Chapter 12
Modeling and Model Simplification to Facilitate Biological Insights and Predictions

Olivia Eriksson and Jesper Tegnér

Abstract Mathematical dynamical models of intracellular signaling networks are continuously increasing in size and model complexity due in large part to the data explosion in biology. However, the sheer complexity of the relationship between state-variables through numerous parameters constitutes a significant barrier against obtaining insight into which parts of a model govern a certain read-out, and the uncertainty in model structure and especially model parameters is here a further complicating factor. To meet these two challenges of complexity and uncertainty, systematic construction of simplified models from complex models is a central area of investigation within systems biology as well as for personalized medicine. Model complexity makes the task of deriving predictions difficult in general and in particular when different read-outs depend on combinations of parameters, since exhaustive computer simulations are not sufficient for understanding nor feasible in practice. Construction of simplified models is therefore an important complementary approach to this end, while also facilitating the identifiability of over-parameterized models. Within this chapter we discuss different methods for model simplification, and we specifically summarize a recently developed simplification method based on an iterative "tearing, zooming and simplifying" approach. We also look into the modeling process in general. In the "tearing, zooming and simplifying" approach the original model is divided into modules (tearing), the modules are considered as input-output systems (zooming), which then are replaced by simplified transfer functions (simplifying). The idea behind the simplification is to utilize biological features such as modularity and robustness as well as abundance of typical dynamical behaviors in biology such as switch-like responses. The methodology is illustrated using a relatively complex model of the cell division cycle, where the resulting simplification corresponds to a piecewise linear system with delay, facilitating an understanding of

O. Eriksson (✉)
Department of Numerical Analysis and Computer Science, Stockholm University, Stockholm, Sweden
e-mail: olivia@mech.kth.se

J. Tegnér
Unit of Computational Medicine, Department of Medicine, Center for Molecular Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden
e-mail: jesper.tegner@ki.se

L. Geris and D. Gomez-Cabrero (eds.), *Uncertainty in Biology*,
Studies in Mechanobiology, Tissue Engineering and Biomaterials 17,
DOI 10.1007/978-3-319-21296-8_12

the underlying core dynamics and enabling the prediction of combinations of parameters that can change different model features like the size of the cell. Hence, the existence of biological organization principles enables a simplified description of intracellular dynamics.

Keywords Model simplification · Model reduction · Data integration · Dynamical models · Ordinary differential equations · Piecewise linear · Delay · Dynamical modules · Switch-like dynamics · Model decomposition · Lumping · Timescale separation · Sensitivity analysis · Identifiability · Tearing-zooming-simplifying

12.1 Introduction

The enormous increase in cellular experimental data of the last decades, resulting from the sequencing of the human genome [76, 81], microarray techniques [63], the FANTOM projects [43], 1000 genome project [74], ENCODE [75] and other high-throughput methods such as proteomics and metabolomics, has provided us with detailed lists on the constituents of the cell as well as their putative interactions. This has however not yet enabled an understanding of the functionality of the cell on the full genome-scale level. We have static descriptions on networks of interactions, in the same way as we have road maps, but still we do not know much about traffic flow [46]. In order to investigate and describe the dynamical behaviour of the huge system that a cell constitutes, we do not only need experimental developments but also improved computational and mathematical methods [9, 21].

The structure and dynamics of cellular networks are inherently complex, containing a large lattice of redundancy and intertwined feedback loops, where the interactions can be described by detailed biochemical reactions. Imagine the complexity of a single signaling pathway, such as the cell cycle, and then put together all possible pathways of the cell and the resulting picture becomes overwhelming. There are however simplifying circumstances in this complexity. Cells seem to have a modular organization in space and time, with a sparse number of interactions between constituents. Cells often display quite simple functions (e.g. on/off circuits) and have a robust behaviour. The key idea of this chapter is that provided that we could use these simple levels of regulation hidden in the mesh of details, then maybe we could get a step closer to retrieving descriptions of the functionality of whole cells.

In order to give precise unambiguous descriptions, and quantitative predictions, mathematically formalized relationships and parameters are needed [39, 77], and dynamical cellular network models, describing the time evolution of e.g. protein concentrations, are the focus of this chapter. A few examples of dynamical models describing different cellular pathways and phenomena include [12, 33, 50, 77, 82]. The size of these models is, on the full genome-scale, quite modest, describing only a small subset of the proteins or genes of the cell e.g. a signaling pathway or regulatory module. However, although being relatively small models they display large complexity, with intertwined feedback-loops and functional redundancy.

A future vision, which is becoming increasingly realistic [42], is to construct dynamical models of whole cells. One approach towards this goal is to combine small mechanistically detailed models into larger and larger contexts. Such combined models tend to have a large number of parameters and complicated rate laws. Within the emerging complexity of these models it can be a difficult task to figure out the functional relationships of the model constituents, e.g. which parts that are essential for a certain read-out. It is therefore challenging to derive predictions on important features without performing extensive *brute-force* computer simulations. This problem becomes even more urgent as most models of biochemical systems today are highly over-parameterized with respect to available in vivo data [65], i.e. there exists a very large number of parameter combinations that give an equally good agreement with the data.

One important tool to deal with larger systems is *model simplification or reduction*. A model simplification process can illuminate the dynamical mechanism behind certain behaviour of the system and identify functional relationships between variables that are not obvious from inspection of a large model. Furthermore, a simplified model could remove over-parameterization and thereby generate an efficient description of the system with improved identifiability.

In this chapter we review different methods for simplifying complicated models of intracellular signaling cascades. We will also discuss in detail an important biological signaling cascade within the cell, namely the network regulating the cell cycle. This well characterized biological system is a useful test case to develop a model simplification process based on the identification, characterization and simplification of dynamical modules, here resulting in a piecewise linear model with delay. Before we go into model simplification and reduction in general and this specific example in detail we will first discuss how intracellular models actually are constructed.

12.2 Data Integration, Experimental Setting and Model Uncertainty

To develop models of intracellular networks data from different sources have to be integrated. The direction of data integration is often described by a *top–down* or *bottom–up* approach [10]. The top–down approach considers data from the whole cell and by an iterative cycle between experiments and modeling retrieves a better and better model resolution of the system. In contrast the bottom–up approach considers detailed descriptions of subparts of the cell and combine these into larger and larger models. The first approach describes the system more phenomenologically while the other line of investigation is geared towards a more mechanistic objective.

The genome sequencing projects [76, 81] and other high-throughput experimental techniques are examples of a top–down approach. Together with computational methods within bioinformatics for combining and analyzing several sources of data these experiments have provided us with a detailed *parts-list* of genes, proteins and

other constituents within the cell. We will denote this type of data as *genome-scale* data. During the last decades we have witnessed how these parts have been combined to large *static networks* of possible interactions, e.g. protein–protein interaction [79], transcriptional regulation [49] and metabolic reaction [20] networks, retrieved from experiments or by computational methods. Finally the *activity* of the parts can also be recorded by e.g. microarrays [63].

In contrast, bottom–up approaches depend on characterizing the interactions between individual components of the system and then integrating them into a larger reaction network [10]. Detailed kinetic and physiochemical properties of the interactions are obtained from experiments, where the experimental setting differ from each other in their level of simplification as compared to native conditions. At one end there are test tube experiments with purified components in optimized conditions, whereas at the other end experiments on living cells. We will denote these settings the dissociated versus the embedded setting, respectively. An example of the dissociated setting is the estimation of the parameters K_M (the Michaelis Menten constant) and k_{cat} (catalytic rate) of an enzyme, using purified components (e.g. proteins) in a test tube in optimized conditions for their function [1]. Examples of the embedded setting include the monitoring of time-course data for protein interaction markers and/or morphological cellular features obtained from the living system under perturbation [62].

