

On the Verification and Correction of Large-scale Kinetic Models in Systems Biology

Attila Gábor¹, Katalin M. Hangos^{2,3}, Gábor Szederkényi^{2,4}, and Julio R. Banga¹

¹ BioProcess Engineering Group, IIM-CSIC, Vigo, Spain
{attila.gabor, julio}@iim.csic.es

² Computer and Automation Research Institute of the Hungarian Academy of Science, Budapest, Hungary
{hangos, szeder}@sztaki.mta.hu

³ University of Pannonia, Veszprém, Hungary

⁴ Pázmány Péter Catholic University, Budapest, Hungary

Abstract. In this paper we consider the problem of verification of large dynamic models of biological systems. We present syntactical criteria based on biochemical kinetics to ensure the plausibility of a model and the positivity of its solution. These criteria include the positivity of the rate functions, their kinetic type dependence on the reactant species concentrations, and the absence of the negative cross-effects that together guarantee the nonnegativity of the dynamics. Further, the stoichiometric matrix of the truncated reaction system is checked against conservation using its algebraic properties. Algorithmic procedures are then proposed for checking these criteria with emphasis on good scaling up properties. In addition to these verification procedures, we also provide, for certain typical errors, model correcting methods. The capabilities and usefulness of these procedures are illustrated on biochemical models taken from the Biomodels database. In particular, a set of 11 kinetic models related with *E. coli* are checked, finding two with deficiencies. Correcting actions for these models are proposed.

Keywords: verification, model checking, kinetic models

1 Introduction

Dynamics play a key role in the explanation of complex phenomena occurring in living systems. Therefore, the dynamic modeling and model analysis of biochemical networks is of high importance in systems biology, as quantitative mathematical models allow the description, analysis and/or manipulation of a wide range of biochemical processes.

The mathematical form of these models varies depending on the aim of modeling and on the quality of measured data available. Petri-net models –both deterministic and stochastic– are widely used for the analysis of qualitative dynamic properties, such as persistency [1], stability [2], etc. Qualitative dynamic

models in the form of nonlinear ordinary differential equations (ODEs) are also widely used when good quality measured data are available for model parameter estimation, model verification, validation and detailed dynamic analysis. More in particular, the class of kinetic models [3] (with mass action type or other rate functions) is a widely accepted description form.

In practice, however, many of the medium and large-scale kinetic models in systems biology show problems when the space of parameters is explored. For example, dynamic simulations for certain parameter values result in negative concentrations (suggesting that mass-balance may not be correct), or simply blow-up. Therefore, careful checks should be performed before the use of a published model. This is routinely done in large biochemical model bases (see e.g. [4]), but these checks cannot detect every deficiency that may arise from the many different uses (simulation, parameter estimation, experiment design, etc.) of these models.

A biologically valid model should be valid both from physical and chemical point of view. There are some tools which help the user to avoid making modelling mistakes e.g. by offering predefined rate-functions, tracking the variables, or supporting measurement units and their consistency. These tools serve mostly for syntactic checking purposes. Furthermore, some tools can also check fundamental model properties, such as e.g. mass balance, the existence of admissible steady states, or the characteristics of the dynamic behaviour near a steady state, among others.

The Systems Biology Markup Language (SBML) [5] is a kind of accepted “standard”, which offers model *syntax* checking, e.g. checking that the measurement units are correct. Systems Biology Toolbox 2 offers in addition (i) moiety conservation, (ii) steady state calculation, (iii) stoichiometry analysis, and (iv) bifurcation analysis.

As another well-known example, COPASI [6] provides a systematic model building framework to reduce the possibility of making modelling errors. Further functions are: steady state analysis, mass conservation, time scale separation, sensitivity calculation etc.

Despite the above efforts to ensure the acceptable quality of a biochemical model, it is easy to find in the literature such models that do not possess very basic properties, like positivity. This is usually the consequence of model simplification based on assumptions [7]. However these assumptions are sometimes forgotten or not known explicitly. Therefore, our aim was to formulate simple syntactical and semantic criteria of biochemical origin that ensure the plausibility of the studied model and the positivity (more precisely, non-negativity) of its solution. Similar ideas of model checking appear in [8] and [9].

The basic properties of dynamic models that describe reaction kinetic systems are used for this purpose. Roughly speaking, the kinetic property of these models means that the individual reaction rates are non-negative and there cannot be negative cross-effects between the dynamics of species [10]. Besides other important features, the kinetic property implies non-negativity which means that the non-negative orthant of the coordinates system remains invariant for the

process dynamics (i.e. the differential variables that typically describe concentrations, always remain non-negative). However, as it is illustrated in this paper, some models published in journals and/or in open access biological databases do not fulfill fundamental kinetic properties, and this can be a serious obstacle in tasks such as parameter estimation, or in the later use or extension of these models.

