

RESEARCH

Open Access

A constraint solving approach to model reduction by tropical equilibration

Sylvain Soliman^{1*}, François Fages¹ and Ovidiu Radulescu²

Abstract

Model reduction is a central topic in systems biology and dynamical systems theory, for reducing the complexity of detailed models, finding important parameters, and developing multi-scale models for instance. While singular perturbation theory is a standard mathematical tool to analyze the different time scales of a dynamical system and decompose the system accordingly, tropical methods provide a simple algebraic framework to perform these analyses systematically in polynomial systems. The crux of these methods is in the computation of tropical equilibrations. In this paper we show that constraint-based methods, using reified constraints for expressing the equilibration conditions, make it possible to numerically solve non-linear tropical equilibration problems, out of reach of standard computation methods. We illustrate this approach first with the detailed reduction of a simple biochemical mechanism, the Michaelis-Menten enzymatic reaction model, and second, with large-scale performance figures obtained on the <http://biomodels.net> repository.

Keywords: Systems biology, Model reduction, Tropical algebra, Tropical equilibration, Constraint programming

Background

Model reduction is a central topic in systems biology and dynamical systems theory, for reducing the complexity of detailed models, finding important parameters, and developing multi-scale models for instance.

Indeed, for many of the problems in computation and analysis of complex systems, the upper limit on the size of the system that can be studied has been reached. This limit can be very low, namely tens of variables for system identification, symbolic calculation or bifurcation of attractors of large dynamical systems. For instance, the complexity of extant symbolic solvers of polynomial equations is exponential in the number of indeterminates and parameters, that sets a drastic limitation to the size of the models that can be analyzed [1,2]. Some examples of computational difficulties that arise when trying to apply standard tools of algebraic geometry to systems biology models can be found in [3]. Model reduction is a way to bypass these limitations by replacing large scale models with models containing less parameters and variables, easier to analyse.

There are mathematical methods, based on singular perturbations or on the theory of invariant manifolds, allowing reduction of fully parametrized systems with separation of time scales. More precisely, in dissipative systems, fast variables relax rapidly to some low dimensional attractive manifold called invariant manifold [4] that carries the slow mode dynamics. A projection of dynamical equations onto this manifold provides the reduced dynamics. Numerical reduction methods, such as computational singular perturbation (CSP, [5]), intrinsic low dimensional manifold (ILDM, [6]) exploit the separation of timescales of various processes and compute approximations of the invariant manifold. Purely structural reduction methods can handle big models possibly with lack of kinetic information [7]. However, the case of biochemical models of intermediate size, with partially known parameters and that ask for symbolic analysis, is more open [8].

While singular perturbation theory is a standard mathematical tool to analyze the different time scales of a dynamical system and decompose the system accordingly, tropical methods provide a simple algebraic framework to perform these analyses systematically in polynomial systems, and in situations when model parameters are known only by their orders of magnitude. Differential

*Correspondence: Sylvain.Soliman@inria.fr

¹Inria, Domaine de Voluceau 78150, Rocquencourt, France
Full list of author information is available at the end of the article

equations describing kinetics of biochemical reactions are polynomial or become polynomial after decomposition of reaction mechanisms into elementary steps. For these models, quasi-equilibrium or quasi-steady state invariant manifolds allowing reductions are given by systems of algebraic equations [3]. A potentially crucial application of tropical mathematics is to enumerate and describe asymptotic solutions of algebraic systems of equations [9]. In particular, tropical solutions of polynomial equations provide the leading terms of their solutions via curves or in other terms via Newton-Puiseux series [10,11]. At the basis of our method lies the idea that equilibration of fast variables on invariant manifolds implies, at lowest order, equilibration of at least two dominant monomials, one positive and the other negative in the right hand side of the differential equations. We have called such a condition, similar to Kapranov's condition [11] for existence of Newton-Puiseux series with specified lowest order terms, tropical equilibration. The crux of our method lies in the computation of tropical equilibrations that define some reduced truncated systems with fewer parameters to identify, thus pointing to fewer experiments to (in)validate the model [12,13]. Our method copes with uncertain parameters by replacing exact values by orders of magnitude and the reduction is performed symbolically in both variables and parameters. With respect to methods based on singular perturbations, this could be less precise at lowest order, but it is more general in implementation.

Solving the tropical equilibration problem boils down to solving a system of equations in the min-plus algebra (also known as the tropical semiring). For solving linear tropical systems there are pseudo-polynomial algorithms, i.e. whose complexity is polynomial in the size of the system and in the absolute values of its coefficients [14]. In the nonlinear case, the existence of tropical equilibrations, which is equivalent to the problem of the intersection of tropical varieties, was shown to be NP-complete [15]. In this paper we show that constraint-based methods, using reified constraints for expressing the equilibration conditions, make it possible to numerically solve non-linear tropical equilibration problems, out of reach of standard computation methods.

We first illustrate this approach with the detailed reduction of a simple biochemical mechanism, the Michaelis-Menten enzymatic reaction model. We detail the general procedure to obtain truncated systems by identification, through equilibration, of fast and slow species, and relate the obtained reduced systems to the usual notions of quasi-steady-state and quasi-equilibrium. Then, we demonstrate that the approach is computationally feasible, and scales up properly, by treating in an automatic way all the curated dynamical models of <http://biomodels.net> repository [16].

Model reduction by tropicalization

We consider networks of biochemical reactions with mass action kinetic laws. The structure of a reaction is defined by a multiset rewriting rule as

$$\sum_{i=1}^n \alpha_{ji} A_i \rightarrow \sum_{k=1}^n \beta_{jk} A_k$$

where A_i , $i = 1, \dots, n$ denote the chemical species and α_{ji} , β_{jk} are positive integers named stoichiometric coefficients defining which species are consumed and produced by the reaction j , $1 \leq j \leq r$, and in which quantities.

The mass action law means that reaction rates are monomial functions of the species concentrations x_i , $1 \leq i \leq n$ and read

$$R_j(\mathbf{x}) = k_j \mathbf{x}^{\alpha_j},$$

where $k_j > 0$ are kinetic parameters and we use the shorthand notation $\mathbf{x}^{\alpha_j} = x_1^{\alpha_{j1}} \dots x_n^{\alpha_{jn}}$.

The network dynamics is then described by the following differential equations

$$\frac{dx_i}{dt} = \sum_{j=1}^r k_j S_{ij} \mathbf{x}^{\alpha_j}. \quad (1)$$

where $S_{ij} = \beta_{ji} - \alpha_{ji}$ are entries of the stoichiometric matrix.

