BioSystems 109 (2012) 78–86

Contents lists available at SciVerse ScienceDirect

BioSystems

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

Solving dynamical inverse problems by means of Metabolic P systems

V. Manca*, L. Marchetti

University of Verona, Department of Computer Science, Strada Le Grazie 15, 37134 Verona, Italy

ARTICLE INFO

Article history:
Received 26 October 2011
Received in revised form
29 December 2011
Accepted 29 December 2011

Keywords:
Metabolic P systems
Dynamical systems
Dynamical inverse problems
Stepwise regression

ABSTRACT

MP (Metabolic P) systems are a class of P systems introduced for modelling metabolic processes. We refer to the dynamical inverse problem as the problem of identifying (discrete) mathematical models exhibiting an observed dynamics. In this paper, we complete the definition of the algorithm LGSS (Log-gain Stoichiometric Stepwise regression) introduced in Manca and Marchetti (2011) for solving a general class of dynamical inverse problems. To this aim, we develop a reformulation of the classical stepwise regression in the context of MP systems. We conclude with a short review of two applications of LGSS for discovering the internal regulation logic of two phenomena relevant in systems biology.

1. Introduction

The main framework analysis for the most part of biological dynamics remains the theory of ordinary differential equations (ODEs). *Metabolic P systems* (*MP systems*), based on Păun's P systems (Păun, 2002), were introduced in Manca et al. (2005) for modelling *metabolic systems* by means of suitable multiset rewriting grammars. They are essentially a particular type of finite difference recurrent equations where "fluxes" (see later) play a role analogous to that of derivatives in ODEs. This change of perspective, from a continuous to a discrete approach, provides in many cases computational and modelling advantages. The following discussion and the results of the present paper intend to argument an important case showing such a kind of advantages.

A Metabolic P system is essentially a multiset grammar where multiset transformations are regulated by functions (Manca, 2010; Păun and Rozenberg, 2010). Namely, a rule like $a + b \to c$ means that a number u of molecules of kind a and u of kind b are replaced by u molecules of type c. The value of u is the *flux* of the rule application. Let us assume to consider a system at some time steps $i = 0, 1, 2, \ldots, t$ ($i \in \mathbb{N}$, the set of natural numbers). Let us also assume that a substance x is produced by rules r_1, r_3 and consumed by rule r_2. If $u_1[i]$, $u_2[i]$, $u_3[i]$ are the fluxes of the rules r_1, r_2, r_3, respectively, in the passage from step i to step $i+1$, then the variation $\Delta_x[i]$ of substance x at step i is given by:

$$\Delta_x[i] = x[i+1] - x[i] = u_1[i] - u_2[i] + u_3[i]. \quad (1)$$

In an MP system, in any state the flux u_l of rule r_l is provided by a state function φ_l, called *regulator* of the rule. A state is essentially determined by the values of the system variables, that is, substances and parameters (quantities which are not transformed by the rules). However, usually only some variables enter as arguments of regulators, therefore if $u_l = \varphi_l(x, y, \ldots)$, the arguments $x, y, \ldots$ of φ_l will be called *tuners* of the regulator.

Substances (also metabolites), rules, initial values and regulators define an *MP grammar* which is easily representable by an *MP graph* (Manca and Bianco, 2008). The set of the rules of an MP grammar can be also represented by a *stoichiometric matrix* $\mathbb{A}$, which gives a sort of "matrix-like representation" of the system stoichiometry (see Fig. 1). An MP system is essentially an MP grammar equipped with a *temporal interval* τ, a conventional *mole size* ν, and substances masses, which specify the time and population (discrete) granularities, respectively (Manca, 2010; Păun and Rozenberg, 2010).

MP systems inherited from P systems the multiset rewriting mechanism as their fundament, by developing a different perspective. In fact, while P systems were essentially unconventional computational models, MP systems are intended to generate dynamics instead of computations. Namely, their aim in modelling biological phenomena is that of finding the multiset rewriting mechanism underlying an observed biological behaviour. They were successfully applied in many modelling contexts (Manca and Marchetti, 2010b,a; Manca et al., 2011; Marchetti and Manca, 2012)

* Corresponding author.
E-mail addresses: vincenzo.manca@univr.it (V. Manca), luca.marchetti@univr.it (L. Marchetti).

0303-2647/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2011.12.006

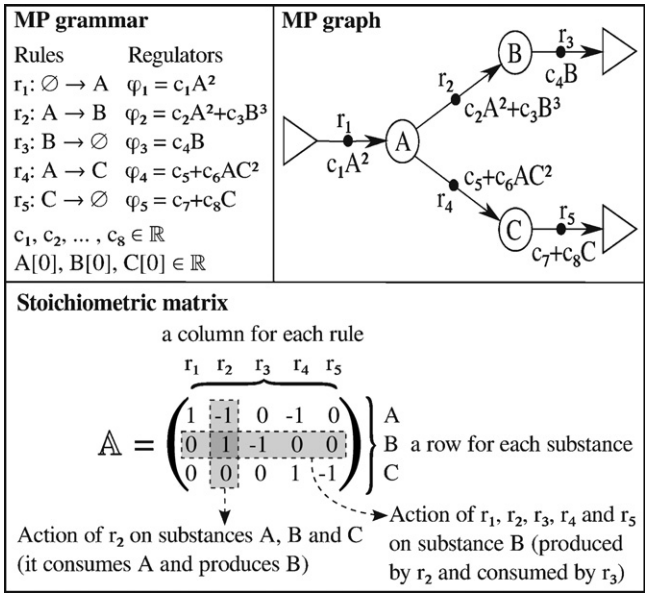

Fig. 1. An example of MP grammar (where ∅ denotes an empty multiset and substance symbols occurring in regulators denote the corresponding substance quantities), the stoichiometric matrix $\mathbb{A}$ is directly deduced by the MP grammar on the top left corner. The MP graph on the top right corner is obtained by translating the rules in the source-target-edge notation.

and in this paper we will present a systematic approach for obtaining MP models from time series of observed dynamics.