Data from several different dissociated characterizations has been combined, in a bottom–up approach, to try to reconstruct the dynamical behavior of the larger system. This has shown some success in the case of modeling of the glycolytic pathway of unicellular organisms like Saccharomyces cerevisiae [66]. The efforts to model this universally conserved pathway in this way began several years ago using each enzymes own optimal conditions for its characterization (and thereby different test tube conditions for different enzymes despite some of them sharing the same intracellular environment). Just recently standardized conditions that resemble the intracellular milieu has been used to characterize all enzymes in the pathway resulting in a more accurate model [1]. While a similar dissociated characterization, bottom–up, approach has been proposed for components of signaling cascades [41] there has not been any systematic effort in this direction. Unlike glycolytic enzymes in unicellular organisms, the components of intracellular signaling cascades have a higher degree of compartmentalization and more interacting partners so that even an approximate recreation of physiological conditions in vitro is far more demanding [32]. Thus, there are relatively few interactions of signaling cascades characterized with purified components in vitro. Besides the limited physiological relevance of the experimental conditions used in these characterizations, most of the resulting estimates correspond to steady state conditions and parameters, despite the fact that fast transient signaling is occurring in many systems. While dissociated characterization is appropriate for ranking the effects of mutations, as well as developing inhibitors with pharmaceutical applications, and studying mechanisms of enzyme action, its usefulness for modeling the dynamics of signaling pathways is limited. The parameter values estimated through dissociated characterization should be taken in most cases as soft constrains [12]. The embedded characterization is becoming prevalent

with the growing trend in the development of cell-based measurement methods. This development is driven by progress in protein labeling with genetically encoded markers which allows to track interactions and compartmentalization in real time [53], high-resolution and two photon microscopy which makes it possible to track discrete events in small cellular compartments with reduced photo damage [84] and high-content high-throughput techniques which enables monitoring the effect of several conditions on a few protein markers and second messengers in a single run [27]. Despite all this progress in data acquisition from cell-based experiments, the amount of available phenotypic data is still far from sufficient for fully constraining the model so that every parameter is identified. More information about experimental data for modeling can be found in [64].

12.3 Cellular Organization and Model Structure

There are different rationales driving modeling approaches. A *predictive* model is mainly used to make predictions about the system while a *descriptive* model stores knowledge. Within technical applications a predictive model often does not need to have any structural similarities with the real system as long as it can anticipate the behavior of the system, thus it is essentially viewed as a black-box with respect to mechanisms. However, within biology, models are often of a more descriptive nature, utilized to precisely collect knowledge about the system, and as a picture to discuss around, even though the final goal may well be predictions. Biological models are thus constructed to have a structural correspondence with the real system. Before we go into details on different model formalisms and simplification methods in the coming sections, we will here first discuss the structural and functional organization of cellular networks.

As described above, high-throughput genome-scale experiments have provided huge amounts of data concerning the cellular constituents, like genes and proteins, which can be organized into different static networks. The topology of such networks has been analyzed by graph theoretical methods [7] and intriguing similarities have been found between different biological systems. Numerous networks display a power-law distribution in the number of connections (edges) a node can have, i.e. most nodes have only a few edges to other nodes at the same time as there exist a few nodes having a large number of edges, denoted *hubs*. Many networks have also been found to have a relatively high clustering-coefficient, indicating the existence of groups of highly interconnected nodes, *topological modules*. Cellular functions have long been suggested to be carried out in a highly modular manner [7, 39, 78], where modular refers to a group of molecules, physically and/or functionally linked together, which perform a distinct function. Most studies so far concern static topological modules. Identifying topological modules is however a non-trivial task. The fact that clustering and hubs coexist indicates that topological modules are not independent and well isolated from each other, but rather that the network has a hierarchical organization [7]. Since the cell is in fact a dynamical system, *dynamical*

modules could have a closer correspondence to the functional or biological properties of the system. Tyson et al [78] have reviewed a set of different dynamical circuits typical for cellular networks.

There is one specific type of circuit behavior we would like to consider in a bit more in detail here, namely, *switch-like dynamics*. This is a recurrent phenomenon in cellular networks and we will illustrate how to exploit this feature in the 'tearing-zooming-simplifying' example provided at the end of this chapter, effectively defining a simplified model. Switch-like dynamics can been found in many biological systems [6, 36, 38, 40], and is often modeled using a steep sigmoid function e.g. a Hill-function with a high Hill coefficient. Biologically, this can be due to cooperative processes, positive feedback, or enzymes operating near saturation [28]. Two explanations have been suggested by James Ferrell as to why a steep sigmoid input/output response should be useful for the cell [28]. The resting state of a cell could be near the upstroke of the input-output curve, or it could be far away. In the first case a small change in input can give a large change in output, an amplification of the signal. In the second case the system would filter out small stimuli and allow the cell to respond decisively to stimuli of a sufficient magnitude. The second case could therefore support cellular robustness against noise.

12.4 Modeling Formalisms and Choice of Model Detail

An important part in the art of modeling is the decision of modeling formalism. This choice is strongly influenced by the nature of the question and the data at hand, and therefore the art in choosing a suitable abstraction or representation of the system at hand. In order to give precise unambiguous descriptions, and quantitative predictions, mathematically formalized relationships and parameters are often needed [39, 77]. In this section we will mainly focus on dynamical models, but in the interest of completeness we will first briefly discuss static models.

12.4.1 Static Models

As described earlier, high-throughput techniques and computational methods have produced an incredible quantity of biological interaction data. Data that is conveniently represented by large networks, or *graphs*. Such graphs consist of *nodes*, representing e.g. genes, proteins or metabolites, connected by *edges*, representing interactions or other relationships. These graphs can be *directed* as in the case of gene-regulatory networks, where the product of one gene is regulating another gene or *undirected* as is the case for protein–protein networks, when only a binding possibility is recorded. As mentioned earlier, the topology of different biological networks have been analyzed by graph theoretical methods and similarities between different types of networks have been discovered. It has, however, become increasingly clear

that these static networks only describe interaction possibilities, and that not all of the edges are active at a certain time and in a certain context (cellular location, external signal) [35, 54]. The data of possible interactions therefore need to be combined with node activity in order to approach a dynamical description of the cellular network.

12.4.2 Dynamical Models

Depending on the purpose of the modeling, the approximations that can be made on the system, and the experimental data available, there are several different mathematical formalisms to choose from when it comes to *dynamical* models of cellular networks [13, 16]. Examples include *Boolean models*, *non-linear ordinary differential equations*, *piecewise linear differential equations*, *partial differential equations*, *delay differential equations*, *stochastic master equations* and *rule-based formalism*. We here describe ordinary differential equation (ODE) and piecewise linear (PL) differential equation models in more detail, and also touch upon delayed differential equation (DDE) models. We look at ODE models because this is the model formalism most widely used within biochemical modeling, and DDE and PL differential equation models since the simplified model in the example at the end of this chapter is a PL model with delay. More information about modeling, and then especially modeling under uncertainty, can be found in [45].

Non-linear ordinary differential equations Modeling cellular networks by ordinary differential equations (ODE:s) uses non-negative, time-dependent variables to describe e.g. the concentrations of proteins or other molecules. Interactions between molecules correspond to functional or differential relations between the variables. The concentrations are thus described by *rate equations*, describing the rate of production of a component of the system as a function of this and other components of the system. We use $\dot{x}$ to denote time-derivative. The rate equations correspond to

$$\dot{x_i} = f_i(\mathbf{x}, \mathbf{u}), \qquad 1 \leq i \leq n, \tag{12.1}$$

where $\mathbf{x} = (x_1, \ldots, x_n) \geq \mathbf{0}$ is a vector of protein or other molecule concentrations internal to the system, $\mathbf{u} = (u_1, \ldots, u_m)$ is a vector of external input signals, e.g. nutrients and f_i a function.

The above formalism can be extended with discrete time-delays, $\mathbf{x}_\tau = (\mathbf{x}(t - \tau_1), \ldots, \mathbf{x}(t-\tau_p))$, where τ_i are positive constants, to deal with e.g. the time required for transcription and translation. The system is thus transformed into a DDE model.