In addition to the model verification procedures, we aim at localizing the reaction, or set of reactions, that cause a particular problem (for example, possible negative solutions), and at giving advice on how to correct them. Note that verification in this paper is used in the sense of checking for the presence of incorrect dynamic behaviour of the model, and not in the sense of its experimental (in)validation.

2 Plausible Biochemical Models

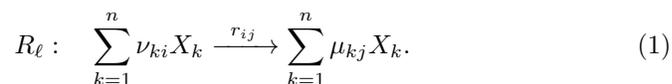
2.1 Mathematical Models of Biochemical Reactions

Biochemical reactions form an important sub-class in chemical reaction kinetics, that are characterized by the generally large number of reaction steps, and by the potentially complex, e.g. non-monotonous nature of the reaction rate functions. The reaction scheme together with the appropriate reaction constants of the most important biochemical reaction systems is collected in large biochemical databases (such as the Biomodels database [4]), and special description languages (such as SBML [5]) are developed for their standardized representation.

In order to develop a model representation of biochemical reaction systems that is suitable for model verification, the model representation of chemical reaction networks [3] can be used with some adjustments.

2.2 Basic Notions for Describing Biochemical Reactions

Complex biochemical reaction schemes are composed of elementary reaction steps that are irreversible. This means, that a reversible reaction step is represented as two irreversible elementary reaction steps. An elementary reaction step R_ℓ can be formally described using n species X_1, \dots, X_n and associated *stoichiometric coefficients*. The species are classified as reactants (with stoichiometric coefficients ν_1, \dots, ν_n) and products (with μ_1, \dots, μ_n).



The non-negative linear combinations of the species $\sum_{k=1}^n \nu_{ki} X_k$ and $\sum_{k=1}^n \mu_{kj} X_k$ are called the *complexes* and are denoted by C_1, \dots, C_m , e.g. $C_1 = 2X_1 + X_3$.

It is important to note that some species may appear on both sides of a given reaction with the same stoichiometric coefficients ($\nu_{ki} = \mu_{ki}$).

A *reaction* (elementary reaction step) R_{ij} is an ordered pair of complexes $C_i, C_j \in \mathcal{C}$, which means that the *reactant complex* C_i is transformed to the

product complex C_j in the chemical reaction network i.e. $R_{ij} = (C_i, C_j)$. Reactant complexes are also called *source complexes*. To each of the reactions, a *reaction rate function* r_{ij} is associated that may depend on the concentration $[X_i] = x_i$ of any species X_i in the biochemical reaction system. The reaction rate is usually measured in units $[\frac{mol}{s}]$ and shows how many moles of a reactant X_k with $\nu_{ki} = 1$ is used, or how many moles of a product X_ℓ with $\mu_{\ell j} = 1$ is produced by the reaction in one second.

2.3 Plausible Reaction Rate Functions

Because of the above chemical meaning of the reaction rate, the reaction rate function should possess the following properties.

1. *Rate positivity.* As the elementary reaction steps are irreversible and the reaction rate is defined as the rate of the consumption (decrease) of the reactant concentrations, the inequality $r_{ij} \geq 0$ should be fulfilled over the entire domain of the reaction rate function, i.e. for all non-negative concentration values in its argument.
2. *Kinetic dependence.* Reaction rate functions in biochemical reactions include the concentrations of the *reactants* such as substrates, which are consumed in the reaction. Some species concentration may not change in the reaction because the same amount is consumed as produced, i.e. $\nu_{ki} = \mu_{kj}$. Further, the reaction function may include other concentrations that modify the reaction rate either in a catalytic way, or in the form of inhibitors, e.g. the concentration of a product of that reaction. However, one only considers the real *reactants* in the source complex which influence the reaction rate in a dominant way that is described by the notion of kinetic dependence. r_{ij} is said to be *kinetic with respect to the species* in the source complex ($X_k \in C_i$) if

$$r_{ij}(x_k = 0) = 0 \quad \text{for all } k = \{1, \dots, n | X_k \in C_i\}, \quad (2)$$

i.e., if the concentration of any species in the source complex is zero, then the reaction rate becomes zero.

2.4 Plausibility of Some Common Biochemical Reaction Rate Functions

Only a limited number of rate function types are usually present in biochemical reaction systems, that are characterized by a functional form and the values of its parameters [11]. A few of the most important ones are analysed for plausibility below.

