Mathematical Biosciences xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Mathematical Biosciences

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

Characterizability of metabolic pathway systems from time series data

Eberhard O. Voit *

The Wallace H. Coulter, Department of Biomedical Engineering at Georgia Tech. and Emory University, 313 Ferst Drive, Suite 4103, Atlanta, GA 30332-0535, United States

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Characterizability
Dynamic flux estimation
Identifiability
Metabolic pathway
Moore–Penrose inverse
Parameter estimation

ABSTRACT

Over the past decade, the biomathematical community has devoted substantial effort to the complicated challenge of estimating parameter values for biological systems models. An even more difficult issue is the characterization of functional forms for the processes that govern these systems. Most parameter estimation approaches tacitly assume that these forms are known or can be assumed with some validity. However, this assumption is not always true. The recently proposed method of Dynamic Flux Estimation (DFE) addresses this problem in a genuinely novel fashion for metabolic pathway systems. Specifically, DFE allows the characterization of fluxes within such systems through an analysis of metabolic time series data. Its main drawback is the fact that DFE can only directly be applied if the pathway system contains as many metabolites as unknown fluxes. This situation is unfortunately rare. To overcome this roadblock, earlier work in this field had proposed strategies for augmenting the set of unknown fluxes with independent kinetic information, which however is not always available. Employing Moore–Penrose pseudo-inverse methods of linear algebra, the present article discusses an approach for characterizing fluxes from metabolic time series data that is applicable even if the pathway system is underdetermined and contains more fluxes than metabolites. Intriguingly, this approach is independent of a specific modeling framework and unaffected by noise in the experimental time series data. The results reveal whether any fluxes may be characterized and, if so, which subset is characterizable. They also help with the identification of fluxes that, if they could be determined independently, would allow the application of DFE.
© 2013 Elsevier Inc. All rights reserved.

1. Introduction

A central challenge of computational systems biology is the translation of biological systems into mathematical models. Addressing this challenge critically depends on two components: data of high quality and effective strategies for model design, diagnostics, and analysis. The translation process itself consists of two steps, namely the determination of suitable mathematical representations and the identification of values for the parameters in these representations. Recent years have witnessed enormous efforts in the area of parameter estimation, indicating that parameter estimation is an unavoidable and very difficult task that is not yet completely solved (*e.g.*, [1–5]). Some of its difficulties are of computational nature, while others are due to the noisiness of biological data and the fact that several computed solutions often lead to similarly good data fits [6–11].

The parameter estimation task is not only difficult; it also makes a fundamental *a priori* assumption, namely, that the mathematical structure of the model representing the given data is known. However, this assumption is seldom entirely true; in fact, one could legitimately ask whether we ever truly know the structural format of a model in biology. The choice of a particular structure for a given modeling task may be rationalized in various ways. The traditional argument has been that certain functions or models had been used frequently and successfully in a particular biological subfield and therefore had developed into default representations. A good example in ecology is the Lotka–Volterra (LV) model, which describes the time-dependent changes in population sizes by linear and binomial terms that represent interactions among the various pairs of populations [12–15]. LV models have been very successful, but no ecologist would claim that they capture the dynamics of ecosystems in their full complexity. A second example is the Michaelis–Menten function [16], which was derived from a conceptual scheme describing the enzyme catalyzed conversion of a substrate into a product under idealized conditions. Although these conditions are seldom present in real cells [17,18], this function has been used as a default in thousands of biochemical studies, and even in cases that have not much to do with enzyme catalysis, such as the uptake of nutrients through the root system of a plant [19].

The choice of an appropriate model becomes more complicated in cases where the processes to be represented are aggregates of several steps [20,21]. An example is the ubiquitous effect of an extracellular ligand or an intracellular process like gene expression. At a coarse level, this effect is direct: if the ligand is present,

* Tel.: +1 (404) 385 5057; fax: +1 (404) 894 4243.
E-mail address: eberhard.voit@bme.gatech.edu

0025-5564/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.mbs.2013.01.008

Please cite this article in press as: E.O. Voit, Characterizability of metabolic pathway systems from time series data, Math. Biosci. (2013), http://dx.doi.org/10.1016/j.mbs.2013.01.008

a gene becomes expressed, and without a ligand, the gene is more or less silent. However, any description of this relationship in detail becomes exceedingly difficult, as it typically would have to account for physical changes in the conformation of the receptor, an entire signaling cascade, the translocation of a transcription factor, as well as the transcriptional machinery.

The alternative to an *a priori* assumption of a particular functional form is the use of a 'canonical' approximation, which is a representation that is based on theory, always leads to the same mathematical structures, and therefore permits streamlined analyses. The LV models mentioned earlier, as well as power-law models of Biochemical Systems Theory (BST; [4,18,22,23]) and linear representations fall into this category. The advantages of these representations include their guaranteed appropriateness at some operating point of choice, generality in applications, mathematical and computational tractability, and the fact that these types of models, at least initially, require very few application-specific assumptions. Nonetheless, canonical models are not necessarily the perfect answer to the question of what an appropriate representation of a process should look like, because they are not mechanistic and their parameters therefore do not have a mechanistic meaning. Also, like all other representations, they have by definition a limited range of valid approximation, and the size of this range is almost always unknown and difficult to assess.