Piecewise linear differential equations Several cellular networks have been modelled by piecewise linear differential equations, e.g. [17, 29, 30]. We here follow the notation of Gonçalves [31]. Piecewise linear systems (PLS) are constructed from a set of affine linear systems,

$$\dot{\mathbf{x}} = \mathbf{A}_\alpha \mathbf{x} + \mathbf{B}_\alpha \tag{12.2}$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A}_\alpha$ an $n \times n$ matrix, $\mathbf{B}_\alpha$ an $n \times 1$ input vector, and α is a *switching rule*,

$$\alpha(\mathbf{x}) \in \{1, \ldots, M\} \tag{12.3}$$

which describes when to switch between the linear systems. The switching rule depends on the present state $\mathbf{x}(t)$ and might also depend on earlier states e.g. $\mathbf{x}(t-\tau)$. We denote a switching rule that only depends on the present state as *memoryless*. A solution to (12.2)–(12.3) are functions $(\mathbf{x}, \alpha)$, satisfying (12.2)–(12.3), where α is piecewise constant. A *switching time* of a solution $(\mathbf{x}, \alpha)$ is a time t where $\alpha(t)$ is discontinuous. A trajectory *switches* at a time t if t is a switching time. Switching occurs at *switching surfaces* in the state space of $\mathbf{x}$. If the switching rule is memoryless and consists of linear inequalities, then these surfaces are hyperplanes of dimension $n-1$,

$$\mathbf{S}_j = \{\mathbf{x} | \mathbf{C}_j \mathbf{x} + d_j = 0\} \tag{12.4}$$

where $\mathbf{C}_j$ is a $1 \times n$ vector and $j \in \{1, \ldots, N\}$. For a PLS let us define a partition of the state space, where a separate linear system is used, as

$$\mathbf{X}_i = \{\mathbf{x} | \alpha(\mathbf{x}) = i\} \tag{12.5}$$

for $i = \{1, \ldots, M\}$. When the switching rule has no memory, $\mathbf{X}_i \cap \mathbf{X}_j = \emptyset, i \neq j$, and in each partition the dynamics is given by the linear system $\dot{\mathbf{x}} = \mathbf{A}_i \mathbf{x} + \mathbf{B}_i$. A phase portrait of a two dimensional piecewise linear system is displayed in Fig. 12.1.

Switching rule with memory If the switching rule has a memory, i.e. it does not only depend on $\mathbf{x}(t)$, but also earlier states, e.g. $\mathbf{x}(t-\tau)$, then the intersection of different $\mathbf{X}_i$ might not give the empty set. One example of this is given in [23], where the switching rule not only depends on the present state but also on a past state $\alpha = \alpha(\mathbf{x}(t), \mathbf{x}(t-\tau))$. In [23], at each phase point $\mathbf{x}$, one out of two linear systems can be used, depending on the value of $\mathbf{x}(t-\tau)$.

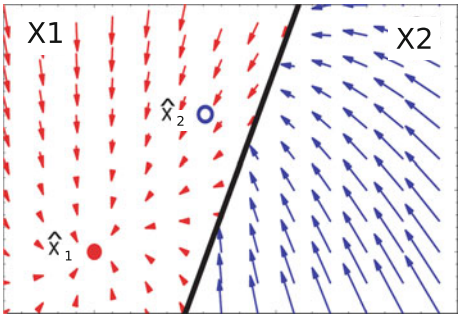

Fig. 12.1 A phase portrait of a piecewise linear system consisting of two linear systems defined in **X1** and **X2** respectively. The fixed points $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ of the respective linear systems are also indicated. From [23]

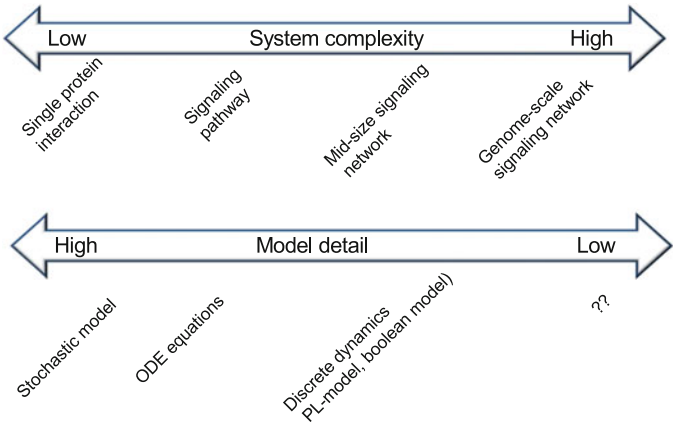

Fig. 12.2 Model detail is determined by system complexity. A single protein interaction can be modelled in great detail using stochastic models. For larger and thus more complex biological systems less detailed models are more informative. Dynamical models on a genome scale are not yet feasible. Modified from [9]

12.4.2.1 Choice of Model Detail

As our focus shifts from modeling single pathways to increasingly complex cellular networks, the computational methodology and formalism used must be carefully considered. Extrapolating the traditional ODE model used for modeling single pathways to whole-cell systems would make the model prohibitively complicated. This is discussed in [9], where it is suggested that finding a "course-grained" level of model description where the molecular details are left out when possible and focusing on the system behaviour could be one strategy to solve this problem, see Fig. 12.2. This theme is also elaborated by de Jong [16] who suggests that whole-cell models could be organized in a hierarchical manner, based on the inherent modularity of the cell. On different levels of abstraction different modeling formalisms could be used, thus on a higher level a more "course-grained" formalism would be appropriate.

12.5 Model Simplification and Reduction

Model simplification and reduction in order to enable analysis is not a new concept within biology. The most famous example is perhaps the Michaelis-Menten equation were a separation in time-scales justifies a "quasi-steady-state" approximation. Throughout this chapter, we use the term model simplification as a more loose description that a model is simplified in some sense, whereas the term model reduction to describe simplifications where the number of degrees of freedom in the model is reduced, while the model formalism remains the same.

Which method to use when simplifying a particular full-scale model depends on the objective of the simplification process and the equation structure of the full-scale model. Before a model simplification process is performed it is useful to decide on which features of the original model to be retained, in [68] named *reduction target* and in [65] denoted as *reference data*. The simplified model must of course be consistent with at least some aspect of the full-scale model. It can, however, in most cases, by definition, not reproduce all possible dynamics of the full model under all possible circumstances. This was investigated in [26], where it was found, that reduced models that reproduced predictions of the full model for a particular set of parameters loosed their predictive capacity when parameters were varied over two-fold ranges.

One can note that the model simplification process, in the same way as ordinary modeling, is an approach to essentially obtain insights on how the underlying system works. Model simplification can therefore be seen as the "modeling of an already existing model". The proper approximations made during the model simplification process could therefore be instrumental in learning about the underlying processes. As was noted earlier, biological systems are *sparse* and seem to have a *modular* structure [39], features that could aid a reduced model description.

There are a number of reduced models described for a variety of biological systems and some early examples include [48, 67, 85]. The model reduction procedures often include *ad hoc* components, which require that the modeler is intimately familiar with the dynamics of the original system. Attempts of more systematic model reduction approaches have been performed e.g. [34, 47, 65, 68]. However, also such approaches are often only applicable to a small subset of biological systems of a particular structure (e.g. systems having one ubiquitous variable) and/or dynamical behavior (e.g. hopf bifurcation). Hence, there is a need for development of more general methods that would enable systematic simplification and evaluation of complex non-linear systems.

12.5.1 Model Simplification and Reduction Approaches

We here give examples of some approaches towards model simplification and reduction. This list is by no means exhaustive. A review can be found in [58]. Some model simplification strategies include a first step of decomposing the system into subsystems, whereas others remove or simplify individual parts (e.g. equations or reactions) one-by-one. As described earlier there is also a difference between methods that remains within the same formalism and methods that translate the model into a new formalism. Examples of the first case include lumping of variables, separation of timescales, and removal of variables based on sensitivity analysis or identifiability, whereas examples of the second case include boolean approximations, introduction of explicit time delay, and hybrid approaches; as the simplified model described at the end of this chapter, a delayed piecewise linear approximation of an ordinary differential model.

Module based model decomposition Cellular functions have, as was described earlier, long been suggested to be carried out in a modular manner [7, 39, 78], and different methodological approaches have tried to utilize this for model decomposition, in order to retrieve model subsystems that can be easier analyzed then the original system. One example can be found in [60, 61] where the original system is divided into subsystems or modules of low *retroactivity*. Loosely defined, retroactivity describes the phenomenon in which a downstream event affects upstream reactions. This concept emerged in the field of electric engineering. When the output of one electric unit is connected as input to another electric unit, this can affect the first electric unit retroactively, for example by draining electrons too fast [4]. A unit would be without retroactive effects if both input and output are unidirectional. This means that the behavior of the unit only depends on its input, and that 'connecting' it to another unit does not change its input-output behavior. The behaviour of retroactive-free modules can be studied uncoupled with the system and analyzed by systems theory's tools. Kinetic insulation is another concept related to modularity and a mechanism suggested to increase "isolation" between modules by the use of different timescales [18] and segregate between different signals that are using common pathway components [8].