In what follows, the kinetic parameters do not have to be known precisely, they are given by their orders of magnitude. A convenient way to represent orders is by considering that

$$k_j = \bar{k}_j \epsilon^{\gamma_j},$$

where ϵ is a positive parameter much smaller than 1, γ_j is an integer or, more generally, a rational number, and \bar{k}_j has order unity. An approximate integer order can be obtained from any real positive parameter by

$$\gamma_j = \text{round}(\log(k_j) / \log(\epsilon)),$$

where round stands for the closest integer. Notice that in this representation, small quantities have large orders. Furthermore, the smaller ϵ , the better the separation between quantities of different orders, indeed $\lim_{\epsilon \rightarrow 0} \frac{k_i}{k_j} = \infty$, if $\gamma_i < \gamma_j$.

We also define orders for species concentrations, using a vector of orders $\mathbf{a} = (a_1, \dots, a_n)$, such that $\mathbf{x} = \bar{\mathbf{x}} \epsilon^{\mathbf{a}}$. We suppose that various (a_1, \dots, a_n) are integers or rational numbers with a common denominator. In our method we will calculate the concentration orders as solutions of the tropical equilibration problem (see below).

First, let us replace Eqs. (1) by their equivalent rescaled versions

$$\frac{d\bar{x}_i}{dt} = \sum_{j=1}^r \epsilon^{\mu_j - a_i} k_j S_{ij} \bar{\mathbf{x}}^{\alpha_j}, \quad (2)$$

where

$$\mu_j = \gamma_j + \langle \mathbf{a}, \alpha_j \rangle, \tag{3}$$

and \langle, \rangle stands for the vector dot product.

The r.h.s. of each equation in (2) is a sum of multivariate monomials in the concentrations. The orders μ_j indicate how large are these monomials, in absolute value. For sufficiently small ϵ , monomials of different orders are well separated. For instance a monomial of smallest order $\mu_j < \mu_{j'}$ dominates the other monomials $k_j |S_{ij}| x^{\alpha_j} \gg k_{j'} |S_{ij'}| x^{\alpha_{j'}}$. One could see all these monomials as “forces” acting on the chemical species. At steady state, the resultant of all these forces should be naught. A consequence of this is that the orders of dominant positive and negative forces should be equal. This is exactly our notion of *tropical equilibration* that we introduced in [13]. More precisely, we say the system (2) is *tropically equilibrated* iff

$$\min(\mu_j, S_{ij} < 0) = \min(\mu_{j'}, S_{ij'} > 0), \text{ for all } i = 1, \dots, n \tag{4}$$

The tropical equilibration problem consists in solving the system (4) for orders a_i , $1 \leq i \leq n$.

Another way to understand the condition (4) is via Newton-Puiseux series. Suppose we want to solve the polynomial equation

$$P(x, \epsilon) = \sum_j b_j \epsilon^{\gamma_j} x^{\alpha_j} = 0, \tag{5}$$

where α_j are positive integers, γ_j are rational powers, and b_j are real coefficients. It is well known [10] that solutions of this equation can be expressed as Newton-Puiseux series, i.e. have the form

$$x(\epsilon) = c_1 \epsilon^{\frac{a_1}{q}} + c_2 \epsilon^{\frac{a_2}{q}} + \dots,$$

where c_i are complex coefficients, $a_1 < a_2 < \dots$ are integers, q is a positive integer. By substituting $x(\epsilon) = c_1 \epsilon^{\frac{a_1}{q}} (1 + x_1(\epsilon))$ (where $x_1(\epsilon)$ collects terms with positive orders in ϵ) in (5) we get

$$P(x, \epsilon) = \sum_j b_j c_1^{\alpha_j} \epsilon^{\gamma_j + a_1 \alpha_j / q} + r_1(\epsilon) = 0,$$

where $r_1(\epsilon)$ collects higher order terms. Necessary conditions for $P(x, \epsilon) = 0$ read at lowest order

$$\sum_{j, \gamma_j + a_1 \alpha_j / q = m} b_j c_1^{\alpha_j} = 0 \tag{6}$$

$$m = \min_j (\gamma_j + a_1 \alpha_j / q) \tag{7}$$

In order to satisfy (6), the minimum in (7) should be attained at least twice. Furthermore, if one looks for real solutions $c_i \in \mathbb{R}$, then from (6) it follows that at least two b_j corresponding to the minimum (7) should have opposite signs. This means that the lowest order a_1/q in the

Newton-Puiseux series solution has to satisfy a tropical equilibration problem.

We must emphasize that the tropical equilibration condition is weaker than the steady state condition, and makes sense also away from a steady state. In systems with slow/fast variables, the fast variables are equilibrated by compensation of dominant forces whose orders result from the tropical condition (4). As a consequence, the fast variables can be expressed as functions of the slow variables. However, both fast and slow variables can slowly evolve under the influence of weaker, higher order forces. This picture is valid as long as the relative dominance relations between various monomial terms in Eqs.(2) are preserved. This is true if the rescaled concentrations \bar{x}_i stay between bounds, whereas ϵ is allowed to tend to zero.

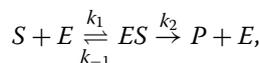
To summarize, the tropical equilibration is a necessary condition for the elimination of fast variables and model reduction. As we showed in [13], in order to become sufficient this condition should be combined with a boundedness condition, called permanency:

Definition 0.1. *The system (1) is permanent, if there are two constants $C_- > 0$ and $C_+ > 0$, a set of renormalization exponents a_i , and a function T_0 of the initial conditions, such that after renormalization we have*

$$C_- < \bar{x}_i(t) < C_+, \text{ for all } t > T_0(x(0)) \text{ and for every } i.$$

A simple example, the Michaelis-Menten reduction

The Michaelis-Menten enzymatic reaction network consists of three reactions:



where S, ES, E, P represent the substrate, the enzyme-substrate complex, the enzyme and the product, respectively.

The corresponding system of polynomial differential equations reads:

$$\begin{aligned} x_1' &= -k_1 x_1 x_3 + k_{-1} x_2 \\ x_2' &= k_1 x_1 x_3 - (k_{-1} + k_2) x_2 \\ x_3' &= -k_1 x_1 x_3 + (k_{-1} + k_2) x_2 \\ x_4' &= k_2 x_2 \end{aligned} \tag{8}$$

where $x_1 = [S], x_2 = [ES], x_3 = [E], x_4 = [P]$.

