Noname manuscript No.
(will be inserted by the editor)

Reaction Network Realizations of Rational Biochemical Systems and Their Structural Properties

Attila Gábor · Katalin M. Hangos · Julio R. Banga · Gábor Szederkényi

Received: date / Accepted: date

Abstract In this paper, a frequently used representation of mass-action type reaction networks is extended to a more general system class where the reaction rates are in rational function form. An algorithm is given to compute a possible reaction graph from the kinetic differential equations. However, this structure is generally non-unique, as it is illustrated through the phenomenon of dynamical equivalence, when different reaction network structures correspond to exactly the same dynamics. It is shown that under some technical assumptions, the so-called dense realization containing the maximal number of reactions, forms a super-structure in the sense that the reaction graph of any dynamically equivalent reaction network is the sub-graph of the dense realization. Additionally, optimization based methods are given to find dynamically equivalent realizations with preferred properties, such as dense - or sparse realizations. The introduced concepts are illustrated by examples.

Keywords dynamic models · biochemical reaction graph · dynamic equivalence · parameter-free model analysis

Attila Gábor and Julio R. Banga
BioProcess Engineering Group, Instituto de Investigaciones Marinas, CSIC, Vigo, Spain
E-mail: {attila.gabor,julio}@iim.csic.es

Katalin M. Hangos
Process Control Research Group, Computer and Automation Research Institute, Budapest, Hungary
Department of Electrical Engineering and Information Systems, University of Pannonia, Veszprém, Hungary
E-mail: hangos@scl.sztaki.hu

Gábor Szederkényi
Faculty of Information Technology and Bionics, Pázmány Péter Catholic University,
Process Control Research Group, Computer and Automation Research Institute, Budapest, Hungary
E-mail: szeder@scl.sztaki.hu

1 Introduction

The fundamentals of chemical reaction network theory (CRNT) were established back in the 1970s by [17] and [8]. This theory aims to consider complex reaction systems from a general perspective, linking the structure of the reaction network with its possible dynamics (such as unstable and/or multiple steady states, oscillations or chaotic behaviour). CRNT uses powerful descriptors, like the reaction graph and its properties. For example the deficiency concept is used to characterize structural, (i.e. parameter independent) dynamic properties (such as stability) of the network. The dynamic model of a CRN is considered in the form of a positive polynomial ordinary differential equation (ODE) that is assumed to be kinetic, and then an algorithm can be used to construct the so called canonical realization of the CRN [15]. Later, the notion of dynamic equivalence has appeared and it became apparent that possibly several equivalent reaction kinetic schemes, i.e. reaction network realizations can be constructed to a given dynamic ODE model by using optimization, see e.g. [30,32]. These realizations offer the possibility to analyse or ensure advantageous dynamic property to a CRN if a suitable realization with desirable structural properties (e.g. zero deficiency and weak reversibility) is found.

It is worth highlighting that although Horn and Jackson [17] already mentioned complex reactions in biological systems as one of the general situations they wanted to embrace with their theory, its applications in biology did not appear until Bailey [1] rescued it, highlighting its potential for the analysis of biochemical networks without calibrating the model with experimental data ("complex biology with no parameters"). In other words, CRNT can be used to characterize kinetic models (multi-stability, oscillations, etc.) without knowing the precise values of the kinetic parameters.

Several important applications regarding the characterization of the dynamics of biochemical reaction networks have appeared since then, including the works of e.g. [3,4,34,24,28,29,21,27,22].

Furthermore, the theory has also been applied to the identification of biological systems. For example, Craciun and Pantea [5] used CRNT to show that, given a (mass action) reaction network and its dynamic equations, it might be impossible to identify its rate constants uniquely (even with perfect measurements of all species). Furthermore, they also concluded that, given the dynamics, it might be impossible to identify the reaction network uniquely.

In [30,32] and [31], CRNT principles were used to pinpoint inherent limitations in the inference of biological networks. These works show that, in addition to the obstacles identified by [25] (lack of data and deficiencies in the inference algorithms), there are fundamental problems related to the uniqueness and distinguishability of these networks. Further, these problems are present even for the utopian case of fully observed networks with noiseless measurements.

Despite of the above mentioned works in the area of biochemical reaction networks (bio-CRNs), no systematic attempt has been made to construct and analyse the structure of bio-CRNs, to characterize their canonical structure as a subset of positive rational ODEs [23], and to link their structural properties to the dynamic properties of the underlying biochemical system.

Therefore, our general aim is to extend the well-known formalism of the chemical reaction networks obeying the mass action law (MAL-CRN) for general bio-

chemical networks where the reaction rate functions often account for more complex mechanism than the simple mass action law, such as the Michaelis-Menten kinetics, Hill kinetics, activating or inhibiting mechanisms. A canonical decomposition of these reaction rate functions being in rational function form lets us define the main structural elements of the reaction network and the complex – reaction graph. The MAL-CRN case then becomes a special case of this biochemical reaction network.

This way the basic structural (i.e. parameter-independent) properties of a bio-CRN structure are easy to define and understand, that include reversibility, weak reversibility, and deficiency. The ordinary differential equations describing the dynamics of the biochemical reaction network (i.e. the concentrations of the species in the network) can be stated in a form that is similar to the MAL-CRN formalism. This new form of the equations let us easily analyse certain properties of the network, for example to prove the non-negativity of the solutions. Furthermore, the proposed structure allows to algorithmically infer a network from the ODEs, and determine alternative bio-CRN structures that are dynamically equivalent to the original network, i.e. although different reactions exist in the networks, the concentrations of the compounds evolve along exactly the same trajectories.

The paper is organised as follows. In Section 2 the structural elements of the biochemical reaction networks are defined and some properties are shown. In Section 3 the kinetic realizability conditions and a canonical realization algorithm is presented for ODEs in rational polynomial form. Section 4 is devoted to find dynamically equivalent realizations with preferred structure of a given network. Finally, Section 5 summarizes and concludes the work.

2 The basic structure of biochemical reaction networks

The classical theory of reaction networks assumes a closed thermodynamic system with constant physico-chemical properties under isothermal and isobaric conditions, where chemical species X_i, $i = 1, \dots n$ take part in chemical reactions of the form

$$\sum_{k=1}^{n} \nu_{ki} X_k \xrightarrow{r_{ij}} \sum_{k=1}^{n} \mu_{kj} X_k \tag{1}$$

with *stoichiometric coefficients* of *reactants* $\nu_{1i}, \dots, \nu_{ni}$ and of *products* $\mu_{1j}, \dots, \mu_{nj}$. Then the specie concentrations $x_i = [X_i]$, $i = 1, \dots n$ form the state vector, the elements of which are non-negative.

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

A finite set of biochemical kinetics $\mathcal{G}_i = \{G_1, G_2, \dots G_{d_i}\}$ is assigned to each complex C_i. A biochemical kinetics is a description that specifies the way the species in the complex can react, for example $G_1 =$ `'Mass action'`, $G_2 =$ `'Michelis Meten kinetics'`, etc.

The reaction rate $r(\cdot) : \overline{\mathbb{R}}_+^n \mapsto \overline{\mathbb{R}}_+$ corresponding to the reaction $C_i \xrightarrow{k_{ijl}, g_{il}} C_j$ is described by

$$r_{ijl}(x) = k_{ijl} \cdot g_{il}(x), \tag{2}$$

where $k \in \mathbb{R}_0^+$ is the non-negative principal reaction rate coefficient and $g_{il}(\cdot)$ is a function associated with the kinetics G_l of the complex C_i. The possible forms and properties of this rate function are described below in subsection 2.1.

To further formalize the above description, similarly to the formalism of MAL-CRNs [9], we characterize bio-CRNs with the following three sets.

