Journal of Theoretical Biology 304 (2012) 16–26

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Complexity reduction preserving dynamical behavior of biochemical networks

Mochamad Apri [a,b,c,*], Maarten de Gee [a,b], Jaap Molenaar [a,b]

[a] Biometris, Wageningen University and Research Center, 6708 PB, Wageningen, The Netherlands
[b] Netherlands Consortium for Systems Biology, 1090 GE, Amsterdam, The Netherlands
[c] Industrial and Financial Mathematics Group, Bandung Institute of Technology, 40132 Bandung, Indonesia

ARTICLE INFO

Article history:
Received 6 December 2011
Received in revised form
13 February 2012
Accepted 15 March 2012
Available online 23 March 2012

Keywords:
Complexity reduction
Dynamical behavior
Admissible region
Parameter estimation
Sloppy parameters

ABSTRACT

The complexity of biochemical systems, stemming from both the large number of components and the intricate interactions between these components, may hinder us in understanding the behavior of these systems. Therefore, effective methods are required to capture their key components and interactions. Here, we present a novel and efficient reduction method to simplify mathematical models of biochemical systems. Our method is based on the exploration of the so-called admissible region, that is the set of parameters for which the mathematical model yields some required output. From the shape of the admissible region, parameters that are really required in generating the output of the system can be identified and hence retained in the model, whereas the rest is removed.

To describe the idea, first the admissible region of a very small artificial network with only three nodes and three parameters is determined. Despite its simplicity, this network reveals all the basic ingredients of our reduction method. The method is then applied to an epidermal growth factor receptor (EGFR) network model. It turns out that only about 34% of the network components are required to yield the correct response to the epidermal growth factor (EGF) that was measured in the experiments, whereas the rest could be considered as redundant for this purpose. Furthermore, it is shown that parameter sensitivity on its own is not a reliable tool for model reduction, because highly sensitive parameters are not always retained, whereas slightly sensitive parameters are not always removable.

1. Introduction

Biochemical systems are usually very complex (Ross and Arkin, 2009), consisting of hundreds to thousands components with intricate interactions. A quantitative description of the dynamics of such systems is canonically formulated in terms of ordinary differential equations (ODEs). However, the usage of such large systems of ODEs is often problematic, since one is faced with the challenge to numerically solve a very large nonlinear set of differential equations. The complexity of this task is often a serious obstacle to get required information about the system. For example, the possible presence of multiple time scales in huge systems may result in unacceptably long computing times. Furthermore, often the question arises how to interpret the behavior of huge networks biologically.

At the level of the modeler, the problems are even more serious, because a large number of interactions gives rise to a large number of parameters. Of course, one may try to find values for these parameters in the literature, but even if they can be found, their reliability and applicability to the specific purpose is often unknown. Furthermore, after all available information from the literature is used to the limit, still a number of parameters may remain that have to be fitted to data. Fitting procedures require the system of ODEs to be solved iteratively, so it has to be evaluated many times which is quite time consuming. Other problems may involve the questions of identifiability, sensitivity, and robustness. Therefore, there is a need for reduction methods that deliver simplified models that still capture the essential dynamical behavior of the original system (Okino and Mavrovouniotis, 1998; Klipp et al., 2009).

Complexity reduction can be carried out in several ways, and the choice for the most appropriate approach depends on the purpose one has in mind. For example, one may try to decompose a large biochemical network into smaller submodules that have relatively little interaction with each other (Hartwell et al., 1999; Saez-Rodriguez et al., 2004, 2005, 2008; Conzelmann et al., 2004).

* Corresponding author at: Biometris, Wageningen University and Research Center, 6708 PB, Wageningen, The Netherlands. Tel.: +31 317 482384.
E-mail address: mochamad.apri@wur.nl (M. Apri).

0022-5193/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2012.03.019

In this way, the network becomes more manageable, easier to analyze, and it could make sense to study and interpret the modules separately. A large network can also be replaced by a functional module that has lower dimension (less ODEs) but still mimics the input–output behavior. This could be done, e.g., by replacing the network by a black box model (Liebermeister et al., 2005) or by lumping together components and/or reactions based on their characteristics (Conzelmann et al., 2004; Danø et al., 2006; Dokoumetzidis and Aarons, 2009; Sunnaker et al., 2010, 2011). A disadvantage of both black box modeling and lumping is that the new reduced model may be structurally different from the original one, e.g., a component in the new reduced model may be a linear combination of several components in the original model, or a component in the original model could be contained in several new components in the reduced model. In practice, this may obstruct us in interpreting the new reduced model.