It can be easily checked that the system has two algebraic invariants: $(x_2 + x_3)' = 0$, which implies

$$x_2 + x_3 = e_0, \tag{9}$$

where e_0 is a positive constant (the total amount of enzyme), and

$$x_1 + x_3 + x_4 = s_0 \tag{10}$$

where s_0 is a positive constant (the total amount of substrate and product). These conservation laws can be used to reduce the model by elimination of x_4 (by (10)) and x_3 (by (9)). It follows:

$$\begin{aligned} x_1' &= -k_1x_1(e_0 - x_2) + k_{-1}x_2 \\ x_2' &= k_1x_1(e_0 - x_2) - (k_{-1} + k_2)x_2 \end{aligned} \quad (11)$$

The constraint $x_2 \leq e_0$ resulting from the elimination is also to consider, but we will see that all equilibrations of the above equations already imply it.

Tropical equilibration equations

By rescaling variables and parameters, we get $x_i = \bar{x}_i \epsilon^{a_i}$, $1 \leq i \leq 2$, $k_1 = \bar{k}_1 \epsilon^{\gamma_1}$, $k_{-1} = \bar{k}_{-1} \epsilon^{\gamma_{-1}}$, $e_0 = \bar{e}_0 \epsilon^{\gamma_e}$.

The tropical equilibration equations for the reduced system read:

$$\gamma_1 + \gamma_e + a_1 = \min(\gamma_1 + a_1, \gamma_{-1}) + a_2 \quad (12)$$

$$\gamma_1 + \gamma_e + a_1 = \min(\gamma_1 + a_1, \min(\gamma_{-1}, \gamma_2)) + a_2 \quad (13)$$

The set of integer (or rational) orders endowed with the $\oplus = \min$ and $\otimes = +$ operations is a semi-field, called min-plus algebra or tropical semi-field [17]. In this semi-field, $-\infty$ plays the role of 0 and 0 plays the role of 1. The multiplicative inverse of a is denoted $-a$. Our tropical equilibration problem means solving a set of polynomial equations in this semi-field. Using these notations and properties of semi-field operation, the tropical equations become:

$$\gamma_1 \otimes \gamma_e \otimes a_1 \otimes (-a_2) = (\gamma_1 \otimes a_1) \oplus \gamma_{-1} \quad (14)$$

$$\gamma_1 \otimes \gamma_e \otimes a_1 \otimes (-a_2) = (\gamma_1 \otimes a_1) \oplus \gamma_{-1} \oplus \gamma_2 \quad (15)$$

Classical Michaelis-Menten reduction

The classical derivation of the Michaelis-Menten reduction is based on the behaviour of the variable x_2 for the complex concentration. Using (8) it follows that:

$$x_4' = V_m x_2 / e_0$$

where $V_m = k_2 e_0$ is the maximum value of the production rate x_4' , since $x_2 \leq e_0$.

The variable x_2 satisfies equilibration relations and can be expressed as a function of a slow variable (either the substrate x_1 when x_2 is small, or the sum $x_1 + x_2$ in general) in two situations: quasi-stationarity and quasi-equilibrium.

The quasi-stationarity corresponds to setting x_2' to zero and is justified by the smallness of x_2 that can be considered a fast species (radical). More precisely one has $k_1 x_1 (e_0 - x_2) - (k_{-1} + k_2) x_2 = 0$, leading to $x_2 = x_1 e_0 / (K_m + x_1)$, where $K_m = (k_{-1} + k_2) / k_1$, i.e.

$$x_4' = V_m x_1 / (K_m + x_1) \quad (16)$$

The quasi-equilibrium corresponds to setting $k_1 x_1 (e_0 - x_2) - k_{-1} x_2 = 0$, meaning zero net flux of the first reaction in the mechanism. This leads to $x_2 = x_1 e_0 / ((k_{-1} / k_1) + x_1)$, i.e.

$$x_4' = V_m x_1 / ((k_{-1} / k_1) + x_1) \quad (17)$$

This is justified by having a very fast transformations in the direct and reverse sense by the first reaction, much faster than the transformations by the second reaction. In this case both x_1 and x_2 are fast, but their sum $x_1 + x_2$ is slow.

We show next, in Section ‘‘Tropical equilibrations and model reductions’’, that analysis of tropical equations provide the conditions for the asymptotic validity of quasi-stationarity and quasi-equilibrium approximations and also the exhaustive list of asymptotic regimes.

Geometrical interpretation

It was discussed in [13] that there is a bijection between the set of solutions of each tropical equation and parts of the tropical curves of the polynomials defining the ordinary differential equations. A tropical curve is defined as the locus of species concentration values (x, y) where at least two monomials of the considered polynomial are equal and larger than all the others. In logarithmic scale, this locus is made of lines, half-lines, or line segments [13,18]. There is one tropical curve for each differential equation. For instance, the tropical curve defined by the polynomial $-k_1 e_0 x_1 + k_1 x_1 x_2 + k_{-1} x_2$ is made of three half-lines with a common origin depicted in Figure 1, namely

$$\log(x_2) = \log(e_0), \quad \log(x_1) > \log(k_{-1} / k_1) \quad (18)$$

$$\log(x_1) = \log(k_{-1} / k_1), \quad \log(x_2) > \log(e_0) \quad (19)$$

$$\log(x_2) = \log(x_1) + \log(e_0 k_1 / k_{-1}), \quad \log(x_1) < \log(k_{-1} / k_1) \quad (20)$$

The solutions of the tropical equation (14) form two branches, corresponding to the two situations $(\gamma_1 \otimes a_1) \oplus \gamma_{-1} = (\gamma_1 \otimes a_1)$ and $(\gamma_1 \otimes a_1) \oplus \gamma_{-1} = \gamma_{-1}$, respectively. These are two half-lines in the plane of concentration orders:

$$a_2 = \gamma_e, \quad \gamma_1 + a_1 < \gamma_{-1} \quad (21)$$

$$a_2 = a_1 + \gamma_1 + \gamma_e - \gamma_{-1}, \quad \gamma_1 + a_1 > \gamma_{-1} \quad (22)$$