1. $\mathcal{S} = \{X_1, \dots X_n\}$ is the set of species or chemical substances.
2. $\mathcal{C} = \{C_1 \dots C_m\}$ is the set of complexes.
3. The set of biochemical reactions is

$$\mathcal{R} = \{(C_i, C_j, G_l) \mid C_i, C_j \in \mathcal{C}, G_l \in \mathcal{G}_i \text{ and } C_i \text{ is transformed to } C_j \text{ by the kinetics } G_l\}. \quad (3)$$

The relation $(C_i, C_j, G_l) \in \mathcal{R}$ is denoted by $C_i \xrightarrow{G_l} C_j$. In this case, C_i is called the reactant or source complex, and C_j is the product complex. Further, a non-negative principal reaction rate coefficient is also assigned to each reaction, which will be indicated as $C_i \xrightarrow{k_{ijl}, g_{il}} C_j$. $k_{ijl} = 0$ means that $C_i \xrightarrow{G_l} C_j \notin \mathcal{R}$.

The set of species, complexes and reactions with the kinetics uniquely determines the biochemical reaction network which is denoted by $\Sigma = (\mathcal{S}, \mathcal{C}, \mathcal{G}, \mathcal{R})$.

In the special case, when the chemical reaction network contains only mass-action type kinetics, the set of kinetics becomes meaningless and the traditional $\Sigma_{\text{MA}} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ is obtained.

2.1 Biochemical kinetics associated functional forms

The species in each complex C_i for $i = 1 \dots m$, may react in different ways that is described by the concept of kinetics. A set of kinetics $\mathcal{G}_i = \{G_1, \dots G_{d_i}\}$ is defined for each complex, and a function $g_{il}(\cdot)$ is associated to each kinetics. This kinetic function characterizes the rate of the reaction (2) and can depend on a set of species. Species can be classified as dominant species and modifiers. The dominant species of a kinetics are the species, the concentration of which strongly effects the reaction rate, i.e. if any of the dominant species concentration is zero, the reaction rate is also zero. On the other hand, the positive concentration of modifier species of a given kinetics is not required for a strictly positive reaction rate. We assume that the species in the source complexes are the dominant species with respect to all the kinetics assigned to that complex. In other words, if any of the species has zero concentration in a complex all the corresponding reaction rate functions are also zero. Further, we make the following assumptions:

(KA1) any kinetic function $g_{il}(\cdot)$ is always non-negative,
(KA2) any kinetic function $g_{il}(\cdot)$ is zero if and only if the concentration of any of its source species is zero.

Due to (KA1) the reversible reactions are represented by two reactions.

We further assume that

(KA3) the elementary reaction rate corresponding to the source complex $C_i = \sum_{l=1}^{n} \nu_{li} X_l$ can be stated as a ratio of two functions:

$$r(x) = k \cdot g(x) = k \frac{M_i(x)}{D(x)}, \quad (4)$$

where

$$M_i(x) = \prod_{l=1}^{n} x_l^{\nu_{li}} \geq 0 \tag{5}$$

is a monomial with non-negative integers coefficients ν_{li} of the source complex; k is the principal reaction rate coefficient. Here $D(x)$ is an element of a set $\mathcal{P}$ of multivariate polynomials with the following form

$$D(x) = 1 + \sum \alpha_{m_1,m_2,\ldots m_n} x_1^{m_1} x_2^{m_2} \ldots x_n^{m_n} > 0 \tag{6}$$

where $\alpha \in \overline{\mathbb{R}}_+$ and m_1, m_2, $\ldots$, m_n are non-negative integers.

Similar decomposition of the reaction rates (4) was used for example in [20] for the verification of biochemical dynamic models and in [26] for model reduction.

2.2 Differential equations of the biochemical system

Motivated by the formalism of the MAL-CRNs dynamics (see, e.g. [9]), let us define the dynamics of the bio-CRNs with the following form

$$\dot{x} = Y \cdot \tilde{A}_{\mathrm{k}} \cdot P(x) \cdot \Psi(x), \tag{7}$$

where $Y \in \mathbb{R}^{n\times m}$ is the complex composition matrix; the matrix $\tilde{A}_{\mathrm{k}} \in \mathbb{R}^{m\times\kappa}$ stores the principal reaction rate coefficients and its structure has a close relationship with the graph of the network, the mapping $P : \mathbb{R}^n \mapsto \mathbb{R}^{\kappa\times m}$ is the rate-weighting matrix valued function that contains the information regarding the kinetics of the reactions and $\Psi : \mathbb{R}^n \to \mathbb{R}^m$ is the monomial vector function

$$\Psi_i(x) = \prod_j x_j^{Y_{ji}} \ . \tag{8}$$

The ith column of Y denoted by $\eta^{(i)}$ contains the composition of complex C_i, i.e. Y_{ji} is the stoichiometric coefficient of C_i corresponding to the specie X_j. As the stoichiometric coefficients are non-negative, it may happen that all of them are equal to zero for a certain complex C_0, i.e. the corresponding column in Y is the zero vector ($\eta^{(i)} = 0$). Such a complex C_0 is called the *zero complex*, and it can be used to describe the case when *the system is not closed but has in-/outflow from/to the environment*, as it is a usual situation in biochemical models.

*The rate-weighting matrix.*The kinetic function $g(x)$ in (4) can be seen as state-dependent weighted form of the monomial function $M_i(x) = \Psi_i(x)$ using the weights $1/D(x)$. For each monomial $\Psi_1(x)$, $\Psi_2(x)$, $\ldots \Psi_m(x)$, define a $P_i(x) \in \mathbb{R}^{d_i}$ vector function component-wise, in such a way, that $P_{i\,l}(x) \cdot \Psi_i(x) = g_{il}(x)$ for $l = 1, \ldots d_i$, where $P_{i\,l}$ is the weight in the l-th kinetics and d_i is the number of distinct kinetics associated to complex C_i. Obviously,

$$P_{i\,l}(x) = \frac{1}{D_{il}(x)} \ . \tag{9}$$

When the vector functions $P_i(x)$ are arranged in a block matrix $P \in \mathbb{R}^{\kappa \times m}$ (where the total number of kinetics is $\kappa = \sum_{i=1}^{m} d_i$), this matrix is

$$P = \begin{bmatrix} [P_1] & 0 & \dots & 0 \\ 0 & [P_2] & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & [P_m] \end{bmatrix}, \tag{10}$$

and the product

$$P(x) \cdot \Psi(x) = [g_{11}(x),\, g_{12}(x),\, \dots g_{md_m}(x)]^T := \varphi(x) \tag{11}$$

gives the vector of kinetic functions $\varphi(\cdot) : \overline{\mathbb{R}}_+^n \mapsto \overline{\mathbb{R}}_+^\kappa$, which simply collects the reaction rates without the principal reaction rate coefficients. This vector also fixes the ordering of the complexes and also the ordering of kinetics in each complex, i.e. the first d_1 elements correspond to the kinetics of the first complex, the following d_2 elements correspond to the kinetics in the second complex and so on. To simplify the indexing, let us define the index variable

$$z_i = \sum_{k=1}^{i-1} d_k \quad \text{for} \quad i = 1 \dots m, \tag{12}$$

which denotes the sum of the number of kinetics originating from the first $i-1$ complexes. Thus, the kinetic functions starting from complex C_i can be easily indexed in the kinetic vector (11) as $\varphi_{z_i+1},\, \varphi_{z_i+2} \dots \varphi_{z_i+d_i}$.

*The modified Kirchhoff matrix.*The matrix $\tilde{A}_\mathrm{k} \in \mathbb{R}^{m \times \kappa}$ contains the principal reaction rate coefficients of the reactions and encodes the structure of the reaction graph. Each row of $\tilde{A}_\mathrm{k}$ corresponds to a complex and each column corresponds to a kinetic function.