The question thus arises whether it is possible to infer mathematical descriptions that adequately represent the true biological process without introducing too much bias. In an attempt to address this question for metabolic pathway systems, we recently proposed the method of Dynamic Flux Estimation (DFE; [24]), which is briefly reviewed in a later section. DFE uses as input the topology of a pathway system, together with time series measurements of the involved metabolites over a sufficiently wide time horizon. Of note is that DFE presupposes no knowledge or assumptions regarding the processes governing a metabolic system, but only of the topology of the network. In ideal cases, the input information is sufficient to prescribe a straightforward strategy for characterizing trends of all processes as they change over time or as they are affected by metabolites and modulators in the system. These resulting trends are not given as numerical functions, but as graphical representations. These plots, in turn, can directly be used for further analysis or allow the testing of numerically specified candidate functions. Thus, in contrast to identifiability tasks, which have the goal of determining optimal numerical settings for a model, the first phase of DFE addresses a characterizability task that precedes the identification of functional forms and parameters in the second phase.

Unfortunately, the ideal conditions allowing such an unbiased flux characterization are not often given. In particular, most metabolic pathway systems contain more fluxes than metabolite pools, and this discrepancy leads to a stoichiometric matrix of the flux system that has less than full rank (see later). Thus, unless additional information on sufficiently many fluxes is available, DFE cannot be applied. It is not even clear which fluxes would need to be identified independently to permit subsequent DFE. Discussion of this issue has led to suggestions for potentially helpful additional information, which could come from different sources. For example, in addition to the metabolic time series one might have measurements of some in- or effluxes. One might also be able to assume a flux representation from generally accepted kinetic knowledge [25]. If sufficiently many fluxes can be numerically characterized in this manner, the remaining fluxes can be computed in a point-wise fashion, as it is done in DFE with a system of full rank. If the data are rich enough, it is also sometimes possible to infer some fluxes from the data themselves [26].

This article presents an extended, general strategy for characterizing fluxes for pathway systems where the original DFE strategy is insufficient. The strategy uses a pseudo-inverse matrix method that reveals which reaction steps in a system are uniquely characterizable if time series data are available, even if the system is underdetermined. Secondly, the method permits the scanning for those reaction steps in a pathway system that, if they could be characterized independently, would be most beneficial for a subsequent DFE analysis. Intriguingly, the characterizability method proposed here is model free and uses only the topology of the pathway system, but no knowledge of regulatory features or specific time series data. The immediate result is a list of all reaction steps that could be uniquely characterized in a DFE sense if time series were available. Of course, the actual characterization of dynamic trends requires data, and a correct interpretation of these trends requires knowledge of the regulatory control patterns of the system.

2. Methods

2.1. Metabolic time series data

Modern ^{13}C- and ^{31}P-NMR methods permit the non-invasive determination of the concentrations of substrates and intracellular metabolites in living cell cultures. These measurements can be made every 30 s or even faster, thereby leading to dense metabolic time series data on the same cells and under the same conditions. In some sense, these data reflect all metabolic activities in these cultures, at least in principle. Examples of such data and their analysis can be found in [9,27–29].

Mass spectrometry (MS) has advanced to a point where very many metabolites in very small quantities can be identified simultaneously. While the method is destructive and requires the running of standards, it can be used to generate time series data as well. As an example, Kinoshita and colleagues measured metabolic time-courses of human red blood cell exposed to hypoxia, using capillary electrophoresis coupled to time-of-flight MS [30]. Other destructive methods, including liquid and gas chromatography, can similarly be used to establish metabolic profiles over relevant time horizons.

2.2. A brief review of dynamic flux estimation (DFE)

2.2.1. Rationale

The generic format of ODE models for metabolic systems is

$$\frac{d\mathbf{X}}{dt} = \dot{\mathbf{X}} = \mathbf{N} \cdot \mathbf{R}. \tag{1}$$

In this generic formulation, **X** is the vector of metabolite concentrations, **N** is the stoichiometric matrix, and **R** is a vector containing the specific reactions in the pathways. The stoichiometric matrix describes which variables are involved in which reaction [31–33]. An example is the branched pathway system in Fig. 1, which consists of one independent variable (X_0), three dependent variables (X_1,X_2,X_3) and five reactions (v_1,...,v_5), and has the stoichiometric matrix

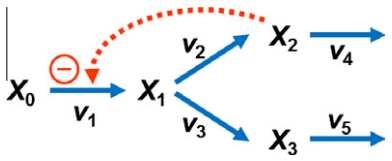

Fig. 1. Branched pathway with one feedback signal.

Please cite this article in press as: E.O. Voit, Characterizability of metabolic pathway systems from time series data, Math. Biosci. (2013), http://dx.doi.org/10.1016/j.mbs.2013.01.008

Download English Version:

https://daneshyari.com/en/article/6372123

Download Persian Version:

https://daneshyari.com/article/6372123

Daneshyari.com