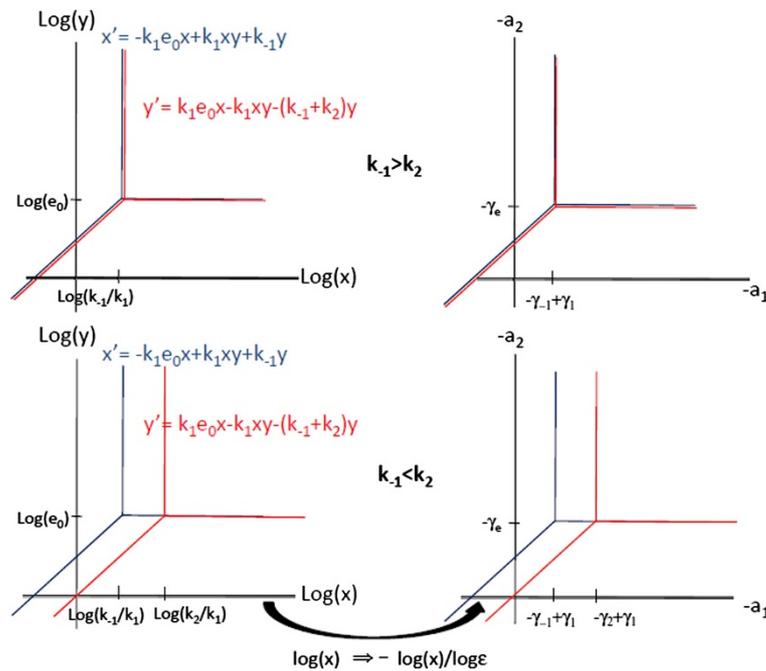


Figure 1 Tropical curves in the planes of concentrations and orders for the two variables Michaelis-Menten model. Tropical curves are defined as the locus of points where two monomials of a polynomial describing a differential equation are equal. The tropical curves for each differential equations are indicated by colors, blue for the first equation and red for the second equation. The vertical half-line of each of the tripods does not carry tropical equilibrations because it corresponds to equality of two monomials of the same sign. The horizontal and the oblique half-lines of the tripods carry tropical equilibrations. We have represented the two situations when $k_{-1} > k_2$ and when $k_{-1} < k_2$. All the tropical equilibrations are double (both variables are equilibrated) in the first case, and can be simple (only one variable is equilibrated) in the latter.

The two branches of solutions can be also related to parts of the tropical curves corresponding to the equilibration of monomials of different signs. More precisely (21) corresponds to (18), and (22) corresponds to (20). The branch (19) of the tropical curve corresponds to the equality of two positive monomials and has no correspondence in the set of tropical equilibrations.

Similarly to computing steady states as intersections of null-clines, we are considering multiple tropical equilibrations as intersections of tropical curves.

We therefore consider the second tropical equation (15), in two situations $\gamma_{-1} \oplus \gamma_2 = \gamma_{-1}$ and $\gamma_{-1} \oplus \gamma_2 = \gamma_2$. In the first case the tropical equation (15) is equivalent to the tropical equation (14) (also, the tropical curves coincide). Therefore, the two solutions (21) and (22) equilibrate both equations. In the second case, the solutions of the tropical equation (15) form two branches, corresponding to $(\gamma_1 \otimes a_1) \oplus \gamma_2 = \gamma_1 \otimes a_1$ and $(\gamma_1 \otimes a_1) \oplus \gamma_2 = \gamma_2$, respectively. They correspond to two half-lines in the plane of orders (a_1, a_2) , namely $a_2 = \gamma_e$, $a_1 < \gamma_2 - \gamma_1$ and $a_2 = a_1 + \gamma_1 + \gamma_e - \gamma_2$, $a_1 > \gamma_2 - \gamma_1$. A simple graphical inspection of the relative positions of these half-lines with respect to the half-lines carrying solutions of the first tropical equation shows that there are four branches of tropical equilibrations:

$$a_2 = \gamma_e, \quad a_1 < \gamma_2 - \gamma_1 \quad (23)$$

$$a_2 = \gamma_e, \quad \gamma_2 - \gamma_1 < a_1 < \gamma_2 - \gamma_1 \quad (24)$$

$$a_2 = a_1 + \gamma_1 + \gamma_e - \gamma_2, \quad a_1 > \gamma_2 - \gamma_1 \quad (25)$$

$$a_2 = a_1 + \gamma_1 + \gamma_e - \gamma_{-1}, \quad a_1 > \gamma_{-1} - \gamma_1 \quad (26)$$

The branch (23) equilibrates the two variables. The branch (25) equilibrates only the second variable, whereas the branches (24), (26) equilibrate only the first variable.

Tropical equilibrations and conservation laws

The reduced Michaelis-Menten mechanism with two dynamical variables has been obtained by elimination of a variable using an exact conservation law. It is interesting to compute the tropical equilibrations directly, in the unreduced model. In this three variables model, two of the equilibrium equations are identical. Like for computation of steady states, we need the conservation law as an extra constraint. If we treat this constraint exactly, we obtain the reduced model. An approximate treatment of Eqs. (8), (9),

considering equilibration of dominant terms, leads to the tropical problem:

$$\gamma_1 \otimes a_1 \otimes a_3 = \gamma_{-1} \otimes a_2 \quad (27)$$

$$\gamma_1 \otimes a_1 \otimes a_3 = (\gamma_{-1} \otimes a_2) \oplus (\gamma_2 \otimes a_2) \quad (28)$$

$$a_2 \oplus a_3 = \gamma_e \quad (29)$$

This tropical problem is different from (14), (15) and leads to different solutions in general. Firstly, let us notice that elimination is not possible in semi-fields, because there is no additive inverse in general. Hence, (27), (28) (29) can not be reduced to an equivalent system of two tropical equations. Secondly, dominant monomial equilibration in the reduced model does not always correspond to monomial equilibrations in the unreduced model. A typical example is the monomial x_1x_3 that becomes the difference $x_1e_0 - x_1x_2$ in the reduced model. The two monomials can equilibrate each other in the reduced model, but the same quantity is a unique, un-equilibrated monomial in the full model.

There are six branches of tropical solutions of the system (27), (28), (29). Two branches are obtained when $\gamma_{-1} \otimes a_2 = \gamma_{-1}$. In this case the two tropical equations (27), (28) are identical. Depending on $a_2 \oplus a_3 = a_2$, or $a_2 \oplus a_3 = a_3$ we get:

$$a_2 = \gamma_e, \quad a_1 < \gamma_{-1} - \gamma_1 \quad (30)$$

$$a_2 = a_1 + \gamma_1 + \gamma_e - \gamma_{-1}, \quad a_1 > \gamma_{-1} - \gamma_1 \quad (31)$$

These branches correspond to equilibrations of all the variables.