The principal reaction rate coefficient k_{ijl} of the reaction (C_j, C_i, G_l) –for $i, j = 1, \dots m;\ i \neq j;\ l = 1, \dots d_j$– is located in $\tilde{A}_{\mathrm{k},i\, z_j+l}$, where z_j is defined as (12). Furthermore, the elements $\tilde{A}_{\mathrm{k},i\, z_i+l}$ for $i = 1, \dots m;\ l = 1, \dots d_i$ contain the negative sum of the corresponding columns of $\tilde{A}_\mathrm{k}$, which makes $\tilde{A}_\mathrm{k}$ a column-conservation matrix. In short:

$$\tilde{A}_{\mathrm{k},i,z_j+l} = \begin{cases} k_{jil} & \text{if } i \neq j, \\ -\sum\limits_{o=1,o\neq i}^{m} k_{jo,l} & \text{if } i = j \end{cases} \quad \text{for } i = 1 \dots m, j = 1 \dots m \text{ and } l = 1 \dots d_i. \tag{13}$$

Using the notation Σ for a given bio-CRN, it is clear from the above description that the system's reaction graph and the corresponding dynamics can be characterized either by the sets $(\mathcal{S}, \mathcal{C}, \mathcal{G}, \mathcal{R})$ or equivalently by the matrix triplet $(Y, \tilde{A}_\mathrm{k}, P)$, therefore we can use the notations $\Sigma = (\mathcal{S}, \mathcal{C}, \mathcal{G}, \mathcal{R})$ or $\Sigma = (Y, \tilde{A}_\mathrm{k}, P)$ and Equation (7) is called the *normal form* of the dynamic equations.

The MAL-CRN case. Note that for the special case, in which only mass-action kinetics is used, the P is the identity matrix and $\tilde{A}_{\mathrm{k}} = A_{\mathrm{k}}$ is a square matrix as follows

$$A_{\mathrm{k},ij} = \begin{cases} -\sum_{l=1,l\neq i}^{m} k_{il} & \text{if } i = j \\ k_{ji} & \text{if } i \neq j \end{cases} . \tag{14}$$

Using the notions of the reaction graph, the diagonal elements $A_{\mathrm{k},ii}$ contain the negative sum of the weights of the edges starting from the node C_i, while the off-diagonal elements $A_{\mathrm{k},ij}$, $i \neq j$ contain the weights of the directed edges (C_j, C_i) going into C_i. In this case, the dynamics (7) is simplified to

$$\dot{x} = M \cdot \Psi(x) = Y \cdot A_{\mathrm{k}} \cdot \Psi(x), \tag{15}$$

where $M \in \mathbb{R}^{n\times m}$ is the monomial coefficient matrix coding the reaction graph weighting/structure.

2.3 Graph representation of biochemical reaction networks

The set of complexes together with the set of reactions give rise to the following directed, weighted graph representation. The reaction graph $D = (V_d; E_d)$ consists of a finite non-empty set V_d of vertices and a finite set of E_d directed edges. The directed edges representing reactions will be defined by triplets of the form $e_{(i,j,l)} = (C_i, C_j, G_l)$ for $i, j = 1, \ldots, m$, $i \neq j$, $l = 1, \ldots, d_i$, where i, j and l are the indices of the source complex, product complex and the kinetics, respectively. The positive real weight of the directed edge $e_{(i,j,l)}$ is denoted by k_{ijl}, and it is the principal reaction rate coefficient (see (4)) of the reaction represented by the directed edge. Based on the above description, we will graphically represent the weighted directed edge $e_{(i,j,l)}$ in the reaction graph as

$$C_i \xrightarrow{k,g_{il}} C_j, \tag{16}$$

where $k = k_{ijl}$. Note that, two complexes C_i, C_j can be connected by multiple edges if the species in C_i can produce the products in C_j by more than one kinetics. When we refer to the *structure of the biochemical reaction network*, we mean the unweighted directed graph.

2.4 The non-negativity of the solutions of bio-CRNs

The dynamic variables x_k of any biochemical model are species concentrations, which are naturally non-negative. Therefore, any plausible biochemical model [20, 12] should have this property, that is mathematically based on the notion of essentially non-negative functions [13,14,2,11].

A function $f = [f_1 \ldots f_n]^T : [0,\infty)^n \mapsto \mathbb{R}^n$ is called essentially non-negative if, for all $i = 1, \ldots, n$ and $x \in [0,\infty)^n$, $f_i(x) \geq 0$, whenever $x_i = 0$. Let us consider an autonomous non-linear system

$$\dot{x}(t) = f(x(t)), \quad x(0) = x_0, \quad t \in [0, t_f) \ . \tag{17}$$

Haddad and Chellaboina [13] showed that for a locally Lipschitz f function, the nonnegative orthant $\overline{\mathbb{R}}^n_+$ is invariant under the system dynamics (17) (i.e. starting from a nonnegative initial condition, all the state variables in x remain nonnegative for all time) if and only if f is essentially nonnegative. For example, it is shown in [2] and in [6] that MAL-CRNs of the form (15) are essentially non-negative, and therefore all the solutions are non-negative.

Similar results can be obtained for the bio-CRN case, too as a consequence of the properties (KA1)-(KA3) of the kinetic functions. As all of the kinetic functions in a biochemical reaction network defined by (7) are locally Lipschitz and essentially non-negative, therefore the concentrations remain non-negative. The proof is given in Appendix C.

2.5 Structural properties of bio-CRNs

Some further properties of the bio-CRN structure are presented in this section that are all simple extensions of the notions defined for MAL-CRNs.

*Reversibility and weak reversibility.*A reaction network is called *reversible*, if whenever the reaction $C_i \xrightarrow{k,g_{il}} C_j$ with *any kinetics* g_{il} exists, then a reverse reaction $C_j \xrightarrow{k',g_{jl'}} C_i$ with any other kinetics $g_{jl'}$ is also present in the network. A reaction network is called *weakly reversible*, if whenever complex C_j is reachable from complex C_i on a directed path in the reaction graph, then there exists a directed path from C_j to C_i, too. In other words, each linkage class of the network forms a strongly connected component in the reaction graph.

Deficiency. The notion of the *deficiency* [10] of a reaction kinetic system is built on the set of *reaction vectors* ($\rho^{(l,k)}$) forming the *stoichiometric subspace* S that is defined as

$$S = \{\rho^{(l,k)} = \eta^{(j)} - \eta^{(i)} \mid (C_i, C_j, G_l) \in E \text{ for any } l \in \{1, \dots d_i\}\} \tag{18}$$

where $\eta^{(i)}$ denotes the ith column of Y. The *deficiency* d of a reaction network is defined as:

$$d = m - \ell - s \tag{19}$$

where m is the number of complexes, ℓ is the number of linkage classes and s is the rank of the stoichiometric subspace, i.e. $s = \text{rank}(S)$.

Feinberg proved important properties of the solutions of kinetic systems with mass action law related to the existence, uniqueness and stability of equilibria based on the deficiency and weak reversibility of the network, particularly in the Deficiency Zero and Deficiency One Theorems [10]. With biochemical kinetics, one can apply those points of the Deficiency Zero Theorem that correspond to arbitrary (not necessarily mass action) kinetics. For example, when the deficiency of the network is zero, but the network is not weakly reversible, there is no strictly positive steady state solution, and there cannot be exist a cyclic trajectory in which all states remain positive [10].