Combining variables, or lumping refers to the process of reducing the number of dimensions of a system by merging states (e.g. protein concentrations) together. This approach is well suited for biological of the often occurring modular structure, and as the remaining states can have a biological meaning, like sums of protein concentrations. This can be illustrated by the following system

$$S \longrightarrow I_1 \longrightarrow P \tag{12.6}$$

$$S \longrightarrow I_2 \longrightarrow P \tag{12.7}$$

where S is a substrate that turns into P through two different intermediates I_1 or I_2. The species I_1 and I_2 can be lumped together to the new pseudo-species $I = I_1 + I_2$, resulting in

$$S \longrightarrow I \longrightarrow P. \tag{12.8}$$

Lumping of variables leads to a simplification in terms of the number of states and reactions of the systems but this might come to the cost of a higher complexity of the remaining rate equations. A description of lumping procedures can also be found in [58]. Lumping can be divided into two categories, in *proper lumping* each of the species of the original model contributes only to one of the lumps of the reduced model, whereas in *improper lumping* species can contribute to more than one lump.

The choice of which variables to be lumped together is often decided on an intuitive basis corresponding to an understanding of the specifics of the system. An example of a more systematic method is the use of *equivalent potentials* [44, 68] used to reduce neuron models. Here the special structure of these neuron models enabled the design of an automated method. This method first uses a nonlinear transformation to

the state variables, in order to find similarities between them. Finally, the variables are sorted into groups based on these similarities. Next, each group is replaced by a new variable approximating the function of the original state variables. In [72, 73] the back-translation properties of a reduction process for biochemical models are emphasized, and they present a method based on lumping of clusters of fast variables, where it is possible to map from reduced model variables and parameters back to original model variables and parameters. A systematic method for proper lumping of the systems state is also described in [19] which is based on a search through all possible combinations of lumps. The combinatorial complexity of the problem is bypassed through a heuristic, greedy algorithm.

Separation of timescales, is another method to reduce the complexity of models. As a rule, biological processes occur over a broad range of time-scales, from milliseconds (e.g. phosphorylation reactions), minutes (e.g. transcription) to hours (e.g. cell cycle), days and years, and the principle of time-scale separation has been widely used within biology. One example is the *quasi-steady-state approximation* (QSS), which is used within enzyme kinetics to derive the well known *Michalis-Menten* equation. In QSS, the variables x are decomposed into two blocks x_f, fast variables and x_s slow variables: $\frac{d}{dt}\begin{pmatrix} x_f \\ x_s \end{pmatrix} = \begin{pmatrix} F_f(x_f, x) \\ F_s(x_f, x_s) \end{pmatrix}$. The reactions including the fast variable x_f are approximated as to be instantaneous compared to the reactions including the slow variable x_s. For given fast species we can simply set $\frac{dx_f}{dt} = 0$, which results in an algebraic relation

$$F_f(x_f, x_s) = 0.$$

If the functions F_f is nice, the Implicit Function Theorem can be applied, that is, there is a unique solution $x_f = G(x_s)$. Substituting it into the original system yields a lower dimension system $\frac{dx_s}{dt} = F_s(G(x_s), x_s)$. This means that we can study the lower dimensional system on the slow manifold $F_f(x_f, x_s) = 0$. The generalization of this technique is difficult since it is no trivial task to divide variables into fast and slow. There are also conditions that have to be satisfied for the method to work. In [37] a nice extension to the QSS is described based on the zero-derivative principle. For a more detailed description on timescale separation see [58] or [25].

Introduction of explicit time delay, is used in a few studies as a mean to simplify biological models e.g. [22, 69]. Consecutive interactions in an ODE model often give rise to a time delay in the system. In the above studies, intermediate variables are removed and instead represented by an explicit delay, reducing the number of variables, and turning the system into a system of delayed differential equations (DDE). It can be noted that DDE systems, however, being infinite dimensional, are in general harder to analyze than ODE systems.

Aside from model reduction, the question whether to use explicit delay or slow intermediate variables, when modeling biological systems has been discussed [55]. Oscillations in biological systems are often assumed to be due to a delayed feedback loop and the delay is often modeled by slow intermediary variables. This may have

the consequence that processes such as transcription and translation are assumed to be instantaneous. In [55] it is shown in a model of the oscillatory expression of Hes1, p53 and NF-κB that no intermediate variables are needed to get oscillations if an explicit transcriptional delay is introduced. This is a principally important result, since it is preferable to include an anonymous delay in the system (that of course could be due to an intermediate variable) than a variable that correspond to a protein that might not exist. If a truly delayed process is modeled as instantaneous (even though slow) this can result in erroneous parameter estimates.

Sensitivity analysis based methods A sensitivity analysis (SA) investigates how the model output depends on the model parameters and can be performed through a *local* SA, describing infinitesimal changes around a nominal point in parameter space, or through a *global* SA, investigating larger parts of the parameter space, often by statistical analysis. More information about sensitivity analysis can be found in [80]. In model reduction SA is often used through two steps. First all parts (e.g. terms, reactions, variables) of a model are ranked according to their influence on the model output. In the next step parts that are considered to be unimportant for the considered behaviour are removed (corresponding to low ranked parts) [11, 58]. This is not a trivial task, however, since individual parts can have a low sensitivity, but combinations of parts can be crucial. Also for a local sensitivity analysis the result depend on the nominal parameter value that the sensitivity analysis starts off from. Therefore, an individual ranking list can only be considered as a guide. Also notable is that the original model structure must be considered during the removal of variables, so that the remaining model is biochemically consistent. One example of model reduction based on SA can be found in [51], where SA together with flux analysis and principal component analysis is used to reduce a model of epidermal growth factor (EGF) mediated signaling and trafficking. SA is also used in [5] to guide the order of removal of parameters in the model.

Identifiability. In [65] the concept of unidentifiability is used to reduce rate expressions. The rationale is that an unidentifiable rate expression has more than one parameter set that can describe data equally well and that simpler expressions therefore could be used. In the systematic method of [65], the rate expressions have to be in rational form, i.e. a fraction between two polynomials. The reduction is next performed in a reaction-wise manner, so that the complexity of the individual rate expressions is reduced, while the structure of the cellular network is conserved. The reduction is performed in relation to a *reference data* set, corresponding to in silico simulations of the original full-scale model, and parts of the reaction rates that are unidentifiable with respect to this data are removed. By this method terms of the reaction rates that are less important for the model behaviour, as represented by the reference data, are removed. This procedure thereby identifies the functionally important interactions. Another important study exploits the profile likelihood to detect structural and practical non-identifiabilities and suggests the use of this methodology for model reduction [59]. We also refer to [45] where issues related to identifiablity are discussed.

Topological filtering and viable parameter spaces. Uncertainties in model structure often makes it necessary to evaluate more than one model topology, where different topologies correspond to different biological hypothesis (see also [70]). In a recent study [71] a "supermodel" is constructed that incorporates all hypothetical mechanisms concerning the underlying biological system. This includes the region of the parameter space of this supermodel that is compatible with the experimental data, called the viable space. The exploration of such a large multidimensional space is done by a procedure that combines an initial coarse grained global sampling of the viable space with a subsequent finer grained exploration [86]. This supermodel is then reduced to a set of new models, with reduced topology, by an iterative procedure where parameters are eliminated.

Another related study [5] is also based on the exploration of a parameter region where the model yields some required output, here called admissible region. Based on the shape of this region, parameters and variables are removed and lumped as long as the dynamical behavior of some target species are preserved. The authors further use sensitivity analysis as guidance for ordering the parameters during the reduction procedure.

Change of modeling formalism. The decision on which modeling formalism to use when describing a system, includes a decision on model detail [9, 16] and a decision on which underlying processes that are important to include in a model. A stochastic model representing all molecules of the system is more detailed than an ODE model which assumes that molecular concentrations are sufficiently large in order to be approximated with a continuous and deterministic description. An ODE model in turn is more detailed than a boolean model. Therefore, as an example, can a boolean model be seen as a reduction of a certain class of ODE models where the variables can be described as being on or off. This technique was used by Albert and co-workers to identify conditions for robustness in a large kinetic ODE model of the Drosophila segmentation [2], and by Davidich and Bornholdt to reduce an ODE model of the cell cycle [15]. Other ways to reduce model complexity by a change of formalism is to transform a nonlinear ODE model to a piecewise linear ODE model [22] or a hybrid model [34].