(i) *Mass action kinetics*

This is the simplest reaction rate function form $r_{MA,i} = k_i \cdot \prod_{l=1}^n x_l^{\nu_{li}}$ where $k_i > 0$ is the reaction rate constant, and the reactant complex is $C_i = \sum_{l=1}^n \nu_{li} X_l$. It is easy to see that $r_{MA,i}$ is kinetic in each of the species in complex C_i .

(ii) *Michaelis-Menten kinetics*

Recall, that elementary reaction steps are irreversible, then the rate function is in the form

$$r_{\text{MM},i} = k_i \cdot \frac{x_i}{(K_i + x_i)} \quad (3)$$

where where $k_i > 0$ and $K_i > 0$ are constant parameters, and the reactant complex $C_i = X_i$. This reaction rate function is kinetic in X_i .

(iii) *Constant level reactions*

Here the rate function is simply a constant, i.e. $r_{C,i} = k_i^M$, where $k_i^M > 0$ is a constant. This rate function does not have kinetic dependence on any specie, thus no reactant species can be associated to this reaction. Consequently *it is not a plausible reaction rate function*, whenever it is a consuming reaction. Note that, when this reaction stands for model input it always occurs with positive sign in the balance equations.

Correcting Non-plausible Reaction Rates. There is unfortunately no general way of correcting non-plausible reaction rates. However, in some cases, such rates can be made plausible. An example of this case is, when a constant level type reaction rate function is present in the kinetic equation of the species X_i with negative sign. Then we can multiply the rate function with x_i that will make this rate function kinetic in X_i .

2.5 Positive (Non-negative) Kinetic Models

The dynamic variables x_k of any biochemical model are species concentrations, that are non-negative. Therefore, any plausible biochemical model should have this property, that is based mathematically on the notion of essentially non-negative functions [12]. A function $f = [f_1 \dots f_n]^T : [0, \infty)^n \mapsto \mathbb{R}^n$ is called essentially non-negative if, for all $i = 1, \dots, n$, $f_i(x) \geq 0$ for all $x \in [0, \infty)^n$, whenever $x_i = 0$. In the context of biochemical models, where the components f_i correspond to the right-hand sides of the kinetic differential equations, the non-negativity of individual rate functions and the lack of negative cross-effects between species together guarantee essential non-negativity of the model [10].

2.6 Component Mass Conservation

Kinetic models are constructed based on the conservation of the masses of species assuming *closed systems* and isothermal conditions.

The conservation equations are constructed for species that are either reactants or products of the chemical reactions in the form

$$\frac{dx_k}{dt} = - \sum_{i=1}^m \nu_{ki} r_i + \sum_{i=1}^m \mu_{ki} r_i = \sum_{i=1}^m s_{ki} r_i, \quad (4)$$

where s is the element of the $S \in \mathbb{R}^{n \times m}$ stoichiometric matrix. No dynamic conservation equations are written to species with only catalytic or inhibitory role.

In open systems one has in addition (i) *input terms*, that have positive sign and may depend on externally set concentrations and/or mass flow of certain non-conserved specie, and (ii) *output terms*, that are linear in one conserved specie, have negative sign and appear only in the dynamic equation of that particular specie. Therefore, *all of the input and output terms should be set to zero when checking the conservation property*: this form of the dynamic model will be called the *truncated model*.

A *truncated stoichiometric matrix* $\tilde{S} \in \mathbb{R}^{n \times m}$ is constructed from the truncated model by associating a column S_i to each complex C_i with $[\tilde{S}]_{ki} = \mu_{ki} - \nu_{ki}$ for only the reactant species (but not to the catalytic or inhibitory ones). The truncated biochemical model has the conservation property taking into consideration all species, if there exists a strictly positive vector m such that $m^T \tilde{S} = 0$ (see [13], [14] and for efficient computation methods [15]).

Some biological models do not obey mass conservation on purpose, otherwise the above property enables us *to check the truncated stoichiometric matrix \tilde{S} against conservation*, that is only a partial verification of the values of the stoichiometric coefficients μ_{ki} and ν_{ki} in the model.

Plausible Model Structure. The model structure is said to be plausible, when the stoichiometric constants in the conservation equations (4) are consistent with the reactants and products of the reactions, i.e. ν_{ki} is strictly positive if reaction r_i consumes the species X_k and μ_{ki} is strictly positive if X_k is a product of the reaction r_i . The stoichiometric coefficient of a reaction which neither consumes nor produces a species should be zero in the corresponding balance equation.