2.6 Simple biochemical reaction network examples

2.6.1 An example with rational kinetics

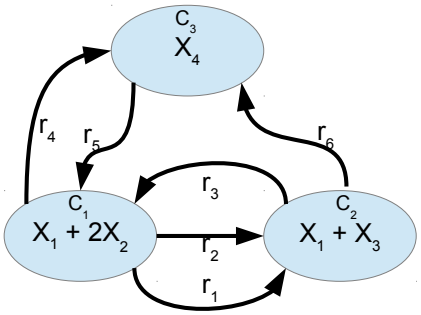

Fig. 1 Reaction graph of a simple bio-CRN

Consider the biochemical network in Figure 1. It is seen that three complexes

$$C_1 = X_1 + 2X_2, \quad C_2 = X_1 + X_3, \quad C_3 = X_4$$

are connected by six irreversible reaction steps. The reaction rate functions are given as follows

$$r_1 = k_1 \frac{x_1 x_2^2}{1 + K_{11} x_1}, \quad r_2 = k_2 \frac{x_1 x_2^2}{1 + K_{21} x_1 + K_{22} x_1 x_2}, \quad r_3 = k_3 \frac{x_1 x_3}{1 + K_{31} x_3}$$
$$r_4 = k_4 \frac{x_1 x_2^2}{1 + K_{41} x_4}, \quad r_5 = k_5 x_4, \quad r_6 = k_6 \frac{x_1 x_3}{1 + K_{31} x_3}.$$

The complex stoichiometric matrix and the corresponding nonlinear vector function are :

$$Y = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Psi(x) = \begin{bmatrix} x_1 x_2^2 \\ x_1 x_3 \\ x_4 \end{bmatrix}. \tag{20}$$

By the factorization of the principal reaction rate coefficients from the reaction rate functions we construct the vector of kinetics $\varphi(x)$. Note that the kinetics of the reactions r_3 and r_6 are the same, therefore the corresponding kinetic function appears only once in the vector

$$\varphi(x) = \left[\frac{x_1 x_2^2}{1 + K_{11} x_1}, \frac{x_1 x_2^2}{1 + K_{21} x_1 + K_{22} x_1 x_2}, \frac{x_1 x_2^2}{1 + K_{41} x_4}, \frac{x_1 x_3}{1 + K_{31} x_3}, x_4 \right]^T.$$

The elements of the vector of kinetics above are ordered according to the complexes from which the corresponding reactions are originating. Then we can decompose

the vector of kinetics to a product $\varphi(x) = P(x) \cdot \Psi(x)$ of the rate weighting matrix $P(x)$ and the monomial vector function $\Psi(x)$ in (20), where

$$P(x) = \begin{bmatrix} \frac{1}{1+K_{11}x_1} & 0 & 0 \\ \frac{1}{1+K_{21}x_1+K_{22}x_1x_2} & 0 & 0 \\ \frac{1}{1+K_{41}x_4} & 0 & 0 \\ 0 & \frac{1}{1+K_{31}x_3} & 0 \\ 0 & 0 & 1 \end{bmatrix} .$$

The modified Kirchoff matrix contains the principal reaction rate coefficients, in our case

$$\tilde{A}_{\mathrm{k}} = \begin{bmatrix} -k_1 & -k_2 & -k_4 & k_3 & k_5 \\ k_1 & k_2 & 0 & -(k_3+k_6) & 0 \\ 0 & 0 & k_4 & k_6 & -k_5 \end{bmatrix} .$$

Therefore the product $f(x) = Y\tilde{A}_k P(x)\Psi(x)$ results in the dynamics of the species of the network:

$$\dot{x}_1 = k_5x_4 - k_4\frac{x_1x_2^2}{1+K_{41}x_4} - k_6\frac{x_1x_3}{1+K_{31}x_3}$$

$$\dot{x}_2 = -2k_1\frac{x_1x_2^2}{1+K_{11}x_1} - 2k_2\frac{x_1x_2^2}{1+K_{21}x_1+K_{22}x_1x_2} - 2k_4\frac{x_1x_2^2}{1+K_{41}x_4} + 2k_3\frac{x_1x_3}{1+K_{31}x_3} + 2k_5x_4$$

$$\dot{x}_3 = k_1\frac{x_1x_2^2}{1+K_{11}x_1} + k_2\frac{x_1x_2^2}{1+K_{21}x_1+K_{22}x_1x_2} - (k_3+k_6)\frac{x_1x_3}{1+K_{31}x_3}$$

$$\dot{x}_4 = k_4\frac{x_1x_2^2}{1+K_{41}x_4} + k_6\frac{x_1x_3}{1+K_{31}x_3} - k_5x_4 \ .$$

The network example is a deficiency 0 network, it is not reversible, but weakly reversible.

2.6.2 A structurally similar example with mass action kinetics.

Let us restrict each kinetics of the previous example to mass action type. Note that, there were two reactions (r_1 and r_2) from C_1 to C_2 with different kinetics, which are now represented by one edge as depicted in Figure 2.

The matrices and the non-linear vector function that characterizes the realization is as follows

$$Y = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad , \quad A_{\mathrm{k}} = \begin{bmatrix} -(k_{12}+k_{13}) & k_{21} & k_{31} \\ k_{12} & -(k_{21}+k_{23}) & 0 \\ k_{13} & k_{23} & -k_{31} \end{bmatrix} \quad , \quad \Psi(x) = \begin{bmatrix} x_1x_2^2 \\ x_1x_3 \\ x_4 \end{bmatrix} .$$

The rate weighting matrix is the identity $P = I_{3\times 3}$. The above elements – based on (15) – define the following differential equation model

$$\dot{x}_1 = -k_{13}x_1x_2^2 - k_{23}x_1x_3$$

$$\dot{x}_2 = -2k_{13}x_1x_2^2 - 2k_{12}x_1x_2^2 + 2k_{21}x_1x_3 + 2k_{31}x_4$$

$$\dot{x}_3 = -k_{23}x_1x_3 - k_{21}x_1x_3 + k_{12}x_1x_2^2$$

$$\dot{x}_4 = -k_{31}x_4 + k_{13}x_1x_2^2 + k_{23}x_1x_3 \ .$$

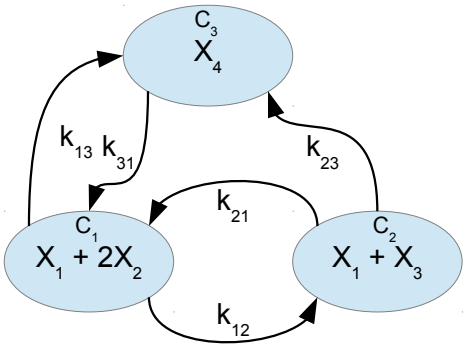

Fig. 2 Reaction graph of a simple MAL-CRN

3 Biochemical reaction network realizations of positive rational ODEs

The problem of constructing a so called reaction network realization to a given a set of nonlinear ordinary differential equations with rational function right-hand sides is considered in this section. This realization problem is different from the realization problem commonly considered in systems theory, where equivalent dynamic models in a pre-specified canonical form, such as with a diagonal coefficient matrix, are searched for.

The question, whether a given set of differential equations can be represented by a biochemical reaction graph is of great interest, since the qualitative properties of the underlying dynamics can possibly be determined based on the properties of the graph. The problem of algorithmically constructing a reaction network realization was first solved by Hárs and Tóth [15] for polynomial differential equations using mass action chemical reaction networks realization.

Since that, many approaches have been developed for defining and investigating realizations of broader classes of non-linear ODEs, that are recently overviewed by the work of Nemcoá and Schuppen [23] (and the references therein) for realization of rational polynomial systems. Recently, Fages et al. [7] considered the realization of biochemical network from a set of ODEs such that each reaction is well-formed. A main difference between that work and ours is the purpose of the methods, i.e. our goal is to infer a *complex-reaction graph* for further computation, while the work of Fages et al. aims at the construction of the *differential influence graph* and *stoichiometric influence graph* (DIG and SIG, respectively).