When $\gamma_{-1} \otimes a_2 = \gamma_2$ the two tropical Eqs. 27, (28) are incompatible. Depending on $a_2 \oplus a_3 = a_2$, or $a_2 \oplus a_3 = a_3$ and further choosing only one of the two tropical Eqs. 27, (28) we get the following branches:

$$a_2 = \gamma_e, \quad a_1 < \gamma_{-1} - \gamma_1 \quad (32)$$

$$a_2 = \gamma_e, \quad a_1 < \gamma_2 - \gamma_1 \quad (33)$$

$$a_2 = a_1 + \gamma_1 + \gamma_e - \gamma_{-1}, \quad a_1 > \gamma_{-1} - \gamma_1, \quad (34)$$

$$a_2 = a_1 + \gamma_1 + \gamma_e - \gamma_2, \quad a_1 > \gamma_2 - \gamma_1 \quad (35)$$

In the branches (32), (34), the variables x_2, x_3 are not equilibrated, whereas in the branches (33), (35), the variable x_1 is not equilibrated.

Comparison of Eqs. (30)-(35) and Eqs. (21)-(26) proves that the tropical equations of the unreduced model have

the same set of solutions as the reduced model. However, the branch of solutions (33) equilibrates all the variables in the reduced model and does not equilibrate the variable x_1 in the reduced model. The reason is exactly the one given above: the monomial x_1x_3 is dominant and un-equilibrated in the unreduced model, becomes $x_1e_0 - x_1x_2$ with equilibrated monomials in the reduced model.

Tropical equilibrations and model reductions

Tropical equilibrations can be used for model reduction. The reduction starts by tropical truncation. We call tropically truncated model the model obtained by elimination of all dominated monomials from the r.h.s. of the ordinary differential equations. The next step is ordering the variables according to the values of the exponents $\mu_i - a_i$. This allows to determine which variables are slow and fast.

An additional construction is needed in the case when the tropically truncated system of fast variables has conservation laws that are not conserved by the un-truncated system. The conservation laws define species pools that are supplementary slow variables. The pools follow differential equations involving previously dominated monomials.

For instance, in the two variables Michaelis-Menten model, we found essentially two types of reduced models, corresponding to quasi-equilibrium and quasi-stationarity approximations [19].

The branch (21) of tropical solutions leads to the following truncated system:

$$\begin{aligned} x'_1 &= -k_1x_1e_0 + k_{-1}x_2 \\ x'_2 &= k_1x_1e_0 - k_{-1}x_2 \end{aligned} \quad (36)$$

This truncated system has conserved quantity $z = x_1 + x_2$. The variable z is not conserved by the full model described by (11). Indeed, addition of Eqs. (11) leads to $z' = -k_{-1}x_2$. As the variable x_2 can be eliminated from $-k_1x_1e_0 + k_{-1}x_2 = 0$ and $x_1 + x_2 = z$ we have the reduced dynamics $z' = -k_zz$, where $k_z = k_{-1}/(1 + k_{-1}/(k_1e_0))$. For small ϵ , we can consider that $k_z \sim \epsilon^{\gamma_z}$, with $\gamma_z = \gamma_{-1} - \min(0, \gamma_{-1} - \gamma_1 - \gamma_e)$. Because $\mu_1 - a_1 = \gamma_1 + \gamma_e$, $\mu_2 - a_2 = \gamma_1 + \gamma_e + a_1 - a_2 = \gamma_{-1}$ the relation $k_z > \mu_1 - a_1$, $k_z > \mu_2 - a_2$ are always satisfied guaranteeing that z is slower than x_1, x_2 . The form (36) of the truncated equations and the conservation of $x_1 + x_2$ by the fast dynamics shows that this case corresponds to quasi-equilibrium of the first reaction in the Michaelis-Menten model, as described in Section "Classical Michaelis-Menten reduction", equation 17.

The branches (23) and (24) lead to quasi-equilibrium with the following truncated system:

$$\begin{aligned} x'_1 &= -k_1x_1(e_0 - x_2) \\ x'_2 &= k_1x_1(e_0 - x_2) \end{aligned} \quad (37)$$

These cases correspond to saturation of the enzyme (x_2 has the same order as e_0). A slow variable $z = x_1 + x_2$ has to be introduced as before, but the reduced dynamics is $z' = -k_{-1}x_2 = -k_{-1}e_0$.

The branch (25) leads to quasi-stationarity of the enzyme/substrate complex. In this case we have the tropical truncated system:

$$\begin{aligned} x_1' &= -k_1x_1e_0 \\ x_2' &= k_1x_1e_0 - k_2x_2 \end{aligned} \quad (38)$$

The variable x_1 is not equilibrated, which still allows for model reduction because this variable is slow. The fast variable x_2 is equilibrated, and the equilibration equation corresponds to the classical notion of quasi-stationary approximation, as described in Section “Classical Michaelis-Menten reduction”, equation 16. In this case, $\mu_1 - a_1 = \gamma_1 + \gamma_e$, $\mu_2 - a_2 = \gamma_1 + \gamma_e + a_1 - a_2 = \gamma_2$. The condition that x_1 is slower than x_2 reads $\mu_1 - a_1 > \mu_2 - a_2$ and we get the additional condition $\gamma_1 + \gamma_e > \gamma_2$, which is a low enzyme concentration condition.

The branch (26) leads to quasi-stationarity of the substrate with the following truncated system:

$$\begin{aligned} x_1' &= -k_1x_1e_0 + k_{-1}x_2, \\ x_2' &= -k_2x_2 \end{aligned} \quad (39)$$

The variable x_2 is not equilibrated, which is allowed only if this variable is slower than x_1 . In this case, $\mu_1 - a_1 = \gamma_1 + \gamma_e$, $\mu_2 - a_2 = \gamma_2$. The condition that x_2 is slower than x_1 reads $\mu_1 - a_1 < \mu_2 - a_2$ and leads to the additional condition $\gamma_1 + \gamma_e < \gamma_2$, which is a high enzyme concentration condition.

Finally, the branch (24) leads to the truncated system:

$$\begin{aligned} x_1' &= -k_1x_1(e_0 - x_2) \\ x_2' &= -k_2x_2 \end{aligned} \quad (40)$$

The variable x_1 is equilibrated, but it can not satisfy permanency. Indeed, at fixed x_2 this variable will converge to zero. Therefore, this tropical equilibration does not lead to a reduced model.