3 Model Checking and Correction in Practice

3.1 Steps of Model Verification

Given a biochemical reaction network in terms of the reaction rate functions and the system of ordinary differential equations. The reaction rate functions are assumed to be smooth functions of the time, some concentrations and parameters: $r_i = r_i(t, x, k)$. The explicit time dependency of the reactions permits to incorporate boundary conditions or model inputs for the dynamic system. The ordinary differential equation form of the model is given by Eq. (4).

Inputs of the Algorithm. We can either start with the list of differential equations and the algebraic equations of the reactions, or the model defined in SBML. Since the SBML model does not contain explicitly the differential equations, in this case the SBML import function of the System Biology Toolbox 2 [16] is used to translate the SBML into MATLAB structure and generate the differential equations. It is important to note that the parameter values of the rate functions are not needed for the verification.

A homogeneous, continuous flow stirred tank bioreactor serves as an tutorial example depicted in Fig. 1. The reaction network consists of three species (A ,

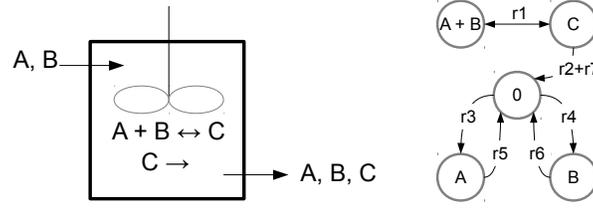


Fig. 1. Continuous flow stirred tank reactor and its reaction graph representation

B , C and their concentrations x_A , x_B and x_C , respectively) and two elementary reactions: a two substrate, one product reversible Michaelis-Menten kinetics (5) and a non-plausible (see subsection 2.4 (iii)) constant reaction (6). The zero complex denoted by a "0" in the reaction graph corresponds to the environment of the system (for clarification see e.g. [17]). The reactor feed contains the substrates A and B with x_A^f and x_B^f constant concentrations, respectively, the corresponding pseudo-reactions [17] are in Eq. (7). The output stream that contains all species is represented by the pseudo-reactions in Eq. (8). The reaction rate functions and the ODEs of the system are

$$r_1 = V_f \frac{\frac{x_A}{K_{x_A}} \frac{x_B}{K_{x_B}}}{1 + \frac{x_A}{K_{x_A}} + \frac{x_B}{K_{x_B}}} - V_r \frac{\frac{x_C}{K_{x_C}}}{1 + \frac{x_C}{K_{x_C}}} \quad (5)$$

$$r_2 = K_d \quad (6)$$

$$r_3 = \zeta x_A^f; \quad r_4 = \zeta x_B^f \quad (7)$$

$$r_5 = \zeta x_A; \quad r_6 = \zeta x_B; \quad r_7 = \zeta x_C \quad (8)$$

$$\frac{dx_A}{dt} = -r_1 + r_3 - r_5 \quad (9)$$

$$\frac{dx_B}{dt} = -r_1 + r_4 - r_6 \quad (10)$$

$$\frac{dx_C}{dt} = r_1 - r_2 - r_7 \quad (11)$$

Splitting the Reversible Reactions. The irreversible forward and backward reactions are created from the original reactions using regular expressions and the Symbolic Math Toolbox of MATLAB. In this example the algorithm finds the subtraction with two operands in Eq. (5) and separates to

$$r_1^f = V_f \frac{\frac{x_A}{K_{x_A}} \frac{x_B}{K_{x_B}}}{1 + \frac{x_A}{K_{x_A}} + \frac{x_B}{K_{x_B}}} \quad \text{and} \quad r_1^b = V_r \frac{\frac{x_C}{K_{x_C}}}{1 + \frac{x_C}{K_{x_C}}} \quad (12)$$

Simultaneously the differential equations are updated to

$$\frac{dx_A}{dt} = -(r_1^f - r_1^b) + r_3 - r_5 \quad (13)$$

$$\frac{dx_B}{dt} = -(r_1^f - r_1^b) + r_4 - r_6 \quad (14)$$

$$\frac{dx_C}{dt} = (r_1^f - r_1^b) - r_2 - r_7 \quad (15)$$

Model Positivity by Checking the Kinetic Property. Next, the stoichiometric matrix (S) is constructed by parsing the string of the ODEs and collecting the coefficients of the rate functions. Whenever the v_{ij} element of S is negative, i.e. reaction r_j consumes the species x_i , r_j must be kinetic with respect to x_i . This can be checked by substituting zeros for the species x_i in the rate function and evaluating it; the result must be zero.