3.1 Kinetic realizability

First a *necessary and sufficient condition for kinetic realizability* of a set of nonlinear ODEs with rational function right-hand sides are proposed that ensures the successful completion of the canonical realization algorithm described in subsection 3.2 below. We follow the method presented by Chellaboina [2] in the derivation of this condition for kinetic realizability.

Consider the following autonomous ordinary differential equations

$$\dot{x}(t) = f(x(t)), \quad x(0) = x_0, \quad t \in [0, t_f),$$

where $f(x) : \mathbb{R}^n \to \mathbb{R}^n$ and $x \in \mathbb{R}^n$. Assume that the right hand side function f is composed of the linear combination of biochemical reaction rate functions (4). Then, there exists a bio-CRN with n species if and only if for each $o = 1, \dots n$ the $f_o([x_1, x_2, \dots, x_{o-1}, 0, x_{o+1} \dots x_n])$ is a *non-negative* linear combination of the biochemical reaction rate functions. The proof of sufficiency and necessity can be found in Appendix D.

3.2 Canonical realization algorithm

Once the kinetic realizability condition is assured for the dynamical equations, the goal is to find the components of the chemical reaction network, i.e. the sets of species, complexes, reactions and kinetics. This is solved by Algorithm 1 that is an extension of the canonical realization algorithm [15], also discussed in [2]. This algorithm transforms the set of ODEs into a biochemical reaction network. The algorithm processes each equation one-by-one and requires the rational function terms on the right hand side of the equations. First, based on the exponents of the monomials in the nominator the corresponding source complex is determined. Then, a product complex is assigned to the reaction. In case of processing the ith equation, the stoichiometric coefficients of the species in the product and source complex are the same for all the species, except the species X_i. Thus, the inferred reaction does not contribute to any other than the i-th equation. Finally, the kinetics is determined by the functional form of the rational term.

The pseudo-code of the canonical realization algorithm for bio-CRNs are given as Algorithm 1.

It is known, that the canonical realization algorithm generally results in a large number of complexes. Further, the inferred biochemical reaction network does not fulfil thermodynamic constraints and almost never fulfil the mass conservation due to the way the product complex is created for each reaction. Therefore, the only purpose of the presented algorithm is to generate *a* biochemical reaction network for kinetic systems. Then, in section 4 optimization methods are formulated to find dynamically equivalent realizations with desired properties starting from an arbitrary reaction network realization.

3.3 An example for constructing the canonical realization for a biochemical reaction network

Consider the following set of ordinary differential equations

$$\dot{x}_1 = -k_{21}x_1^3 - \frac{k_{22}x_1^3}{1 + K_2x_1} + 3k_{11}x_2^3 + \frac{3k_{12}x_2^3}{1 + K_1x_2}$$
$$\dot{x}_2 = k_{21}x_1^3 + \frac{k_{22}x_1^3}{1 + K_2x_1} - 3k_{11}x_2^3 - \frac{3k_{12}x_2^3}{1 + K_1x_2}.$$

It is easy to check that the necessary and sufficient conditions hold for this system, i.e. $f_1([0, x_2])$ and $f_1([x_1, 0])$ are positive linear combination of biochemical reaction rate functions.

Table 1 contains the detailed realization procedure and the inferred reactions in each step.

Algorithm 1 Algorithm for realization of biochemical reaction network from ODEs.

Require: a set of ODEs of n variables; each f_i is a linear combination of elementary reaction rate functions, i.e.

$f_i(x) = \sum_{l=1}^{m} \left(c_l \frac{\prod_{j=1}^{n} x_j^{\nu_{jl}}}{D_l(x)} \right) = \sum_{l=1}^{m} c_l g_l(x);$

$\mathcal{R} = \emptyset$; # set of reactions
$\mathcal{C} = \emptyset$; # set of complexes
$\mathcal{S} = \{X_1 \dots X_n\}$; # set of species
for $i = 1$ **to** n **do**
  for each $c_l g_l(x)$ **do**
    $C_s = \sum \nu_{jl} X_j$; # source complex
    if $C_s \notin \mathcal{C}$ **then**
      add C_s to $\mathcal{C}$;
    end if
    if $\nu_{il} \geq 1$ **then**
      $C_p = (\nu_{il} + \mathrm{sign}(c_i))\, X_i + \sum_{j=1, j\neq i}^{n} \nu_{jl} X_j$; # product complex
      $r = |c_i| \frac{\sum_{j=1}^{n} x_j^{\nu_{jl}}}{D_l(x)}$; # reaction rate
      $g = \frac{\sum_{j=1}^{n} x_j^{\nu_{jl}}}{D_l(x)}$; # kinetics
    else
      $C_p = X_i + \sum_{j=1, j\neq i}^{n} \nu_{jl} X_j$; # product complex
      $r = c_i \frac{\sum_{j=1}^{n} x_j^{\nu_{jl}}}{D_l(x)}$; # reaction rate
      $g = \frac{\sum_{j=1}^{n} x_j^{\nu_{jl}}}{D_l(x)}$; # kinetics
    end if
    if $C_p \notin \mathcal{C}$ **then**
      add C_p to $\mathcal{C}$;
    end if
    add reaction (C_s, C_p, g) to $\mathcal{R}$;
  end for
end for

In conclusion, Algorithm 1 created the complexes

$$\mathcal{C} = \{3X_1,\ 2X_1,\ 3X_2,\ X_1 + 3X_2,\ 3X_1 + X_2,\ 3X_2\}$$

and the kinetics

$$g_{11} = x_1^3 \qquad g_{12} = \frac{x_1^3}{1 + K_2 x_1}$$

$$g_{31} = x_2^3 \qquad g_{32} = \frac{x_2^3}{1 + K_1 x_2} \quad .$$

The corresponding network can be seen in Figure 3.

4 Dynamically equivalent realizations

It has been known that different reaction graph structures in the mass-action case may lead to the same kinetic differential equations. In other words, the reaction

Table 1 Detailed procedure of the simple realization example. Each line of the table corresponds to a realized reaction. The first column contains the right hand side component of the ODEs, the second column shows a reaction term of the right hand side function. The third column contains the assigned reaction: the source and the product complexes are written on the sides of the arrow, the principal reaction rate coefficient and the reaction kinetic function is written on the arrow.

Equation	Term	Realized reaction	New complexes
$f_1(x)$	$-k_{21}x_1^3$	$3X_1 \xrightarrow{\lvert -k_{21}\rvert \left(\frac{x_1^3}{1}\right)} (3+\mathrm{sign}(-k_{21}))X_1$	$C_1 = 3X_1$, $C_2 = 2X_1$
$f_1(x)$	$-\frac{k_{22}x_1^3}{1+K_2x_1}$	$3X_1 \xrightarrow{\lvert -k_{22}\rvert \left(\frac{x_1^3}{1+K_2x_1}\right)} (3+\mathrm{sign}(-k_{22}))X_1$	-
$f_1(x)$	$3k_{11}x_2^3$	$3X_2 \xrightarrow{\lvert 3k_{11}\rvert \left(\frac{x_2^3}{1}\right)} X_1+3X_2$	$C_3 = 3X_2$, $C_4 = X_1+3X_2$
$f_1(x)$	$\frac{3k_{12}x_2^3}{1+K_1x_2}$	$3X_2 \xrightarrow{\lvert 3k_{12}\rvert \left(\frac{x_2^3}{1+K_1x_2}\right)} X_1+3X_2$	-
$f_2(x)$	$k_{21}x_1^3$	$3X_1 \xrightarrow{\lvert k_{21}\rvert \left(\frac{x_1^3}{1}\right)} X_2+3X_1$	$C_5 = X_2+3X_1$
$f_2(x)$	$\frac{k_{22}x_1^3}{1+K_2x_1}$	$3X_1 \xrightarrow{\lvert k_{22}\rvert \left(\frac{x_1^3}{1+K_2x_1}\right)} X_2+3X_1$	-
$f_2(x)$	$-3k_{11}x_2^3$	$3X_2 \xrightarrow{\lvert -3k_{11}\rvert \left(\frac{x_2^3}{1}\right)} (3+\mathrm{sign}(-3k_{11}))X_2$	$C_6 = 2X_2$
$f_2(x)$	$-\frac{3k_{12}x_2^3}{1+K_1x_2}$	$3X_2 \xrightarrow{\lvert -3k_{12}\rvert \left(\frac{x_2^3}{1+K_1x_2}\right)} (3+\mathrm{sign}(-3k_{12}))X_2$	-