12.5.2 "Tearing, Zooming and Simplifying"—an Example of a Model Simplification Procedure

We close this chapter by describing a specific example of a model simplification procedure developed in [23]. The core concept is to utilize biological properties such as modularity in order to identify a simplified description of the system and secondly to develop this methodology using a well characterized biological system. Hence, we start off from a medium sized non-linear ODE-model of the cell division cycle which has been pioneered by the work of Tyson and Novak [57]. The method effectively produces a smaller (in terms of variables and parameters) model in the

mathematical form of a delayed piecewise linear (DPL) system. The fact that the simplified model is piecewise linear facilitates the analysis and enables detailed predictions on parameter relationships that regulate model output such as the cell size. The tearing-zooming-simplifying approach consists of three steps:

1. tearing or subdividing the original model into subsystems (modules),
2. zooming out and characterizing the modules by input/output transfer functions
3. replacing the modules by simplified descriptions.

This is an iterative process where e.g. information from the second step can suggest a new subdivision in the first step. This simplification method is akin to the system theoretical approach to model large engineering systems by tearing and zooming, see e.g. [14, 83] . The concept of zooming has also been used recently in [72, 73].

The variables of the cell cycle model correspond to protein concentrations, and before we go into the details of the tearing-zooming-simplifying approach we will first shortly describe the cell division cycle.

12.5.2.1 The Cell Division Cycle

Cells reproduce by duplicating their contents and then dividing into two.This process contains two parts, the chromosomic cycle and the cytoplasmic cycle [3]. The chromosomic cycle consists of the exact duplication of DNA, *DNA synthesis*, and the subsequent separation of the two copies, *mitosis*. In parallel to this process, during the cytoplasmic cycle, all other constituents are doubled in quantity, the cell grows and the whole cell is divided into two cells. There needs to be a coordination between, and interactive control of, the chromosomic and cytoplasmic cycles. This is achieved by the cell-cycle genes and proteins. Failure in regulation of the cell cycle can result in uncontrolled cell growth and the initiation of cancer.

Several mathematical models have been constructed to account for this system e.g. the pioneering work encoded in the experimentally constrained models of Novak and Tyson [77], one of which [57] is used in this model simplification example. During the eukaryotic cell cycle in fission yeast, the cell grows, DNA is replicated (*S-phase*), and divided into two daughter cells (*M-phase*). Between the S-phase and the M-phase there are also two gap-phases, referred to as *G1* and *G2*.

12.5.2.2 Tearing or Subdividing the Original Model into Subsystems

The final goal of the simplification procedure is to find more or less isolated units (modules), with a dynamical behavior that can be replaced by simpler descriptions. The first step in achieving this consists of dividing the original model into subsystems, based on the topology of the interactions. These potential modules—in the first iteration—consists of a subset of original variables, and have a corresponding set of coupled ODEs. It is important that the coupling between the ODEs within the module is intact, and not torn apart by the subdivision. Therefore a graph describing

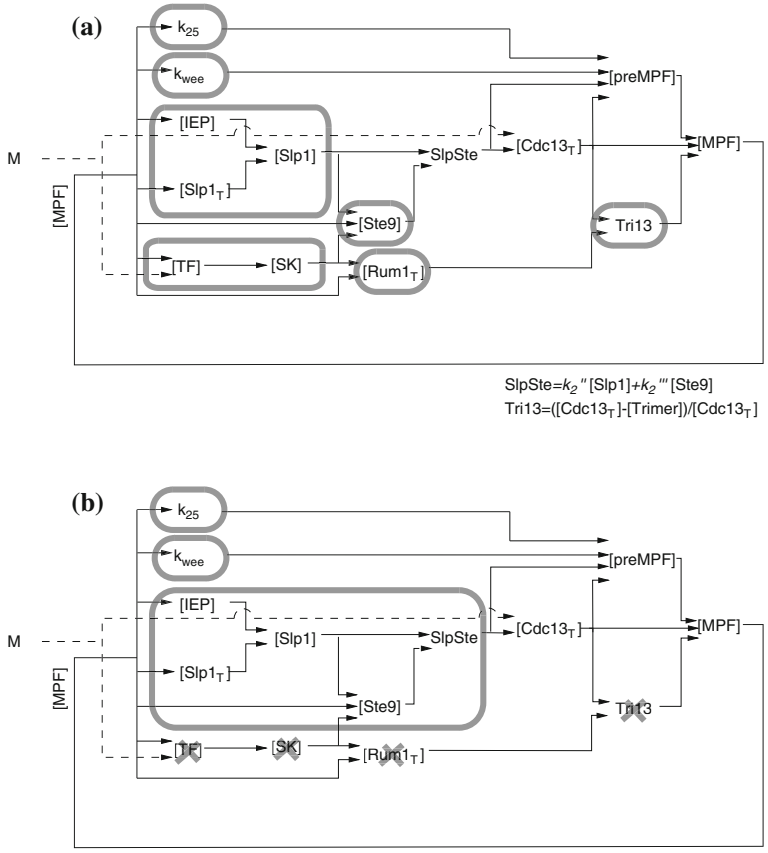

Fig. 12.3 Graph describing the coupling between ODEs, and subdivision into modules (from [23]). A node in this graph corresponds to a variable of the original model, and there is an edge from node j to i if j directly affects i (i.e. is on the *right hand side* of a differential equation defining i). This graph was used to divide the system into potential modules. **a** Full DPL-model, subdivision into switching modules when $M \geq 0$. **b** Small DPL-model, subdivision into switching modules when $M > 0.8$. Some of the variables can be replaced by constant parameters (indicated by crosses), when $M > 0.8$

the coupling structure of the ODEs is constructed, see Fig. 12.3. The graph illustrates how the dynamics of different variables depend (directly) on other variables in the model. Interestingly, this kind of graphical approach for decomposition was recently utilized in a study on observability of biochemical systems [52].

Putative modules are chosen so that (i) each module has one output-variable, and (ii) all nodes within the module are connected to the output by a connected path. Some of the nodes within a module correspond to ODEs which depend on variables coming from outside the module, denoted input-variables. In addition to defining modules, the graph is also useful for representing the coupling relevant

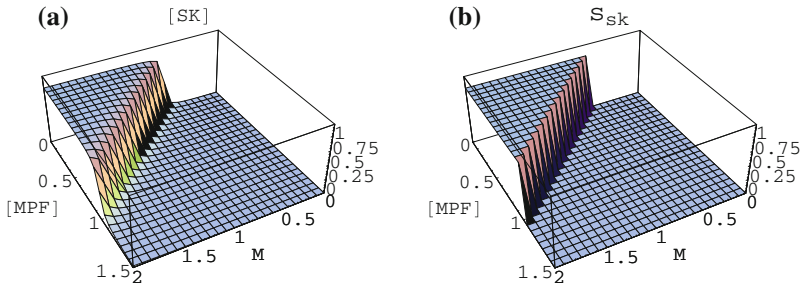

Fig. 12.4 Steady-state response and step function approximation. **a** The steady state response of the output $[SK]$ (concentration of Starter Kinase) to the inputs M (*Mass*) and $[MPF]$ (Mitosis Promoting Factor). **b** The same function as in **a** but approximated by the step function S_{sk}. From [23], for a colored version see the online version of the book

for elucidating dynamical motifs. In the case of the cell cycle model, the feedback structure within the network and the central role of the protein complex $[MPF]$ is made more transparent by our procedures (Fig. 12.3).

12.5.2.3 Zooming Out and Characterizing the Module Dynamics by Input/output Functions

In the next step the dynamical behavior of the potential modules is analyzed by investigating the input/output relationship. This is performed for each module in isolation by two complementary procedures. First by describing the *steady-state response* of the output in response to different inputs (for an example see Fig. 12.4), and secondly to consider the *response time*, i.e. the time it takes for the module to reach steady state after a significant change of input.

Based on such a characterization and the additional constraint that we avoid modules to be overlapping, leads to the final modules being selected so that (including the earlier defined criteria) (i) each module has one output-variable, (ii) all variables within the module is connected to the output by a connected path, (iii) the modules are non-overlapping, and (iv) the output has a switching input/output behaviour. By switching input/output behaviour we mean a steep, sigmoid-like, response curve (e.g. Fig. 12.4). Most of the variables in this cell cycle example have a dynamics based on *Michaelis-Menten* or *Hill* kinetics, and modules can therefore readily be identified which have a steep sigmoid almost step-like steady-state dependency on the input.