In our example the model Eqs. (13)-(15) give rise to the stoichiometric matrix

$$S = \begin{bmatrix} -1 & 1 & 0 & 1 & 0 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \quad (16)$$

and the irreversible reaction vector $R = [r_1^f, r_1^b, r_2, r_3, r_4, r_5, r_6, r_7]^T$. Considering the location of the negative entries of S reaction r_1^f and r_5 must be kinetic to species A , r_1^f and r_6 with respect to B and r_1^b, r_2 and r_7 with respect to C . By substituting zeros for the reactant species in the rate functions –e.g. $r_1^f(x_A = 0)$, $r_5(x_A = 0)$, $r_1^f(x_B = 0)$ etc. – the plausible ones give zeros. At this point reaction r_2 is found to be non-kinetic to the species C , and thus it is a non-plausible reaction. We may correct the rate function by multiplying with its reactant species concentration: $r_2^* = K_d x_C$. This reaction can be regarded as a model output, too.

Component Mass Conservation. The truncated model without the input reactions (Eqs. (7)) and the output reactions (Eqs. (8) and the corrected r_2^*) is represented by the first two columns of S . This sub-matrix is rank deficient and has the $m = [1 \ 1 \ 2]^T$ strictly positive vector in the left nullspace indicating the mass conservation law.

3.2 Verified Models

We have checked 11 *E. coli* curated models of the Biomedb database, and some of them turned out to contain non-plausible reactions. Table 1 shows the unique identifiers of the models in the database. The number of species, the number of reactions and the computation time of the algorithm is also included in the following columns. The 5th column contains the non-plausible reaction, while the last column shows whether the truncated model admits mass conservation. In the next section, the verification of two of these models will be presented in detail.

Table 1. Verified models

BioModel ID	No. of species	No. of reactions	Time [s]	Non-plausible reaction	Mass conservation
BIOMD296	4	10	man.*	plausible	no
BIOMD413	5	9	0.3	plausible	no
BIOMD200	22	46	2.3	plausible	yes
BIOMD217	12	22	23	plausible	yes
BIOMD051	18	62	5	reaction vMURSYNTH is not kinetic w.r.t. species CF6P	no
BIOMD066	11	10	man.*	reaction vATPASE is not kinetic w.r.t. species ATP	yes
BIOMD012	6	12	0.8	plausible	no
BIOMD067	7	16	0.6	plausible	no
BIOMD221	8	22	1.9	reaction vSYN is not kinetic w.r.t. species AKG	no
BIOMD222	8	22	1.9	reaction vSYN is not kinetic w.r.t. species AKG	no
BIOMD065	8	16	0.5	plausible	no

*the separation of some reaction rate function needed manual manipulation

3.3 Case Study 1: Central Carbon Metabolism of *E. coli*

Chassagnole et al. [18] describe the central carbon metabolism of the *Escherichia coli*. Although we could reproduce the results in the paper [18] with the published model (BIOMD0000000051), numerical simulations with CVODES [19] during parameter estimation tasks gave errors, because negative concentrations appeared.

About the Model. The metabolism is described by 48 reactions which are grouped into kinetic types: reversible and irreversible Michealis-Menten kinetics, two-substrate reversible and irreversible Michaelis-Menten kinetics, allosteric enzyme reactions, allosteric regulation, allosteric activation, ordered uni-bi mechanism, Hill kinetics, constant level reaction and reversible mass action kinetics. Appendix A contains examples of these reactions. The mass balance equations for the 18 metabolites are in the following form

$$\frac{dC_i}{dt} = \sum_j v_{ij} r_j(C, k) - \mu C_i, \quad (17)$$

where C is the concentration of the metabolite, v_{ij} is the (i, j) th element of the stoichiometric matrix, $r_j(C, k)$ denotes the j -th reaction rate function, which depends on the concentrations and the k rate function parameters. Finally, μ is the growth factor. The detailed equations are listed in [18] Tables I. and IV.

Checking the Rate Expressions. The first criteria of a plausible model is the non-negativity of the reaction rate functions, for which the reversible reactions have to be cut into a forward and a backward reaction. The separation of

the reactions are straightforward in this case study. It is also easy to see that the reactions are always non-negative since the rate expressions contain only such mathematical operators that preserve the positivity. The model have three constant reactions: the Mureine synthesis, the Tryptophan synthesis and the Methionine synthesis

$$r_{\text{MurSynth}} = r_{\text{MurSynth}}^{\max}, \quad (18)$$

$$r_{\text{TrpSynth}} = r_{\text{TrpSynth}}^{\max} \quad \text{and} \quad r_{\text{MetSynth}} = r_{\text{MetSynth}}^{\max}, \quad (19)$$

but only r_{MurSynth} is not kinetic to its source specie, the others are input terms.