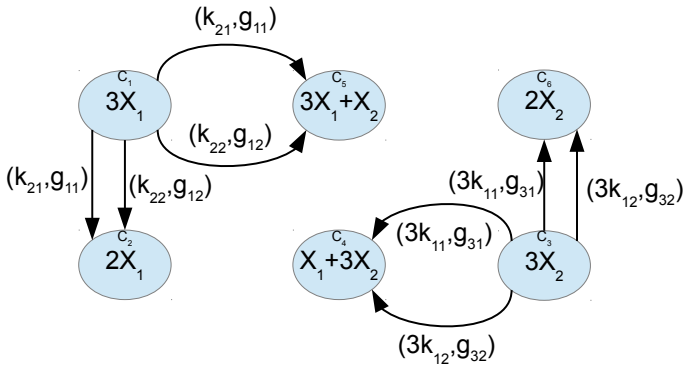

Fig. 3 The reaction graph of the example constructed by the canonical realization algorithm.

graph structure corresponding to a given kinetic ODE system is non-unique. This phenomenon is called macro-equivalence, dynamical equivalence or confoundability [5,17,30,18]. Since the mass action case is a special case of the generalized kinetic description presented in this paper, dynamical equivalence necessarily emerges for the bio-CRN structure, too.

4.1 Dynamic equivalence for MAL-CRNs

It is well-known that given a coefficient matrix M in (15) with a stoichiometric matrix Y one may find more than one Kirchhoff matrix A_k satisfying $M = YA_\mathrm{k}$. Thus, the reaction graph and its corresponding realization (Y, A_k) is not unique. Two *realizations* (Y, A_k) and (Y, A'_k) are called *dynamically equivalent* if they give rise to the same M, i.e. $M = YA_\mathrm{k} = YA'_\mathrm{k}$.

Important structural properties of chemical reaction networks, such as deficiency, reversibility and weak-reversibility mentioned in subsection 2.5 are not encoded uniquely in the differential equations of the chemical reaction networks, i.e. they are *realization dependent properties*. Since some realizations –with specific properties– are more suitable for the analysis of the solutions of the dynamic equations, the need for developing computational methods in order to find dynamically equivalent realizations appeared and solved for MAL-CRNs [30,32,19].

4.2 Dynamical equivalence in bio-CRNs

Two biochemical reaction networks $\Sigma_1 = (Y^1,\, \tilde{A}^1_\mathrm{k},\, P^1(x))$ and $\Sigma_2 = (Y^2,\, \tilde{A}^2_\mathrm{k},\, P^2(x))$ are said to be *dynamically equivalent*, if the two networks give rise to the same dynamic equations (7), i.e.

$$f(x) = Y^1 \tilde{A}^1_\mathrm{k} P^1(x) \Psi^1(x) = Y^2 \tilde{A}^2_\mathrm{k} P^2(x) \Psi^2(x) \quad \text{for } \forall x \in \overline{\mathbb{R}}^n_+ \ , \tag{21}$$

where $Y^{\{1,2\}}$ are integer type matrices, $\tilde{A}^{\{1,2\}}_\mathrm{k}$ are modified Kirchoff matrices as in (13), $P(x)^{\{1,2\}}$ are rate weighting mappings as defined in (9-10) and $\Psi^{\{1,2\}}$ are computed from $Y^{\{1,2\}}$ according to (8).

Let denote the complex composition matrix by Y, which stores the stoichiometric coefficients of the complexes of *both* networks and let P be the common rate weighting matrix, which also contains the reaction kinetics of both networks by forming the *union* of the two corresponding sets. In this case the modified Kirchhoff matrix of each network will contain some zero rows and columns corresponding to the other network kinetics and complexes. Further let $\Psi_i(x) = \prod_{j=1}^m x_j^{Y_{ji}}$. Using the common matrices and the kinetic vector (11) the condition of dynamically equivalent networks can also be written as

$$Y \tilde{A}^1_\mathrm{k} \varphi(x) = Y \tilde{A}^2_\mathrm{k} \varphi(x). \tag{22}$$

It is easy to see that if two networks are dynamically equivalent but different, then there are infinitely many equivalent networks. For this purpose let $0 \geq a \geq 1$, then

$$Ya\tilde{A}^1_\mathrm{k} \varphi(x) + Y(1-a)\tilde{A}^2_\mathrm{k} \varphi(x) = Y \tilde{A}^1_\mathrm{k} \varphi(x) \ . \tag{23}$$

Let assume that the components of the kinetic vector $\varphi(x)$ are linearly independent, i.e.

$$\sum_{i=1}^{\kappa} \alpha_i \varphi_i(x) = 0, \text{ for all } x > 0 \text{ if and only if } \alpha_i = 0 \text{ for } \forall i = 1, \dots \kappa. \tag{24}$$

Then it is apparent that the *kinetic coefficient matrix* $\tilde{M} = Y\tilde{A}^1_{\mathrm{k}} = Y\tilde{A}^2_{\mathrm{k}}$ is an *invariant* for all realizations.

4.3 Sparse and dense realizations

When the dynamic system has more than one realizations, it has infinitely many. Some realizations show structural differences, i.e. the corresponding un-weighted reaction graphs are different, while others are different only in the numerical weights of the graph. When the complexes and their kinetics are fixed, the reaction networks with the least number of edges are called the *sparse realizations*, while the realizations with the highest number of edges are called the *dense realizations*.

Some interesting properties of these special realisations were shown in [33] for the mass-action network case. These results are adapted for biochemical reaction networks here.

Let $\Sigma = (Y, \tilde{A}_{\mathrm{k}}, P)$ be a kinetic system with fixed complexes Y and kinetics P, and let $\Sigma^s = (Y, \tilde{A}^s_{\mathrm{k}}, P)$ and $\Sigma^d = (Y, \tilde{A}^d_{\mathrm{k}}, P)$ be the dynamically equivalent sparse and dense realizations, respectively. Then

P1 the un-weighted graph of any dynamically equivalent realization of Σ is a subgraph of the dense realization Σ^d,
P2 the dense realisation Σ^d is structurally unique,
P3 a realization of the kinetic system is structurally unique if and only if the sparse and the dense realizations are structurally identical.

The proof can be found in Appendix E. A simple implication of the theorems is that the dense realization contains all the possible reactions.

4.4 Optimization methods for the computation of realizations with preferred properties

There are many recent results on the optimization based computation of equivalent reaction networks with given properties for the chemical reaction networks obeying the mass action law. For example, in [30] the optimisation problem for the computation of sparse and dense realizations are stated as mixed integer linear problem, which can be efficiently solved even for hundreds of chemical species. In [32] the procedure is adapted for finding complex and detailed balance realizations. In [33] the problem to find equivalent realizations with minimum or maximum number of complexes are considered. In this section we show, how the mixed integer linear programming procedure can be adapted for the biochemical reaction network case.