Tropical equilibration as a constraint satisfaction problem

As explained in Section “Model reduction by tropicalization”, given a biochemical reaction system with its Mass-Action kinetics, and a small ϵ , the problem of tropical equilibration is to look for a rescaling of the variables such that the dominating positive and negative term in each ODE *equilibrate* as per Eqs. (4), i.e., are of the same degree in ϵ .

Section “A simple example, the Michaelis-Menten reduction” showed that it is possible to iteratively reduce the equilibration problem to a linear system of equations for each possible pair of positive and negative dominating

monomial. It is actually possible to consider fewer pairs by restricting that search to the pairs denoting edges of the Newton polygon [13]. Nevertheless, the number of linear systems to consider remains exponential in the number of species, and may lead to redhibitory computational costs, especially when handling biochemical systems with hub molecules, i.e., molecules involved in a high number of reactions (e.g., E2F, p53, cMyc in cell-cycle control or NF κ B in signalling), which corresponds to a Newton polygon with many vertices.

In order to tackle that complexity, we propose a numerical approach based on Constraint Programming, that encodes the equilibration problem as a Constraint Satisfaction Problem (CSP) [20-22] and uses reified constraints to prune the search space. Constraint Programming is a paradigm that has showed great success at solving combinatorial decision or optimization problems, in particular for real-world instances of NP-hard problems, e.g., in the field of production planning and scheduling. It is therefore an interesting way to approach the combinatorial explosion described above.

In presence of invariants (conservation laws) in the original system, Section “Conservation law constraints” has shown that some constraints related to rescaling need be added. We have shown in [23] that finding these conservation laws can be efficiently solved by constraint methods. Here we will thus assume that the conservation laws are given in input. In our prototype implementation, both the computation of conservation laws and the following equilibration are performed for a given system.

Reified constraints

Key to the modeling of tropical equilibration problems as CSP are reified constraints. Reified constraints are special constraints that link in a bidirectional way the value of a boolean variable to the satisfaction of another constraint. They allow for powerful cuts in the search space by propagating the truth value of some constraints of the problem to the truth value of the Boolean variable, and vice versa.

For instance, the reified constraint

$$B\# \iff X\# = Y$$

states that the Boolean variable B is true (i.e. equal to 1) if and only if the constraints $X = Y$ is satisfied. That constraint posts the constraint $X = Y$ (resp. $X \neq Y$) as soon as B gets value 1 (resp. 0), and vice versa, sets $B = 1$ (resp. $B = 0$) as soon as $X = Y$ (resp. $X \neq Y$ i.e. when the domains of X and Y become disjoint).

For the tropical equilibration problem, these constraints are at the core of our representation of the minimum constraints as they enforce the propagation of existing knowledge before branching on the two possible values. Indeed, if A is the minimum of B and C , you can derive

many things on the domains of A , B and C before eventually trying $A = B$ or $A = C$. For instance it is safe to add that $A \leq B$ and $A \leq C$, but also if you have, from other equations, that $B \geq \min_B$ and $C \geq \min_C$ then you can add the fact that $A \geq \min(\min_B, \min_C)$. If later you obtain that actually $A = B$ then you can enforce $C \geq B$, etc. Section “Minimum constraints” shows in more detail how reified constraint do precisely this kind of conditional addition of cuts and can therefore be used to handle minimum constraints while postponing enumerative search as much as possible.

Variables and domains

For practical reasons, mainly the lack of an efficient solver over rationals with reified constraints, we use a finite domain solver and therefore only look for integer solutions (whereas solutions are rational). In practice this did not seem to change much the nature of results, see Figure 2. Extensions of the approach to cope with half-integer solutions or with rational solutions with a common, small denominator are straightforward.

For each original equation dx_i/dt , $1 \leq i \leq n$ is introduced a variable $a_i \in \mathbb{Z}$ that is used to rescale the system by posing $x_i = \epsilon^{a_i} \tilde{x}_i$. These are the variables of our CSP. Note that they require a solver handling \mathbb{Z} like for instance SWI-Prolog [24,25] with the `clpfd` library by Markus Triska, which we used for our implementation.

Tropical equilibration constraints

For each differential equation that should be equilibrated is a list of positive monomials M_i^+ , and a list of negative monomials M_i^- . The degrees in ϵ of all these monomials are integer linear expressions in the a_i . Now, to obtain an equilibration one should enforce for each i that the minimum degree in M_i^+ is equal to the minimum degree in M_i^- . This corresponds to the Eqs. 4. We will see how they can be implemented with reified constraints, but for now,

let us assume a constraint $\min(L, M)$ that enforces that the variable M of \mathbb{Z} is the minimum value of a list L of linear expressions over variables of \mathbb{Z} . We have in our CSP, for each $1 \leq i \leq n$, $\min(\text{PositiveMonomialDegrees}, M)$ and $\min(\text{NegativeMonomialDegrees}, M)$.

Conservation law constraints

The second kind of constraint comes from conservation laws. Each conservation law is an equality between a linear combination of the x_i and a constant c_i . By rescaling, we obtain a sum of rescaled monomials equal to $\epsilon^{\log(c_i)/\log(\epsilon)} \tilde{c}_i$. We want this equality to hold when ϵ goes to zero, which implies that the minimal degree in ϵ in the left hand side is equal to (the round of) the degree of the right hand side. Since once again the degrees on the left are linear combinations of our variables a_i , this is again a constraint of the form: $\min(\text{ConservationLawDegrees}, K)$ where K is equal to $\text{round}(\log(c_i)/\log(\epsilon))$. This corresponds to the tropical equation (29).