During this process we also remove variables for which the signal response curve are constant or almost constant, by re-defining them into constant parameters. In the cell cycle example this can be done since some of the variables are not "active" in the regime we are interested in, i.e. when the cell mass is larger than 0.8 ($M > 0.8$). In Fig. 12.3 the final modules are illustrated, and in Fig. 12.4 an example of a input-output curve can be seen.

12.5.2.4 Replacing the Modules by Simplified Descriptions

In the final step, the modules are replaced by step functions (Fig. 12.4) and a time delay corresponding to the measured response time (for details see supplementary material of [23]). This transforms the original system into a piecewise linear system (with delay). Not all variables are included into modules, some are removed as described in the previous section, and some become state-variables of the new simplified system. It can be noted that it is not obvious that replacing the modules with step functions would turn the remaining system into a DPL-model. This is possible in this case due the special form of the remaining ODEs of the original model. Whether this is true also for other models remains to be explored.

12.5.2.5 Elucidating the Core Dynamics by a Simplified Model and Making Predictions

The simplified (DPL) model of the cell cycle emphasizes some important features of the system. In the DPL-model, one normal division of the cell, corresponds to the subsequent move between four linear subsystems (Fig. 12.5). Interestingly, these

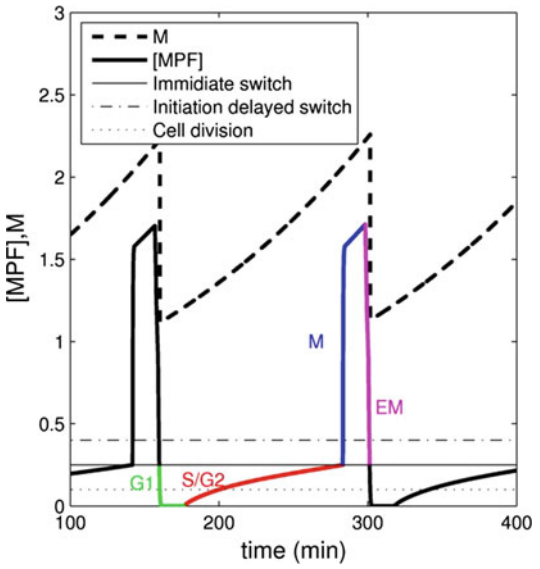

Fig. 12.5 Numerical simulation of the of the DPL-model, showing the time evolution of the cell mass (M), as well as the concentration of the protein complex $[MPF]$ (Mitosis Promoting Factor). During different parts of the cell division cycle, different linear systems are used, as indicated on the time course of the variable $[MPF](t)$ with green, *red*, *blue* and *magenta*. The linear systems correspond to the four cell cycle phases G1, S/G2, M and EM, where EM is the ending of Mitosis. Which linear system that is used at a certain time t depends on $[MPF](t)$ and $[MPF](t-\tau)$ as detailed in [24]. From [24]

subsystems can in turn be directly mapped to different phases of the cell cycle (G1, S/G2, Mitoses, and ending of Mitoses). This means that during each cell cycle phase only one (linear) subsystem is active, corresponding to a subpart of the whole system, and thus, to a subset of system parameters. Therefore, during that phase only a few of the system parameters are "active", and can influence the behavior of the cell. The other parameters are "silent", and can be changed without changing the behavior, as long as they are changed back when it is their turn, i.e. when their linear system is active. This also means that different parameters can only act as system-controls during specific timings of the cell cycle, there is a separation in time between different functional modules.

The change between linear system (and thus "active" system parameters), occurs when the system trajectory (i.e. protein concentrations), passes certain thresholds (phase space switching lines), either immediately or after a certain delay. Based on the DPL model it can be calculated when this occurs, and which and how different parameters can affect this. As an example, the switch between the G1 and the S/G2 phase is important for the final size of the cell, and by the DPL-model it can be predicted how large the cell will be if there is a change in one or more of the model parameters. It can also be predicted which parameters that can change the size of the cell, and which combination should be most effective. Other mechanistic principles can also be found, like the observation that the length of the G1 phase corresponds to the time delay of one of the system modules. These predictions are qualitatively, and semi-quantitatively verified in the original model.

The analysis of the DPL-model is based on the calculation of different important dynamical features like stability, fixed points, switching thresholds, eigenvalues and eigenvectors, and analytical conditions for these. The fact that the DPL-model consists of linear systems enables this, even though the inclusion of delay is a complicating factor. In the study presented in [23] it is shown that the dynamics of the DPL-model very well could be approximated by the slow eigenvectors of the different linear systems (Fig. 12.6), and this facilitated a proof on the global stability of the system. Finally, a further interesting feature were observed in the original model and explained in the simplified model [24]. The effect of different parameter perturbations was explored through two different characteristics; an essential effect of a perturbation made the system stop working (stop oscillating), whereas a modulatory

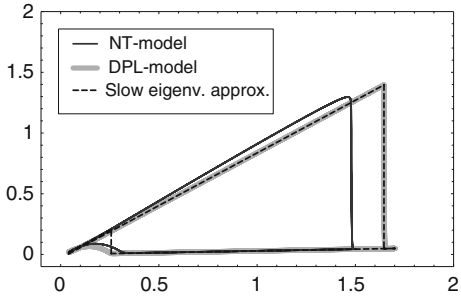

Fig. 12.6 Validation of the simplified DPL-model as well as the slow eigenvector approximation. From [23], for a colored version see the online version of the book

effect had a more minor but significant effect (change in cell size). It was noted that several parameters with no modulatory effect whatsoever for smaller perturbations could in fact be essential to the system, at larger perturbations.

12.5.2.6 What Is Lost by the Simplification?

Three important assumptions are made in the simplification process described above, (i) variables have time to get sufficiently close to their steady-state before there is a significant change of input, and (ii) that the transient behaviour is not important and (iii) that the exact form of the steady-state is not critical for the system, for example, a sigmoid function can be substituted by a step-function. If this is not the case then the simplification will fail. One idea behind this is that biological systems appear to be robust to many biochemical details [56, 82], and that this can be utilized to retrieve simplified coarse-grained descriptions.

The simplification of the original model was performed in a nominal point in the parameter space. In an extended investigation [24] it was analyzed how well the simplified model reproduced the dynamical behavior of the original model for other parameter values than the nominal point. It was noted that the models agreed well for smaller parameter perturbation, but for larger there was a disagreement.

12.6 Conclusions

We have within this chapter described how the huge increase in experimental cellular data of the last decades, and the following increase in the size and number of dynamical models, has been followed by an increasing need for and development of methods for model analysis through simplification. Since dynamical cellular models can consist of several hundreds of parameters, representing interactions between large numbers of species, connected in a nonlinear way by intertwined feedback-loops it is difficult to determine which parts of the models are important for different output. Another complicating factor is the model uncertainty, both when it comes to parameter values as well as topology. New experimental data are both qualitative, e.g. describing existence of interactions between species, as well as quantitative, e.g. describing interaction strength, but there is a mismatch between the amounts of qualitative versus quantitative data, resulting in not fully constrained or non-identifiable models.

Model simplification and reduction has herein been presented as a method to meet these challenges of complexity and identifiability of large-scale cellular models. The idea is to retrieve a smaller or in some other sense simpler model, with increased transparency (easier to understand intuitively), increased identifiability and/or increased predictability. It must also be possible to map the parts of the simplified model back to the biological entities or features of the original model, for predictions to be relevant.

Other utilizations of simpler models, which are not discussed so much in this chapter, are reduced computational times, as well as a means to compare models to each other.

Many different approaches have been taken towards simplification and reduction of models. There are 'horizontal' approaches, working on the same scale or level in the system hierarchy, for example lumping entities of the same kind together, like proteins or reactions, or decomposing the system into subsystems and then simplifying or analyzing these separately. Other 'vertical' approaches are approximating features on other scales than the one considered; like approximating a stochastic model by the average number of species to receive an ODE model; or approximating continuous functions with discrete functions, when transforming an ODE model to a Boolean or piecewise linear model; or the common theme of reducing variables working on faster time-scales. The different approaches differ in methodology, where, for example, horizontal approaches use model topology, sensitivity, identifiability, etc., vertical approaches often use approximations of different kinds, like averages.