Given a realization invariant coefficient matrix $\tilde{M} \in \mathbb{R}^{m\times\kappa}$, the set of complexes $Y \in \mathbb{R}^{n\times m}$ and the number of possible reaction kinetics in each complex

$(d_1, \dots d_m)$, such that the total number of kinetics $\kappa = \sum_{i=1}^{m} d_i$. The goal is to find valid modified Kirchhoff matrix or matrices with properties in (13), which fulfils the following matrix equation:

$$\tilde{M} = Y\tilde{A}_{\mathrm{k}}. \tag{25}$$

The only difference here, comparing to the existing optimization methods in [30, 32,33] is the properties of the matrix $\tilde{A}_{\mathrm{k}}$. In the MAL-CRN case $\tilde{A}_{\mathrm{k}}$ is a square matrix with negative diagonal elements such that the sum of each column is zero. But matrix $\tilde{A}_{\mathrm{k}}$ is typically rectangular in the bio-CRN case, and the location of the negative elements depends on the number of kinetics in each complex.

The entries of the $\tilde{A}_{\mathrm{k}}$ are written in the following way utilizing the indexing $z_i = \sum_{k=1}^{i-1} d_k$, for $i = 1 \dots m$ ($z_1 = 0$ and $z_m + d_m = \kappa$) as in (12):

$$\tilde{A}_{\mathrm{k}}(a) = \begin{pmatrix} -a_{1,1} & \dots & -a_{1,d_1} & a_{1,z_2+1} & \dots & a_{1,z_2+1} & \dots & a_{1,z_m+d_m} \\ a_{2,1} & \dots & a_{2,d_1} & -a_{2,z_2+1} & \dots & -a_{2,z_2+d_2} & \dots & a_{2,z_m+d_m} \\ & & & & \vdots & & & \\ a_{m-1,1} & \dots & & & & & & a_{m-1,z_m+d_m} \\ a_{m,1} & \dots & & & & & & -a_{m,z_m+d_m} \end{pmatrix}. \tag{26}$$

With this explicit notation of the negative elements we can restrict the decision variables (a_{ij}) of the optimization problem to the non-negative orthant. The column conservation property of $\tilde{A}_{\mathrm{k}}$ can be expressed as κ number of equations as $1_m^T \tilde{A}_{\mathrm{k},(\cdot,i)} = 0$, for $i = 1 \dots \kappa$, where 1_m is the m dimensional one vector and $\tilde{A}_{\mathrm{k},(\cdot,i)}$ is the ith column of the matrix.

The non-zero elements of the matrix, which do not have the negative sign, i.e. a_{i,z_j+l} for $i = 1 \dots m$, $j = 1 \dots m$, $l = 1 \dots d_i$, $i \neq j$, define the reactions in the network. When each complex has only one kinetics (all $d_i = 1$) then $\tilde{A}_k$ is a square matrix and these are the off-diagonal elements. Let us introduce binary decision variables w_{ij} for each entry of the matrix $\tilde{A}_{\mathrm{k}}$. This binary variable equals to one if and only if the corresponding entry is larger than zero (in practice, larger than some small threshold value ϵ). Furthermore, let a_{ij}^{ub} be practical upper bounds for the elements of the matrix.

Then, the following mixed integer linear optimization problem finds a sparse realization

$$\underset{y,a}{\text{minimize}} \sum_{i=1}^{m} \sum_{j=1}^{\kappa} w_{ij} \tag{27}$$

$$\text{subject to: } \tilde{M} - Y\tilde{A}_{\mathrm{k}}(a) = 0 \tag{28}$$

$$1_m^T \tilde{A}_{\mathrm{k},(\cdot,i)} = 0 \text{ for } i = 1 \dots \kappa \tag{29}$$

$$0 \leq a_{ij} \leq a_{ij}^{\mathrm{ub}} \text{ for } i = 1 \dots m, j = 1 \dots \kappa \tag{30}$$

$$0 \leq a_{ij} - \epsilon w_{ij} \text{ for } i = 1 \dots m, j = 1 \dots \kappa \tag{31}$$

$$0 \leq -a_{ij} + a_{ij}^{\mathrm{ub}} w_{ij} \text{ for } i = 1 \dots m, j = 1 \dots \kappa \tag{32}$$

$$w_{ij} \text{ are binary variables, for } i = 1 \dots m, j = 1 \dots \kappa \tag{33}$$

where (27) defines the goal of finding the sparse realization by minimizing the number of edges, (29) are the constraints ensuring dynamic equivalence, (30) bounds

the principal reaction rate coefficients, and (31) - (33) define the binary variables associated to the reactions (edges). The dense realization can be simply found by changing the sign of the objective function (27). One can provide further equality constrains to incorporate a-priory knowledge about existing reactions or to exclude possible reactions.

Note that, the exact form of the reaction kinetics does not appear in the optimization, thus the complexity of the reaction rate functions do not influence the performance of the optimization framework for finding dynamic equivalent networks. Further, the structural differences between the networks also have no influence on the optimization problem, since the structural differences are encoded by the values of the binary decision variables, the number of decision variables is the same.

4.5 Example for dynamically equivalent realizations

Consider the networks depicted in Figure 4 with species $\mathcal{S} = \{X_1, X_2, X_3\}$ and complexes $C_1 = 3X_2$, $C_2 = 3X_1$ and $C_3 = 2X_1 + X_2$. The complex composition matrix and the corresponding monomial vector are

$$Y = \begin{bmatrix} 0 & 3 & 2 \\ 3 & 0 & 1 \end{bmatrix} \quad , \quad \Psi(x) = \begin{bmatrix} x_2^3 \\ x_1^3 \\ x_1^2 x_2 \end{bmatrix} \quad .$$

The species in each complex can react with two different kinetics ($d_1 = d_2 = d_3 = 2$):

$$\begin{aligned} &g_{1\,1} = x_2^3, && g_{1\,2} = \frac{x_2^3}{1 + K_1 x_2}, && g_{2\,1} = x_1^3, \\ &g_{2\,2} = \frac{x_1^3}{1 + K_2 x_1}, && g_{3\,1} = x_1^2 x_2, && g_{3\,2} = \frac{x_1^2 x_2}{1 + K_3 x_1 x_2} \quad , \end{aligned}$$

where the first index indicates the source complex and the second index identifies the kinetics. Thus, the vector of kinetics can be written as the product of the monomial vector and the rate weighting matrix

$$\varphi(x) = P(x)\Psi(x) = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{1+K_1 x_2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{1+K_2 x_1} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \frac{1}{1+K_3 x_1 x_2} \end{bmatrix} \cdot \begin{bmatrix} x_2^3 \\ x_1^3 \\ x_1^2 x_2 \end{bmatrix} = \begin{bmatrix} x_2^3 \\ \frac{x_2^3}{1+K_1 x_2} \\ x_1^3 \\ \frac{x_1^3}{1+K_2 x_1} \\ x_1^2 x_2 \\ \frac{x_1^2 x_2}{1+K_3 x_1 x_2} \end{bmatrix} \quad .$$