Checking the Model Structure. The positivity condition from Section 2.5 together with the model Eq. (17) give rise to the model specific positivity condition

$$\frac{dC_i}{dt} = \sum_j v_{ij} r_j - \mu C_i \geq 0 \quad \text{whenever} \quad C_i = 0, \quad \text{for all} \quad i = 1, \dots, 18.$$

This condition holds for plausible reaction rate functions which have the source kinetic property according to Eq. (2). Furthermore, whenever a reaction is not kinetic w.r.t. its source species and has negative stoichiometric coefficient, depending on the numerical values of the parameters it can violate the condition and cause negative concentrations during simulations.

This is exactly what happens in some parameter domain of this *E.coli* model. From the following model equation ([18] Table I. Eq.(3)):

$$\frac{dC_{\text{f6p}}}{dt} = r_{\text{PGI}} - r_{\text{PFK}} + r_{\text{TKb}} + r_{\text{TKa}} - 2r_{\text{MurSynth}} - \mu C_{\text{f6p}} \quad (20)$$

one can see that the stoichiometric coefficient of the Mureine synthesis r_{MurSynth} is negative, but it is not kinetic w.r.t. any metabolite. This may result in the appearance of negative concentrations and thus in a non-plausible model.

Correction of the Non-plausible Reaction. There are several ways to correct the non-plausible reaction. A switching function can be included, which turns off the reaction, whenever the concentration of **f6p** reach zero. This procedure does not influence the model dynamics in the plausible concentration domain, but the switching function may result in mathematical or numerical simulation issues. Alternatively, one can make the reaction source kinetic by multiplying it with C_{f6p} : $r_{\text{MurSynth}}^{\text{cured}} = r_{\text{MurSynth}}^{\max} C_{\text{f6p}}$. It changes the dynamics of the system, but results in a smooth, plausible reaction rate function.

Mass Conservation. The truncated model is created by omitting the reactions which are either stand for inflows or outflows. We have found three linearly independent non-negative vectors for which $m_i^T S = 0$, for $i = 1, 2, 3$. This implies three moiety conservation laws, but there is no strictly positive m vector in the left kernel of S , and thus the model does not obey to the total mass conservation.

3.4 Case Study 2: Verification of the Model BIOMD0000000221

Singh et al. [20] present two kinetic models of the tricarboxylic acid cycle and glyoxylate bypass in the *Mycobacterium tuberculosis*. Both models are based on a validated *E. coli* model, which is in the focus of this case study. The kinetic model contains 12 metabolites and 12 reactions. The reactions and the differential equations are listed in Appendix B.

The reaction rate functions can be categorized into three kinetic reaction types: one substrate reversible Michaelis-Menten kinetics; two substrate reversible Michaelis-Menten kinetics and ordered uni-bi mechanism. The separation of them to irreversible forward and backward reactions is straightforward. The irreversible forms fulfil the rate positivity condition.

For the positivity of the model the kinetic property of the rate expressions should be analysed. The reactions must be kinetic to the species, which are consumed in that reactions. However, due to a modelling assumption, the r_{11} reaction

$$r_{11}^f = 0.0341 \cdot r_3^f = V_{\text{cell}} \cdot 0.0341 \frac{V_{11}^f \frac{C_{\text{icit}}}{K_{11,\text{icit}}}}{1 + \frac{C_{\text{icit}}}{K_{11,\text{icit}}} + \frac{C_{\text{akg}}}{K_{11,\text{akg}}}} . \quad (21)$$

is not kinetic with respect to the **akg** source specie, which is consumed in this reaction according to the balance equation:

$$\frac{dC_{\text{akg}}}{dt} = r_3^f - r_3^b - r_4^f + r_4^b - r_{11}^f + r_{11}^b \quad (22)$$

Note that r_{11}^f has a negative stoichiometric coefficient.

Thus, our algorithm detected the consequence of the modelling assumption which lead the model out from the kinetic model system class.

The model has four boundary species, the concentrations of which are hold constant. The omission of the reactions containing these species leads to the truncated model and the corresponding reduced stoichiometric matrix S . There is no strictly positive vector in the left null-space of S : actually, all species but the glyoxylate participate in the mass conservation.

4 Conclusion

Using the syntax and semantics of biochemical models, simple syntactical criteria were formulated in this paper that ensure the plausibility of the studied model and the positivity of its solution.

First, the plausibility of reaction rate function was defined that include its positivity, and its kinetic dependence on the real reactants of the reaction. The absence of the negative cross-effects in the dynamic equations were used to ensure the positivity of the species concentration functions. The stoichiometric matrix of the truncated reaction system was checked against conservation using its algebraic properties.