Minimum constraints

Furthermore, if the system under study is not at steady state, the minimum degree should not be reached only once, which would lead to a constant value for the corresponding variable when ϵ goes to zero, but at least twice. This is the case for the example treated in [12]. The constraint we need is therefore slightly more general than $\min/2$: we need the constraint $\min(L, M, N)$ which is true if M is smaller than each element of L and equal to N elements of that list. Note that using CLP notation, we have:

$$\min(M, L) : - \quad C\# \geq 1, \quad \min(M, L, C).$$

In order to enforce that the minimum is reached at least a required number of times, one obvious solution is to try all pairs of positive and negative monomials and

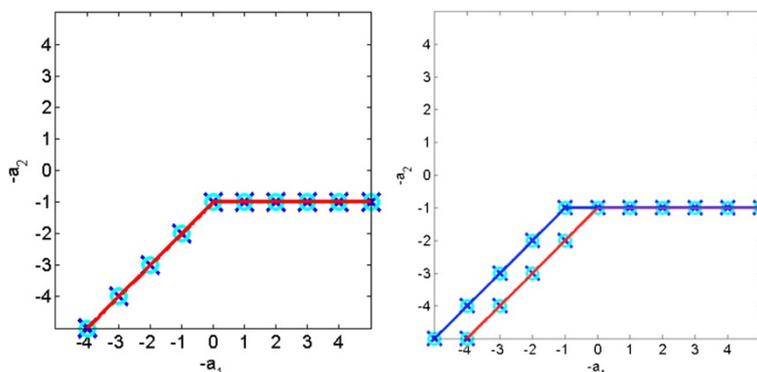


Figure 2 Comparison of the theoretical and computed equilibrations in the cases $k_{-1} > k_2$ and $k_{-1} < k_2$. The circles are equilibrations computed for the simplified two variables Michaelis-Menten model, the crosses are for the full three variables model. The lines indicate the theoretical equilibrations.

count the successful pairs [26]. However, this is not necessary, the $\min(L, M, N)$ constraint directly expresses the cardinality constraint on the minimums and can be implemented using *reified constraints* to propagate information between L , M and N in all directions, without enumeration. Using SWI-Prolog notations, the implementation of $\min/3$ by reified constraints is as follows:

```
min([], 0).
min([HT], M, C) : -M# =< H, B# <==>
M# = H, C# = B + CC, min(T, M, CC).
```

The translation of this predicate into words is roughly as follows, first ignoring the counts: M is smaller than a list with head H and tail T , if it is smaller than the tail T and it is smaller than the head, i.e., $M \leq H$. Now, we also impose that the value M is reached C times as follows: it is reached CC times in the tail and $C = B + CC$ where B is a variable equal to 1 iff M is equal to the head and 0 otherwise. Note that this latest statement is enforced through a reified constraint, it will therefore not lead to immediate branching but to the propagation of as much information as possible (e.g., if some values in the list are already known to be strictly greater than others, the corresponding boolean for each of them will be set to 0, and thus the sum will by necessity enforce some other values to be equal to the required minimum).

This concise and portable implementation will probably improve when the minimum and \min_n global constraints are available (see [27] for a reference). However it already proves very efficient as demonstrated in the next section.

When C is equal to one, we can fall back to using the built-in \min construct in a constraint (e.g., $M \# = \min(L1, \min(L2, L3))$). Some preliminary benchmarking showed that the reified version is more efficient if the length of the list is greater than 3 or 4.

Enumeration strategy

Constraints over finite domains come with domain filtering algorithms which dynamically prune the domain of variables when the domain of other variables change in a constraint. However this strategy is not complete and must be combined with a search procedure for virtually enumerating all possible values of the variables. For this application we obtained good performances with dichotomic search by bisecting the domain of the variables (bisect option in SWI-Prolog) without any particular heuristics for choosing the variables.

Note that since this approach is numerical, contrary to solving *symbolically* an exponential but finite number of linear systems as done in Section “A simple example, the Michaelis-Menten reduction” and in [13], there can be an infinite number of solutions. This situation denotes an

under-constrained linear system and remains to be interpreted biologically. In practice bounds are put on variables in order to force finiteness. This is not a restriction in practice since biochemical species’ concentrations usually do not vary by more than a hundred of magnitude orders.

Furthermore, in order to speed-up the computation of all solutions in such large domains, we used an iterative domain expansion strategy: the problem is first tried with a domain of $[-2, 2]$ for all variables, i.e., equilibrations are searched by rescaling in the $10^{-2}, 10^2$ interval. If that fails, the domain is doubled and the problem tried again until a limit of $10^{-128}, 10^{128}$.

Computation results on Biomodels.net

To benchmark our approach, we applied it systematically to all the dynamical models of the curated part of the <http://biomodels.net> repository [16] of biological systems, with ϵ set arbitrarily to 0.1.

We used release *r24* from 2012-12-12 which includes 436 curated models. Among them, only 55 models have non-trivial purely polynomial kinetics (ignoring *events* if any). Our computational results on those are summarized in Table 1, where the first column indicates whether a complete equilibration was found, and the times are in seconds.

The domain expansion strategy coupled with dichotomic search by domain bisections allowed us to gain two orders of magnitude of computation time on the biggest models.

Only one of the models (number 002) used values far from 0 in the equilibration (up to ϵ^{40}) and has no complete equilibration if the domain is restricted to $[-32, 32]$. This is because the model is written with units such that the initial concentrations are of the order 10^{-21} , translating the search accordingly. We thus do not believe that enlarging the domains even more would lead to more equilibrations. Nevertheless, choosing a smaller ϵ might increase the number of equilibrations.

18 of the 23 models for which there is a complete equilibration are actually under-constrained and appear to have an infinity of such solutions (typically linear relations between variables). For the 5 remaining ones, we computed all complete equilibrations as shown in Table 2.

Table 1 Number of models of the BioModels repository, with a polynomial kinetics, for which tropical equilibrations were found or not, with corresponding size of the model and computation time

Found	# models	Variables (avg/min/max)	Time in seconds (avg/min/max)
yes	23	17.348/3/ 86	0.486/0.004/2.803
no	32	17.812/1/194	0.099/0.000/1.934

Table 2 Number of equilibrations and computation time for the models of the BioModels database with finitely many numerical solutions

Model	# variables	# equilibrations	Total time (s)
BIOMD0000000002	13	18	53
BIOMD0000000122	14	9	4.1
BIOMD0000000156	3	1	<1ms
BIOMD0000000229	7	1	0.07
BIOMD0000000413	5	5	0.4

Conclusions

In this paper we have shown that constraint-based methods can be efficiently used to numerically solve tropical equilibration problems in biological models of real-size in the BioModels.net repository. These calculations are important for model reduction and for determining the unknown orders of the variables. Once the orders of the variables are known, the rapid variables can be identified and the system reduced to a simpler one. This truncation, described in Section “Tropical equilibrations and model reductions” coupled with the proposed constraint-based method for finding equilibrations therefore provides an automatic way to reduce models and to identify fast/slow variables. We have started the application of such technique on non-trivial models provided by biologists and modellers and hope to be able to improve both the understanding, through that identification of fast/slow variables, and the analysis, through the size reduction, of those models.