It can also be noted that there is a difference between "local" and "global model simplification" approaches, i.e. some approaches are performing the simplification at a specific point in parameter space (e.g. time scale separation), whereas others are taking into account the full (or parts of the full) admissible parameter space, i.e. the parameter values for which the model output is consistent with experimental data (e.g. methods based on identifiability).

Traditional means of analyzing nonlinear dynamical models, like bifurcation analysis, do not suffice in order to understand these new large-scale models, rather new approaches have to be taken. Here model simplification and reduction have an important part to play.

Acknowledgments We would like to thank Yishao Zhou and Omar Gutierrez-Arenas for valuable discussions and important input, Sara Maad Sasane for commenting on the manuscript and the Swedish e-Science Research Center (SeRC) and the the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no 604102 (HBP) for financial support.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Adamczyk, M., van Eunen, K., Bakker, B.M., Westerhoff, H.V.: Enzyme kinetics for systems biology: When, why and how. Methods. Enzymol. **500**, 233–257 (2011)
2. Albert, R., Othmer, H.G.: The topology of the regulatory interactions predicts the expression pattern of the Drosophila segment polarity genes. J. Theor. Biol. **223**, 1–18 (2003)
3. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.D.: Molecular biology of the cell. Garland, New York (1994)

4. Alexander, R.P., Kim, P.M., Emonet, T., Gerstein, M.B.: Understanding modularity in molecular networks requires dynamics. Sci. Signal. **2**, pp. pe44 (2009)
5. Apri, M., de Gee, M., Molenaar, J.: Complexity reduction preserving dynamical behavior of biochemical networks. J. Theor. Biol. **304**, 16–26 (2012)
6. Bagowski, C.P., Besser, J., Frey, C.R., Ferrell Jr, J.E.: The JNK cascade as a biochemical switch in mammalian cells: ultrasensitive and all-or-none responses. Curr. Biol. **13**, 315–320 (2003)
7. Barabasi, A.L., Oltvai, Z.N.: Network biology: understanding the cell's functional organization. Nat. Rev. Genet. **5**, 101–113 (2004)
8. Behar, M., Dohlman, H.G., Elston, T.C.: Kinetic insulation as an effective mechanism for achieving pathway specificity in intracellular signaling networks. Proc. Nat. Acad. Sci. **104**, 16146–16151 (2007)
9. Bornholdt, S.: Less is more in modeling large genetic networks. Science **310**, 449–451 (2005)
10. Bruggeman, F.J., Westerhoff, H.V.: The nature of systems biology. Trends Microbiol. **15**, 45–50 (2007)
11. Cedersund, G.: Core-box modelling, theoretical contributions and applications to glucose homeostasis related systems. Ph.D. thesis, Chalmers University of Technology, Sweden, 2006
12. Chen, W.W., Schoeberl, B., Jasper, P.J., Niepel, M., Nielsen, U.B., Lauffenburger, D.A., Sorger, P.K.: Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. Mol. Syst. Biol. **5**, 239 (2009)
13. Christensen, C., Thakar, J., Albert, R.: Systems-level insights into cellular regulation: inferring, analysing, and modelling intracellular networks. IET Syst. Biol. **1**, 61–77 (2007)
14. Csete, M.E., Doyle, J.C.: Reverse engineering of biological complexity. Science **295**, 1664–1669 (2002)
15. Davidich, M., Bornholdt, S.: The transition from differential equations to boolean networks: A case study in simplifying a regulatory network model. J. Theor. Biol. **255**, 269–277 (2008)
16. de Jong, H.: Modeling and simulation of genetic regulatory systems: a literature review. J. Comput. Biol. **9**, 67–103 (2002)
17. de Jong, H., Gouzé, J.L., Hernandez, C., Page, M., Sari, T., Geiselmann, J.: Qualitative simulation of genetic regulatory networks using piecewise-linear models. Bull. Math. Biol. **66**, 301–340 (2004)
18. Del Vecchio, D., Ninfa, A.J., Sontag, E.D.: Modular cell biology: Retroactivity and insulation. Mol. Syst. Biol. **4**, (2008)
19. Dokoumetzidis, A., Aarons, L.: Proper lumping in systems biology models. IET Syst. Biol. **3**, 40–51 (2009)
20. Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R., Palsson, B.Ø.: Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc. Nat. Acad. Sci. **104**, 1777–1782 (2007)
21. Ehrenberg, M., Elf, J., Aurell, E., Sandberg, R., Tegnér, J.: Systems biology is taking off. Genome Res. **13**, 2377–2380 (2003)
22. Eriksson, O., Zhou, Y., Tegnér, J.: 'Modeling complex cellular networks - robust switching in the cell cycle ensures a piecewise linear reduction of the regulatory network'. CDC: IEEE Conf. Decis. Contr. **2004**(1), 117–123 (2004)
23. Eriksson, O., Brinne, B., Zhou, Y., Björkegren, J., Tegnér, J.: Deconstructing the core dynamics from a complex time-lagged regulatory biological circuit. IET Syst. Biol. **3**, 113–129 (2009)
24. Eriksson, O., Andersson, T., Zhou, Y., Tegnér, J.: Decoding complex biological networks - tracing essential and modulatory parameters in complex and simplified models of the cell cycle. BMC Syst. Biol. **5**, 123 (2011)
25. Ermentrout, B.: Simplifying and reducing complex models. In Bower, J.M., Bolouri, H. (Eds.) Computational Modeling of Genetic and Biochemical Networks, MIT Press, Cambridge, MA, pp. 307–323
26. Faeder, J.R., Blinov, M.L., Goldstein, B., Hlavacek, W.S.: Combinatorial complexity and dynamical restriction of network flows in signal transduction. Syst. Biol. (Stevenage) **2**, 5–15 (2005)

27. Feng, Y., Mitchison, T.J., Bender, A., Young, D.W., Tallarico, J.A.: Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. Nat. Rev. Drug Discov. **8**, 567–578 (2009)
28. Ferrell Jr, J.E.: Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. Trends Biochem. Sci. **21**, 460–466 (1996)
29. Glass, L.: Classification of biological networks by their qualitative dynamics. J. Theor. Biol. **54**, 85–107 (1975)
30. Glass, L., Kauffman, S.A.: The logical analysis of continuous, non-linear biochemical control networks. J. Theor. Biol. **39**, 103–129 (1973)
31. Gonçalves, J.: Constructive global analysis of hybrid systems. Ph.D. Thesis, Massachusetts Institute of Technology, 2000
32. Gureasko, J., Galush, W.J., Boykevisch, S., Sondermann, H., Bar-sagi, D., Groves, J.T., Kuriyan, J.: Membrane-dependent signal integration by the Ras activator Son of sevenless. Nat. Struct. Mol. Biol. **15**, 452–461 (2008)
33. Gutierrez-Arenas, O., Eriksson, O., and Hellgren Kotaleski, J.: Segregation and crosstalk of D1 receptor-mediated activation of ERK in striatal medium spiny neurons upon acute administration of psychostimulants. PLoS Comput. Biol. **10**, pp. e1003445 (2014)
34. Halász, A., Kumar, V., Imieliński, M., Belta, C., Sokolsky, O., Pathak, S., Rubin, H.: Analysis of lactose metabolism in E.coli using reachability analysis of hybrid systems. IET Syst. Biol. **1**, 130–148 (2007)
35. Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P., Vidal, M.: Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature **430**, 88–93 (2004)
36. Hardie, D.G., Salt, I.P., Hawley, S.A., Davies, S.P.: AMP-activated protein kinase: an ultrasensitive system for monitoring cellular energy charge. Biochem. J. **338**, 717–722 (1999)
37. Härdin, H.M., Zagaris, A., Krab, K., Westerhoff, H.V.: Simplified yet highly accurate enzyme kinetics for cases of low substrate concentrations. FEBS J. **276**, 5491–5506 (2009)
38. Harper, J.W.: A phosphorylation-driven ubiquitination switch for cell-cycle control. Trends Cell Biol. **12**, 104–107 (2002)
39. Hartwell, L.H., Hopfield, J.J., Leibler, S., Murray, A.W.: From molecular to modular cell biology. Nature **402**, c47–c52 (1999)
40. Huang, C.Y., Ferrell, J.E. Jr.: Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc. Natl. Acad. Sci. U.S.A. **93**, pp. 10078–10083 (1996)
41. Iyengar, R.: Why we need quantitative dynamic models. Sci. Signal. **2**, eg3 (2009)
42. Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival Jr, B., Assad-Garcia, N., Glass, J.I., Covert, M.W.: A whole-cell computational model predicts phenotype from genotype. Cell **150**, 389–401 (2012)
43. Kawai, J., et al.: Functional annotation of a full-length mouse cDNA collection. Nature **409**, 685–690 (2001)
44. Kepler, T.B., Abbott, L.F., Marder, E.: Reduction of conductance-based neuron models. Biol. Cybern. **66**, 381–387 (1992)
45. Kirk, P., Silk, D., Stumpf, M.P.H.: Reverse engineering under uncertainty. In: Uncertainty in Biology, A Computational Modeling Approach. Springer, Chem (2016, this volume)
46. Kitano, H.: Systems biology: A brief overview. Science **295**, 1662–1664 (2002)
47. Koschorreck, M., Conzelmann, H., Ebert, S., Ederer, M., Gilles, E.D.: Reduced modeling of signal transduction - a modular approach. BMC Bioinform. **8**, 336 (2007)
48. LeBeau, A.P., Robson, A.B., McKinnon, A.E., Sneyd, J.: Analysis of a reduced model of corticotroph action potentials. J. Theor. Biol. **192**, 319–339 (1998)
49. Lee, T.I., et al.: Transcriptional regulatory networks in Saccharomyces cerevisiae. Science **298**, 799–804 (2002)
50. Leloup, J.C., Goldbeter, A.: A model for circadian rhythms in Drosophila incorporating the formation of a complex between the PER and TIM proteins. J. Biol. Rhythms. **13**, 70–87 (1998)
51. Liu, G., Swihart, M.T., Neelamegham, S.: Sensitivity, principal component and flux analysis applied to signal transduction: the case of epidermal growth factor mediated signaling. Bioinformatics **21**, 1194–1202 (2005)