Algorithmic procedures were proposed for checking these criteria that scale up well with the size of the biochemical model. For certain typical errors, model correcting methods were also proposed.

The developed notions and tools are illustrated on biochemical kinetic models of *E. coli*.

Acknowledgement. This research was supported in part by the Hungarian Research Fund through grant K83440 and by the funding from EU FP7 ITN "NICHE", project no. 289384.

References

1. Angeli, D., De Leenheer, P., Sontag, E.: A Petri net approach to the study of persistence in chemical reaction networks. *Mathematical Biosciences* **210** (2007) 598–618
2. Sontag, E.: Structure and stability of certain chemical networks and applications to the kinetic proofreading model of T-cell receptor signal transduction. *IEEE Transactions on Automatic Control* **46**(7) (2001) 1028–1047
3. Feinberg, M.: On chemical kinetics of a certain class. *Arch. Rational Mech. Anal.* **46** (1972) 1–41
4. Li, C., Donizelli, M., Rodriguez, N., Dharuri, H., Endler, L., Chelliah, V., Li, L., He, E., Henry, A., Stefan, M.I., Snoep, J.L., Hucka, M., Le Novère, N., Laibe, C.: BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology* **4:92** (2010)
5. Hucka, M., et al.: The Systems Biology Markup Language SBML: A medium for representation and exchange of biochemical network models. *Bioinformatics* **19** (2003) 524–531
6. Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., Kummer, U.: COPASI a COmplex PATHway SIMulator. *Bioinformatics* **22**(24) (2006) 3067–3074
7. Hangos, K.M.: Engineering model reduction and entropy-based Lyapunov functions in chemical reaction kinetics. *Entropy* **12** (2010) 772–797
8. Fages, F., Gay, S., Soliman, S.: Automatic curation of SBML models based on their ODE semantics. Technical report, INRIA Research Report 8014, hal00723554 (2012)
9. Kaleta, C., Richter, S., Dittrich, P.: Using chemical organization theory for model checking. *Bioinformatics* **25**(15) (2009) 1915–22
10. Érdi, P., Tóth, J.: *Mathematical Models of Chemical Reactions. Theory and Applications of Deterministic and Stochastic Models.* Manchester University Press, Princeton University Press, Manchester, Princeton (1989)
11. Klipp, E., Herwig, R., Kowald, A., Wierling, C., Lehrach, H.: *Systems biology in practice: concepts, implementation and application.* Wiley-Blackwell (2008)
12. Haddad, W.M., Chellaboina, V., Hui, Q.: *Nonnegative and Compartmental Dynamical Systems.* Princeton University Press (2010)
13. Famili, I., Palsson, B.O.: The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. *Biophysical Journal* **85**(1) (2003) 16–26

14. Hangos, K.M., Szederkényi, G.: The underlying linear dynamics of some positive polynomial systems. *Physics Letters A* **376** (2012) 3129–3134
15. Sauro, H.M., Ingalls, B.: Conservation analysis in biochemical networks: computational issues for software writers. *Biophysical Chemistry* **109**(1) (2004) 1–15
16. Schmidt, H., Jirstrand, M.: Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics* **22**(4) (2006) 514–515 <http://www.sbtoolbox2.org>.
17. Feinberg, M.: Chemical Reaction Network Structure and the Stability of Complex Isothermal Reactors – I. The deficiency Zero and Deficiency One Theorems. *Chemical Engineering Science* **42**(10) (1987) 2229–2268
18. Chassagnole, C., Noisommit-Rizzi, N., Schmid, J.W., Mauch, K., Reuss, M.: Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnology and Bioengineering* **79**(1) (2002) 53–73
19. Hindmarsh, A.C., Brown, P.N., Grant, K.E., Lee, S.L., Serban, R., Shumaker, D.E., Woodward, C.S.: SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software (TOMS)* **31**(3) (2005) 363–396
20. Singh, V.K., Ghosh, I.: Kinetic modeling of tricarboxylic acid cycle and glyoxylate bypass in *Mycobacterium tuberculosis*, and its application to assessment of drug targets. *Theoretical Biology and Medical Modelling* **3:27** (2006)

Appendix

A Reaction Rate Functions of the First Case Study

Some examples of the reaction rate functions and their irreversible form can be found in the following list.