Even with the progress of high-throughput technologies, having more focused models, with fewer species and parameters to measure, will definitely permit an improvement in the quality and speed of development of the models. Furthermore, the structural methods for comparing models in model repositories, such as [7], can be refined by filtering the structural reduction relationships according to the kinetics of the reactions and the tropical reasoning on the magnitude orders.

In many cases, it makes sense biologically to only look for partial equilibrations. Strategies to decide when such decision has to be made remain unclear. Nevertheless the framework of partial constraint satisfaction and more specifically Max-CSP [28] would allow us to easily handle the maximization of the number of equilibrated variables.

One of the limits of this approach, is that it is not particularly well suited to equilibration problems with an infinite number of solutions. As discussed at the end of previous section, in such situations symbolic solutions would be more appropriate. Nevertheless, even the approximate detection of such a case by the very high number of (bounded) numerical solutions was shown to be not very costly in practice.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

FF and OR designed the study. SS devised the algorithm and conducted the experiments. All authors equally contributed to the writing of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work has been supported by the French ANR BioTempo, CNRS Peps ModRedBio, EPIGENMED Excellence Laboratory and OSEO Biointelligence projects.

Author details

¹Inria, Domaine de Voluceau 78150, Rocquencourt, France. ²University of Montpellier 2, Place Eugene Bataillon, 34095 Montpellier, France.

Received: 4 April 2014 Accepted: 4 November 2014

Published online: 04 December 2014

References

1. Grigoriev D, Vorobjov N: **Solving systems of polynomial inequalities in subexponential time.** *J Symbolic Computat* 1988, **5**:37–64.
2. Grigoriev D: **Complexity of quantifier elimination in the theory of ordinary differential equations.** *Lect Notes Comput Sci* 1989, **18**:11–25.
3. Pantea C, Gupta A, Rawlings JB, Craciun G: **The QSSA in chemical kinetics: as taught and as practiced.** In *Discrete and Topological Models in Molecular Biology*. Berlin: Springer; 2014:419–442.
4. Gorban A, Karlin I: **Invariant manifolds for physical and chemical kinetics.** *Lect Notes Phys* 2005, **660**:1–491.
5. Lam S, Goussis D: **The CSP method for simplifying kinetics.** *Int J Chem Kinet* 1994, **26**(4):461–486.
6. Maas U, Pope SB: **Simplifying chemical kinetics: intrinsic low-dimensional manifolds in composition space.** *Combustion Flame* 1992, **88**(3):239–264.
7. Gay S, Soliman S, Fages F: **A graphical method for reducing and relating models in systems biology.** *Bioinformatics* 2010, **26**(18):i575–i581. [Special issue ECCB'10].
8. Radulescu O, Gorban AN, Zinovyev A, Noel V: **Reduction of dynamical biochemical reactions networks in computational biology.** *Front Genet* 2012, **3**:131. [http://www.frontiersin.org/bioinformatics_and_computational_biology/10.3389/fgene.2012.00131/abstract]
9. Sturmfels B: *Solving systems of polynomial equations, Volume 97.* American Mathematical Soc: Providence; 2002.
10. Walker RJ: *Algebraic curves.* New York: Springer; 1978.
11. Einsiedler M, Kapranov M, Lind D: **Non-archimedean amoebas and tropical varieties.** *J für die reine und angewandte Mathematik (Crelles J)* 2006, **2006**(601):139–157.
12. Noel V, Grigoriev D, Vakulenko S, Radulescu O: **Tropical geometries and dynamics of biochemical networks application to hybrid cell cycle models.** *Electron Notes Theor Comput Sci* 2012, **284**:75–91.
13. Noel V, Grigoriev D, Vakulenko S, Radulescu O: **Tropicalization and tropical equilibration of chemical reactions.** In *Tropical and Idempotent Mathematics and Applications, Volume 616 of Contemporary Mathematics*. Edited by Litvinov G, Sergeev S: American Mathematical Society; 2014:261–277.
14. Grigoriev D: **Complexity of solving tropical linear systems.** *Comput Complexity* 2013, **22**:71–88.
15. Theobald T: **On the frontiers of polynomial computations in tropical geometry.** *J Symbolic Comput* 2006, **41**(12):1360–1375.
16. le Novère N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, Schilstra M, Shapiro B, Snoep JL, Hucka M: **BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems.** *Nucleic Acid Res* 2006, **1**(34):D689–D691.
17. Cohen G, Gaubert S, Quadrat J: **Max-plus algebra and system theory: where we are and where to go now.** *Ann Rev Control* 1999, **23**:207–219.
18. Viro O: **From the sixteenth Hilbert problem to tropical geometry.** *Jpn J Math* 2008, **3**(2):185–214.

19. Gorban AN, Radulescu O, Zinovyev AY: **Asymptotology of chemical reaction networks**. *Chem Eng Sci* 2010, **65**(7):2310–2324. [International Symposium on Mathematics in Chemical Kinetics and Engineering].
20. Mackworth AK: **Consistency in networks of relations**. *Artif Intell* 1977, **8**:99–118.
21. Meseguer P: **Constraint satisfaction problems: an overview**. *A.I. Commun* 1989, **2**:3–17.
22. Kumar V: **Algorithms for constraint- satisfaction problems: a survey**. *A.I. Mag* 1992, **13**:32–44.
23. Soliman S: **Invariants and other structural properties of biochemical models as a constraint satisfaction problem**. *Algorithms Mol Biol* 2012, **7**(15):15.
24. Wielemaker J, Schrijvers T, Triska M, Lager T: **SWI-Prolog**. *Theory Pract Logic Program* 2012, **12**(1-2):67–96.
25. Wielemaker J: *SWI-Prolog 6.3.15 Reference Manual*; 1990. [<http://www.swi-prolog.org/pldoc/refman/>]
26. Radulescu O, Gorban A, Zinovyev A, Noel V: **Reduction of dynamical biochemical reaction networks in computational biology**. *Front Bioinformatics Comput Biol* 2012, **3**:131.
27. Beldiceanu N, Carlsson M, Demassey S, Petit T: **Global constraints catalog**. Tech. Rep. T2005-6, Swedish Institute of Computer Science 2005.
28. Freuder EC, Wallace RJ: **Partial constraint satisfaction**. *Artif Intell* 1992, **58**:21–70.

doi:10.1186/s13015-014-0024-2

Cite this article as: Soliman et al.: A constraint solving approach to model reduction by tropical equilibration. *Algorithms for Molecular Biology* 2014 **9**:24.

Submit your next manuscript to BioMed Central
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