Alternatively, complexity reduction can be carried out by selecting only those components and/or reactions of the network that determine the required dynamics of the system. This could be done, e.g., by exploiting possible time-scale differences that are often present in biological and chemical systems (Alon, 2007). However, the classical time-scale separation such as in Roussel and Fraser (2001), Kooi et al. (2002), and Radulescu et al. (2008) for example, raises the question how to obtain the prior knowledge that some reactions are fast and others are slow. On the other hand, the automatic time-scale separation in Maas and Pope (1992), Zobeley et al. (2005), and Surovtsova et al. (2009) requires that the original system of equations is mathematically transformed into another system before it can be reduced. This transformation impedes the biological interpretation of the reduced model.

One could also omit components and/or reactions that seem to contribute little to the behavior of the network. These components or reactions are usually selected via an optimization approach (Petzold and Zhu, 1999; Androulakis, 2000; Bhattacharjee et al., 2003) or a sensitivity analysis (Turányi et al., 1989; Turányi, 1990; Tomlin et al., 1995; Smets et al., 2002; Liu et al., 2005). In the first approach, given a nominal parameter values, it is investigated whether some parameters can be set to zero without adjusting the other ones. In the example given in Section 2.2 below, we show that this approach may not always be successful, even if the system is extremely simple. In the latter approach, the importance of a particular parameter is measured by evaluating the effect of variations in this parameter on the dynamics of all concentrations. If the effect is large, the system is said to be highly sensitive to this parameter. If, on the other hand, the sensitivity is low, this parameter is considered unimportant and removable. However, one should be careful with this kind of conclusions, since, e.g., the omission of a low-sensitivity parameter may lead to useless results. An example of this phenomenon is found in Chassagnole et al. (2002) from a model of the central carbon metabolism of *Escherichia coli*, for which the authors show that the sensitivity of flux concentrations to some enzymes, e.g., aldolase, is nearly zero. However, the removal of this reaction would result in the shutdown of the whole network, which is of course undesirable. For completeness's sake, we also mention the somewhat different way of reduction in which complicated mathematical expressions in the ODE equations are replaced with simplified ones (Schmidt et al., 2008; Ropers et al., 2011).

We present a novel reduction method that yields a biochemical model with less equations and parameters and still generates some required output. In this context, this output is interpreted as the dynamical behavior of the concentrations of a number of selected network components that are considered important for biological questions at hand and are responsible for its functional behavior. So, as data we take measured time series of some constituents and we look for a reduced model that generates these data. The reduction method proposed in this paper is based on the exploration of the so-called "admissible region", that is the region in the parameter space where the model outcomes match the output data within some given tolerance. From the shape of this admissible region important conclusions can be drawn. For example, if this region includes a part of one of the parameter axes, this parameter can apparently be set to zero. If, on the other hand, this region extends to infinity in some direction, this indicates that lumping of nodes might be allowed. These insights form our starting point to obtain reduced networks. The proposed method does not need to transform the original equations. It can be applied to any system of equations, linear or nonlinear. Contrary to the classical time-scale separation technique, this method does not rely on prior biological knowledge; therefore, it can be automated appropriately. On the other hand, the method can be tuned easily to incorporate any available prior knowledge. Our method also conserves the network structure and maintains the dynamics of the system's output and shows in this sense similarities with the method from (Radulescu et al., 2008). Once a reduced biologically plausible model has been obtained, parameter identification can be carried out more efficaciously.

This paper is organized as follows. In Section 2, we first introduce the concept of an admissible region which forms the basic concept in our reduction method. For illustrational purposes, we use a very simple artificial metabolic network with only three metabolites and three parameters. In spite of its simplicity, this system appears to be rich enough to show all the basic ingredients involved in the reduction process. Next we discuss how the reduction can be carried out effectively, and we conclude this section by formulating our algorithm. In Section 3, the reduction method is applied to a signaling network model taken from (Kholodenko et al., 1999). We show that this network can be considerably reduced with regard to reproducing the time-series of the key proteins. Furthermore, we demonstrate the surprising fact that parameters with high sensitivity are not always necessary and may be removed without any consequence for the purposes we have in mind.

2. Methods

A mathematical model is considered good if it is able to describe and predict the phenomena for which it has been designed. Here, we assume that the dynamical behavior of some components are essential for the predictive power of a model and that their dynamics has been measured. This dynamics should be preserved by any reduction method. In the following, we refer to system components that are measured as "target species". These target species play a pivotal role in the concept of admissible region, which is the basis idea of our reduction method.

2.1. The concept of admissible region

Consider a collection of n chemical species that form a biochemical network, the dynamics of which is modeled by a system of ordinary differential equations (ODEs)

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x},\mathbf{k}_0),$$
$$\mathbf{y}(t,\mathbf{k}_0) = \mathbf{B}\mathbf{x}(t,\mathbf{k}_0), \quad (1)$$

with initial values

$$\mathbf{x}(0) = \mathbf{x}_0. \quad (2)$$

Here $\mathbf{x}(t) \in \mathbb{R}^n$ denotes the vector of biochemical concentrations, $\mathbf{f}$ is the vector valued function representing the interactions, and