52. Liu, Y.-Y., Slotine, J.-J., Barabási, A.-L.: Observability of complex systems. Proc. Natl. Acad. Sci. U.S.A. **110**, 2460–2465 (2013)
53. Lohse, M.J., Nikolaev, V.O., Hein, P., Hoffmann, C., Vilardaga, J.P., Bünemann, M.: Optical techniques to analyze real-time activation and signaling of G-protein-coupled receptors. Trends Pharmacol. Sci. **29**, 159–165 (2008)
54. Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M., Teichmann, S.A., Gerstein, M.: Genomic analysis of regulatory network dynamics reveals large topological changes. Nature **431**, 308–312 (2004)
55. Monk, N.A.M.: Oscillatory expression of Hes1, p53, and NF-κB driven by transcriptional time delayes. Curr. Biol. **13**, 1409–1413 (2003)
56. Morohashi, M., Winn, A.E., Borisuk, M.T., Bolouri, H., Doyle, J., Kitano, H.: Robustness as a measure of plausibility in models of biochemical networks. J. Theor. Biol. **216**, 19–30 (2002)
57. Novak, B., Pataki, Z., Ciliberto, A., Tyson, J.J.: Mathematical model of the cell division cycle of fission yeast. Chaos **11**, 277–286 (2001)
58. Okino, M.S., Mavrovouniotis, M.L.: Simplification of mathematical models of chemical reaction systems. Chem. Rev. **98**, 391–408 (1998)
59. Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J.: Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics **25**, 1923–1929 (2009)
60. Saez-Rodriguez, J., Kremling, A., Conzelmann, H., Bettenbrock, K., Gilles, E.D.: Modular analysis of signal transduction networks. IEEE Control Syst. Mag. **24**, 35–52 (2004)
61. Saez-Rodriguez, J., Kremling, A., Gilles, E.D.: Dissecting the puzzle of life: modularization of signal transduction networks. Comput. Chem. Eng. **29**, 619–629 (2005)
62. Sanghvi, J.C., Regot, S., Carrasco, S., Karr, J.R., Gutschow, M.V., Bolival Jr, B., Covert, M.W.: Accelerated discovery via a whole-cell model. Nat. Methods **10**, 1192–1195 (2013)
63. Schena, M., Shalon, D., Davis, R., Brown, P.: Quantitative monitoring of gene expression patterns with a complementary dna microarray. Science **270**, 467–470 (1995)
64. Schliemann-Bullinger, M., Fey, D., Bastogne, T., Findeisen, R., Scheurich, P., Bullinger, E.: The experimental side of parameter estimation. In: Uncertainty in Biology, A Computational Modeling Approach. Springer, Chem (2016, this volume)
65. Schmidt, H., Madsen, M.F., Danø, S., Cedersund, G.: 'Complexity reduction of biochemical rate expressions'. Bioinformatics **24**, 848–854 (2008)
66. Smallbone, K., et al.: A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes. FEBS Lett. **587**, 2832–2841 (2013)
67. Smolen, P., Baxter, A.D., Byrne, J.H.: A reduced model clarifies the role of feedback loops and time delays in the Drosophila circadian oscillator. Biophys. J. **83**, 2349–2359 (2002)
68. Sorensen, M.E., DeWeerth, S.P.: An algorithmic method for reducing conductance-based neuron models. Biol. Cybern. **95**, 185–192 (2006)
69. Srividhya, J., Gopinathan, M.S., Schnell, S.: The effects of time delays in a phosphorylation-dephosphorylation pathway. Biophys. Chem. **125**, 286–297 (2007)
70. Sunnåker, M., Stelling, J.: Model extension and model selection. In: Uncertainty in Biology, A Computational Modeling Approach. Springer, Chem (2016, this volume)
71. Sunnåker, M., Zamora-Sillero, E., Dechant, R., Ludwig, C., Busetto, A. G., Wagner, A., and Stelling, J.: Automatic generation of predictive dynamic models reveals nuclear phosphorylation as the key Msn2 control mechanism. Sci. Signal. **6**, ra41 (2013)
72. Sunnåker, M., Schmidt, H., Jirstrand, M., Cedersund, G.: Zooming of states and parameters using a lumping approach including back-translation. BMC Syst. Biol. **4**, 28 (2010)
73. Sunnåker, M., Cedersund, G., Jirstrand, M.: A method for zooming of nonlinear models of biochemical systems. BMC Syst. Biol. **5**, 140 (2011)
74. The 1000 Genomes Project Consortium: An integrated map of genetic variation from 1,092 human genomes. Nature **491**, 56–65 (2012)
75. The ENCODE Project Consortium: An integrated encyclopedia of DNA elements in the human genome. Nature **489**, 57–74 (2012)

76. The International Human Genome Mapping Consortium: A physical map of the human genome. Nature **409**, 934–941 (2001)
77. Tyson, J.J., Chen, K., Novak, B.: Network dynamics and cell physiology. Nat. Rev. Mol. Cell. Biol. **2**, 908–916 (2001)
78. Tyson, J.J., Chen, J.C., Novak, B.: Sniffers, buffers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. Curr. Opin. Cell. Biol. **15**, 221–231 (2003)
79. Uetz, P., et al.: A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature **403**, 623–627 (2000)
80. Van Schepdael, A., Carlier, A., Geris, L.: Sensitivity analysis by design of experiments. In: Uncertainty in Biology, A Computational Modeling Approach. Springer, Chem (2016, this volume)
81. Venter, J., et al.: The sequence of the human genome. Science **291**, 1304–1351 (2001)
82. von Dassov, G., Meir, E., Munro, E.M., Odell, M.O.: The segment polarity network is a robust developmental module. Nature **406**, 188–192 (2000)
83. Willems, J.C.: Paradigm and puzzles in the theory of dynamical systems. IEEE Trans. Automatic Control **36**, 259–294 (1991)
84. Yasuda, R.: Imaging intracellular signaling using two-photon fluorescent lifetime imaging microscopy. Cold Spring Harb. Protoc. **2012**, 1121–1128 (2012)
85. Yildirim, N., Santillan, M., Horike, D., Mackey, M.C.: Dynamics and bistability in a reduced model of the lac operon. Chaos **14**, 279–292 (2004)
86. Zamora-Sillero, E., Hafner, M., Ibig, A., Stelling, J., Wagner, A.: Efficient characterization of high-dimensional parameter spaces for systems biology. BMC Syst. Biol. **5**, 142 (2011)