Now, let us consider the MP grammar of Fig. 1. If we suppose to start from a given initial state, then the fluxes of our system are given by calculating $\varphi_1(Z[i])$, $\varphi_2(Z[i])$, $\varphi_3(Z[i])$, $\varphi_4(Z[i])$ and $\varphi_5(Z[i])$ for some suitable values of $c_1, c_2, \ldots, c_8$ where we denote by $Z[i]$ the state vector of substances and parameters at time i.

This means that if $U[0]$ denotes the (column) vector of fluxes at time 0, then the substance variation vector

$$\Delta[0] = (\Delta_A[0], \Delta_B[0], \Delta_C[0])^T$$

is given by the following matrix (row by column) product

$$\Delta[0] = \mathbb{A} \times U[0]$$

that is

$$\Delta[0] = \begin{pmatrix} 1 & -1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix} \times \begin{pmatrix} u_1[0] \\ u_2[0] \\ u_3[0] \\ u_4[0] \\ u_5[0] \end{pmatrix}.$$

Therefore, a Metabolic P system with n substances, m reactions of regulators $\varphi_1, \varphi_2, \ldots, \varphi_m$, and stoichiometric matrix $\mathbb{A}$, has a dynamics given by the following *Equational Metabolic Algorithm* (*EMA*):

$$\Delta[i] = \mathbb{A} \times U[i] \tag{2}$$

where the flux vector $U[i]$ is computed by applying the regulators to the state vector $Z[i]$ of substances and parameters at time i.

1.1. The Dynamical Inverse Problem

Given an MP system, the recurrent equation system EMA (2) generates the evolution of substances, according to the MP grammar of the system (starting from an initial metabolic state, and with the knowledge of parameter evolution in time). In other words, when regulators $\varphi_1, \varphi_2, \ldots, \varphi_m$ are given, then fluxes can be computed and then the substance variations follow easily from the stoichiometry.

The *dynamical inverse problem* is the opposite verse of this process. In fact, let us assume to "observe" a metabolic system for a number of steps (separated by a temporal interval). How can we discover an MP system which provides the dynamics which we observe? The substances and the reactions in question are given by the particular phenomenon we want to describe. Moreover, the stoichiometry can be deduced by a basic knowledge of the phenomenon, but how to know the fluxes of matter transformation which are responsible of the observed evolution? This is the problem of discovering the flux regulation maps, or simply, the *Regulation Discovery Problem*. This problem can be considered as the "MP version" of a more general problem very important in systems biology (Ideker et al., 2001; Kitano, 2002), that is, how to define dynamical systems that explain observed dynamics of phenomena under investigation, by taking into account what is already known about each phenomenon. When such kind of systems are defined, then we can hope of discovering something new about the phenomena under investigation (Bailey, 1998). In ODE terms, this corresponds to the right determination of the kinetic constants of the model. However, even when differential models are prohibitive, due to the lack of information about the internal mechanism driving the dynamics, our method provides a grammatical formalization of the phenomenon under investigation, by discovering at same time numerical parameters and algebraic forms of regulators.

The Regulation Discovery Problem has been widely discussed in Manca and Marchetti (2011) where a powerful regression algorithm, called *LGSS* (*Log-gain Stoichiometric Stepwise regression*), has been introduced for solving the problem. The LGSS algorithm combines the log-gain principles developed in the MP system theory (Manca, 2009) with an extension of the classical method of Stepwise Regression (Hocking, 1976), which is a statistical regression technique based on Least Squares Approximation (Luenberger, 1969) and a statistical F-test (Draper and Smith, 1981). The method can be correctly applied independently from any knowledge about reaction rate kinetics, and can provide, with respect to differential models, different and even simpler mathematical formulations (Manca and Marchetti, 2010b,a; Manca et al., 2011; Marchetti and Manca, 2012).

In this paper we will extend the concepts introduced in Manca and Marchetti (2011) by formulating the stepwise regression method in order to be correctly applied to LGSS. The starting point of this formulation is the concept of stoichiometric expansion which will be explained in the next section.

2. The ADA Stoichiometric Expansion

If we know some time series of global states giving the state of the system at regular time intervals, that is, the vector sequence

$$(Z[i] | i \in \mathbb{N}),$$

then we can read Eq. (2) by reversing the known values with the unknown ones. In fact, by writing the substance variation vector

$$Z[i+1] - Z[i] = \Delta[i]$$

and assuming n substances and m reactions, then we get the following system *ADA* (Avogadro and Dalton Action, see Manca, 2010; Păun and Rozenberg, 2010)

$$\mathbb{A} \times U[i] = \Delta[i] \tag{3}$$

of n equations and m unknowns (the m components of the flux vector $U[i]$). Since usually the number of reactions in the system is greater than the number of substances, the number of unknowns of (3) is usually grater than the number of its equations and this makes the ADA system not univocally solvable. For this reason, in previous papers (Manca, 2008, 2010; Păun and Rozenberg, 2010)