1. Reversible mass action kinetics, e.g. Ribose phosphate isomerase reaction

$$r_{R5P1} = r_{R5P1}^{max} \left(C_{ribu5p} - \frac{C_{rib5p}}{K_{R5P1,eq}} \right),$$

the forward and backward reactions of which are:

$$r_{R5P1}^f = r_{R5P1}^{max} C_{ribu5p} \quad \text{and} \quad r_{R5P1}^b = r_{R5P1}^{max} \frac{C_{rib5p}}{K_{R5P1,eq}}.$$

2. Irreversible Michaelis-Menten kinetics, for example Serine synthesis

$$r_{SerSynth} = \frac{r_{SerSynth}^{max} C_{3pg}}{K_{SerSynth,3pg} + C_{3pg}}.$$

3. Allosteric enzyme activation, for example Glucose 1-phosphate adenylyltransferase reaction

$$r_{GIPAT} = \frac{r_{GIPAT}^{max} C_{glp} C_{atp} \left(1 + \left(\frac{C_{fdp}}{K_{GIPAT,fdp}} \right)^{n_{GIPAT,fdp}} \right)}{(K_{GIPAT,glp} + C_{glp})(K_{GIPAT,atp} + C_{atp})}$$

4. Constant level reaction, such as Mureine synthesis

$$r_{MurSynth} = r_{MurSynth}^{max}$$

B Model Equations of the Second Case Study

This section contains the reaction rate functions and the dynamic model equations of the tricarboxylic acid cycle and glyoxylate bypass model of *E. coli*. The 1st and 10th reactions have two substrate reversible Michaelis-Menten kinetics:

$$r_i = V_{\text{cell}} \frac{V_i^f \frac{S_1}{K_{i,S1}} \frac{S_2}{K_{i,S2}} - V_i^r \frac{P_1}{K_{i,P1}} \frac{P_2}{K_{i,P2}}}{\left(1 + \frac{S_1}{K_{i,S1}} + \frac{P_1}{K_{i,P1}}\right) \left(1 + \frac{S_2}{K_{i,S2}} + \frac{P_2}{K_{i,P2}}\right)} \text{ for } i = \{1, 10\} ,$$

where S_- and P_- denote the concentrations of the substrates and products respectively, $K_{-,}$ are constant parameters and $V_-^{r/b}$ are the maximal rates of forward and backward reactions.

The 2nd – 8th and 11th reactions belong to the one substrate Michaelis-Menten kinetics

$$r_i = V_{\text{cell}} \frac{V_i^f \frac{S_1}{K_{i,S1}} - V_i^r \frac{P_1}{K_{i,P1}}}{1 + \frac{S_1}{K_{i,S1}} + \frac{P_1}{K_{i,P1}}} \text{ for } i = \{2, \dots, 8, 11\} .$$

Finally, the 9th reaction has the form:

$$r_i = V_{\text{cell}} \frac{V_i^f \frac{S_1}{K_{i,S1}} - V_i^r \frac{P_1}{K_{i,P1}} \frac{P_2}{K_{i,P2}}}{\left(1 + \frac{S_1}{K_{i,S1}} + \frac{P_1}{K_{i,P1}} + \frac{P_2}{K_{i,P2}} + \frac{S_1}{K_{i,S1}} \frac{P_1}{K_{i,P1}} + \frac{P_1}{K_{i,P1}} \frac{P_2}{K_{i,P2}}\right)} \text{ for } i = 9 .$$

The system of differential equations expressed in terms of the reactions is

$$\begin{aligned} \frac{dC_{\text{aca}}}{dt} &= 0, & \frac{dC_{\text{aaa}}}{dt} &= 0, & \frac{dC_{\text{coa}}}{dt} &= 0, & \frac{dC_{\text{biosyn}}}{dt} &= 0 \\ \frac{dC_{\text{cit}}}{dt} &= r_1^f - r_1^b - r_2^f + r_2^b \\ \frac{dC_{\text{icit}}}{dt} &= r_2^f - r_2^b - r_3^f + r_3^b - r_9^f + r_9^b \\ \frac{dC_{\text{akg}}}{dt} &= r_3^f - r_3^b - r_4^f + r_4^b - r_{11}^f + r_{11}^b \\ \frac{dC_{\text{sca}}}{dt} &= r_4^f - r_4^b - r_5^f + r_5^b \\ \frac{dC_{\text{suc}}}{dt} &= r_5^f - r_5^b + r_9^f - r_9^b - r_6^f + r_6^b \\ \frac{dC_{\text{fa}}}{dt} &= r_6^f - r_6^b - r_7^f + r_7^b \\ \frac{dC_{\text{mal}}}{dt} &= r_7^f - r_7^b + r_{10}^f - r_{10}^b - r_8^f + r_8^b \\ \frac{dC_{\text{gly}}}{dt} &= r_9^f - r_9^b - r_{10}^f + r_{10}^b \end{aligned}$$