *Bioinformatics*, 15:72–84, 1999. [cited at p. 112]

[205] A.H. Tong, M. Evangelista, A.B. Parsons, H. Xu, G.D. Bader, N. Page, M. Robinson, S. Raghibizadeh, C.W. Hogue, H. Bussey, et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, 294:2364–8, 2001. [cited at p. 4]

[206] A.H.Y. Tong, G. Lesage, G.D. Bader, H. Ding, H. Xu, X. Xin, J. Young, G.F. Berriz, R.L. Brost, M. Chang, et al. Global Mapping of the Yeast Genetic Interaction Network. *Science*, 303:808–813, 2004. [cited at p. 5]

[207] K.Y. Tsai and F.S. Wang. Evolutionary optimization with data collocation for reverse engineering of biological networks. *Bioinformatics*, 21:1180–1188, 2005. [cited at p. 10]

[208] J.J. Tyson, K.C. Chen, and B. Novak. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol*, 15:221–231, 2003. [cited at p. 14, 79, 122, 129, 133]

[209] J.J. Tyson and H.G. Othmer. The dynamics of feedback control circuits in biochemical pathways. *Progress in theoretical biology*, 5:1–62, 1978. [cited at p. 124]

[210] P. Uetz, L. Giot, G. Cagney, T.A. Mansfield, R.S. Judson, J.R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, et al. A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae. *Nature*, 403:623–627, 2000. [cited at p. 5]

[211] S. Vaidyanathan, G.S. Harrigan, and R. Goodacre. *Metabolome Analyses: Strategies for Systems Biology*. Springer, 2005. [cited at p. 3]

[212] N.A.W. van Riel. Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Briefings in Bioinformatics*, 7:364, 2006. [cited at p. 10]

[213] M. Vass, N. Allen, C.A. Shaffer, N. Ramakrishnan, L.T. Watson, and J.J. Tyson. The JigCell Model Builder and Run Manager. *Bioinformatics*, 20:3680–3681, 2004. [cited at p. 112]

[214] E.O. Voit and A.E.N. Ferreira. *Computational Analysis of Biochemical Systems:: a Practical Guide for Biochemists and Molecular Biologists*. Cambridge University Press, 2000. [cited at p. 9]

[215] C. von Mering, R. Krause, B. Snel, M. Cornell, S.G. Oliver, S. Fields, and P. Bork. Comparative assessment of large-scale data sets of protein- protein interactions. *Nature*, 417:399–403, 2002. [cited at p. 5]

[216] D.J. Watts and S.H. Strogatz. Collective dynamics of'small-world'networks. *Nature*, 393:409–10, 1998. [cited at p. 6]

[217] U. Wittig, M. Golebiewski, R. Kania, O. Krebs, S. Mir, A. Weidemann, S. Anstein, J. Saric, and I. Rojas. SABIO-RK: Integration and Curation of Reaction Kinetics Data. In *Proceedings of the 3rd International workshop on Data Integration in the Life Sciences 2006 (DILS'06)*. Hinxton, UK . *Lecture Notes in Bioinformatics*, 4075:94–103, 2006. [cited at p. 10]

[218] C.R. Yang, B.E. Shapiro, E.D. Mjolsness, and G.W. Hatfield. An enzyme mechanism language for the mathematical modeling of metabolic pathways. *Bioinformatics*, 21:774–780, 2005. [cited at p. 111]

[219] L. You, A. Hoonlor, and J. Yin. Modeling biological systems using Dynetica, a simulator of dynamic networks. *Bioinformatics*, 19:435–436, 2003. [cited at p. 112]

Appendices

Definition of concepts and acronyms used in text

Agent - an entity that is capable of perception and action. In the modeling approach presented in this work, each molecular species is treated as an individual agent, with a specific interaction menu and a resulting set of sub-species.

AMA - Approximated Mass Action; a non-spatial, kinetic ODE formalism describing molecular interactions with forms derived from an EMA description based on additional assumptions simplifying the conservation relationships (like negligibility of the enzyme-substrate complex) or dynamics of some variables in the system (like the steady state approximation for the enzyme substrate complex).

BioPAX - Biological Pathways Exchange; a XML-based format for biological pathway data exchange.

Combinatorial complexity - a property of MIN, where various combinations of multimolecular complexes with different properties can occur. It results from both multivalent binding (different structures of multimolecular assemblies) and multivalent enzymatic modification (different properties of these assemblies). For instance, a protein with n binding or covalent modification sites can have up to 2^n distinct states.

Deterministic models - models based on ODE.

EMA - Elementary Mass Action; a non-spatial, kinetic ODE formalism describing molecular interactions in terms of elementary association-dissociation

reactions in the form of law of mass action.

Hub - a species having a large number of connections to other species in a MIN.

Interaction - physical contact of molecules resulting in binding or enzymatic modification. In the modeling approach presented in this work, interactions are defined as any type of relation that can be measured with a biochemical rate constant, in particular: binding of two molecules, enzymatic modification, synthesis and degradation and are represented as links between interfaces.

Interface - in the modeling approach presented in this work it is a Boolean variable telling if a given interaction is taking place or not in a specific combination of molecules

Kinetic models - a class of dynamical models of MIN, where variables can have real-number (concentrations) or integer values (number of molecules) and change according to deterministic or probabilistic rules. Kinetic models rely on the law of mass action.

Law of mass action - an experimental chemical law, which states that the velocity of a reaction is proportional to the quantity of the reacting substances.

MIN - Molecular Interaction Network; a system of coupled biochemical reactions between proteins, genes and small molecules, especially binding and enzymatic reactions.

ODE - Ordinary Differential Equation; a relation that contains functions of only one independent variable, e.g. time, and one or more of its derivatives (i.e. an instantaneous rate of change) with respect to that variable.

Phenomena - in the modeling approach presented in this work, it is a special category of model components that allows incorporating meaningful biological functions such as 'cell mass' or 'cell division' into the model. Phenomena can have either a quantitative or a qualitative character and thus correspond to a real-valued or a Boolean system variable.

PSI MI - Proteomics Standards Initiative Molecular Interaction; a XML-based format for molecular interaction data exchange.

Rate constant - a coefficient expressing influence of several factors on the velocity of a chemical reaction, including temperature and the probability that reaction occurs because substrates are properly arranged in space and posses sufficient energy to form an intermediate complex.

Regulation - in the modeling approach presented in this work, it is any type of relation between two interfaces of the same species, especially activation or inhibition of one interaction by the other

SBML - Systems Biology Markup Language; a XML-based format for representing models of biochemical reaction networks.

Sensitivity - change in the numerical value of system's response relative to the change in the value of the signal, where any of system's parameters/variables or combination thereof can be treated as a signal or a response.

Species - any basic type of molecule (for example a type of protein) included in the MIN. In the modeling approach presented in this work, species are represented as collections of interfaces.

Sub-species - in the modeling approach presented in this work, it is any possible combination of the interface values of a given species.

Ultrasensitivity - a steep sigmoidal sensitivity pattern, where response changes from values close to 0 to values close to 1 (or reverse) within a very small range of signal values.

XML - Extensible Markup Language; a general-purpose markup language; i.e. a language combining text and information about the text in form of markup.