Now we can solve the MILP optimization problem (27)-(33) to find all the sparse realizations, and the other optimization problem with the changed objective function but the same constraints (28)-(33) to obtain the dense realization. The six $\tilde{A}_{\mathrm{k}}$ matrices that correspond to the obtained six dynamically equivalent networks

depicted in Figure 4 are

$$\tilde{A}_{\mathrm{k}}^{A} = \begin{bmatrix} -k_{11} & -k_{12} & 0 & 0 & 0 & 0 \\ k_{11} & k_{12} & -k_{21} & -k_{22} & 0 & 0 \\ 0 & 0 & k_{21} & k_{22} & 0 & 0 \end{bmatrix}, \qquad \tilde{A}_{\mathrm{k}}^{B} = \begin{bmatrix} -\frac{3}{2}k_{11} & -k_{12} & \frac{1}{3}k_{21} & 0 & 0 & 0 \\ 0 & k_{12} & -\frac{1}{3}k_{21} & -k_{22} & 0 & 0 \\ \frac{3}{2}k_{11} & 0 & 0 & k_{22} & 0 & 0 \end{bmatrix}$$

$$\tilde{A}_{\mathrm{k}}^{C} = \begin{bmatrix} -k_{11} & -\frac{3}{2}k_{12} & 0 & \frac{1}{3}k_{22} & 0 & 0 \\ k_{11} & 0 & -k_{21} & -\frac{1}{3}k_{22} & 0 & 0 \\ 0 & \frac{3}{2}k_{12} & k_{21} & 0 & 0 & 0 \end{bmatrix}, \quad \tilde{A}_{\mathrm{k}}^{D} = \begin{bmatrix} -\frac{3}{2}k_{11} & -\frac{1}{2}3k_{12} & \frac{1}{3}k_{21} & \frac{1}{3}k_{22} & 0 & 0 \\ 0 & 0 & -\frac{1}{3}k_{21} & -\frac{1}{3}k_{22} & 0 & 0 \\ \frac{3}{2}k_{11} & \frac{3}{2}k_{12} & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\tilde{A}_{\mathrm{k}}^{E} = \begin{bmatrix} -k_{11} & -k_{12} & \frac{k_{21}}{3} & \frac{k_{22}}{3} & 0 & 0 \\ k_{11} & k_{12} & -\frac{k_{21}}{3} & -\frac{k_{22}}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\tilde{A}_{\mathrm{k}}^{F} = \begin{bmatrix} (-k_{11} - \frac{l_{11}}{3}) & (-k_{12} - \frac{l_{12}}{3}) & \frac{1}{3}(k_{21} - l_{21}) & \frac{1}{3}(k_{22} - l_{22}) & l_{31} & l_{32} \\ (k_{11} - \frac{2l_{11}}{3}) & (k_{12} - \frac{2l_{12}}{3}) & (-\frac{k_{21}}{3} - \frac{2l_{21}}{3}) & (-\frac{k_{22}}{3} - \frac{2l_{22}}{3}) & 2l_{31} & 2l_{32} \\ l_{11} & l_{12} & l_{21} & l_{22} & -3l_{31} & -3l_{32} \end{bmatrix}.$$

One can easily check that all networks give rise to the same dynamics $\dot{x} = Y \cdot \tilde{A}_{\mathrm{k}}^{i} \cdot \varphi(x)$ with $i = A, B, C, D, E, F$. The dynamic equations read as

$$\dot{x}_1 = -k_{21}x_1^3 - \frac{k_{22}x_1^3}{1 + K_2x_1} + 3k_{11}x_2^3 + \frac{3k_{12}x_2^3}{1 + K_1x_2}$$
$$\dot{x}_2 = k_{21}x_1^3 + \frac{k_{22}x_1^3}{1 + K_2x_1} - 3k_{11}x_2^3 - \frac{3k_{12}x_2^3}{1 + K_1x_2} \quad .$$

Note, that the networks A-E are all sparse realizations. Each of them contains 4 edges, which is the minimal number of reactions, that can represent the dynamics. In network E, the complex $2X_1 + X_2$ is isolated (no incoming nor outgoing reaction), therefore it is not shown.

Realization F is the dense realization, which contains all the possible reactions. The dense realisation is obviously *structurally unique*, since in this case it contains all the possible edges. However, there are continuum many dense realizations, with the following conditions on the weights: $l_{11} < \frac{3k_{11}}{2}$, $l_{12} < \frac{3k_{12}}{2}$, $l_{21} < k_{21}$, $l_{22} < k_{22}$, $l_{31} > 0$, $l_{32} > 0$, which guarantees that the matrices $\tilde{A}_{\mathrm{k}}^{i}$, $i = A, B, C, D, E, F$ are proper modified Kirchoff matrices.

Further note, that network A,B,C,D are neither reversible nor weakly reversible, but networks E and F are reversible realizations.

5 Summary and conclusions

A canonical decomposition of biochemical reaction rate functions being in a rational function form is proposed in this paper that contains elementary reaction rates with a reaction monomial in the nominator, and a positive polynomial in the denominator. Such a decomposition is meaningful for the majority of known biochemical reaction rate functions, and gives rise to a set of irreversible elementary biochemical reactions in a biochemical reaction system.

Given such a set of elementary biochemical reactions, we have proposed a unified biochemical reaction network (bio-CRN) structure as a generalization of the well established chemical reaction network structure by defining the complexes, the complex composition matrix Y and the modified Kirchoff matrix $\tilde{A}_k$ of the structure that can be represented as a generalized reaction graph. It was proved that

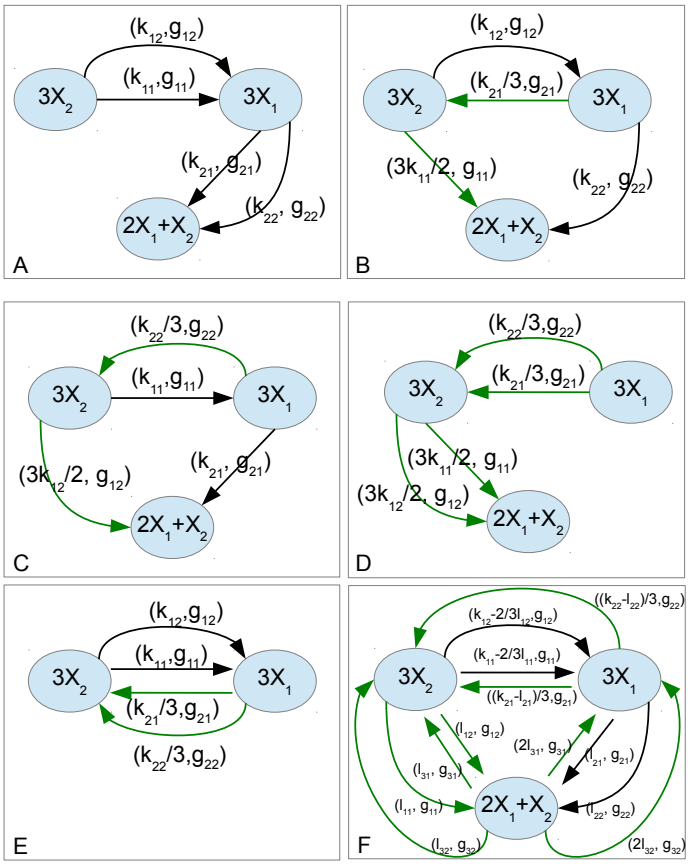

Fig. 4 Reaction graphs of dynamically equivalent biochemical reaction networks. The differences from the network A are depicted with green edges.

a bio-CRN model is essentially non-negative. The basic parameter-independent structural properties, such as reversibility, weak reversibility and deficiency were also illustrated using the bio-CRN structure.

The notion of dynamically equivalent biochemical networks was introduced as a generalization of that defined for CRNs, and an algorithm was proposed for constructing a realization of a biochemical reaction network from its ODE model. Then the mixed integer linear optimization problem was formulated for computing dynamically equivalent alternative sparse and dense bio-CRN structures to a given dynamics that enables to apply similar computational methods that were developed for the CRN case by Szederkényi [30]. The methods and tools proposed by the bio-CRN case were illustrated by simple examples.

Acknowledgements AG and JRB are supported by the funding from EU FP7 ITN "NICHE", project no. 289384. KMH is thankful for the funding from the Hungarian Scientific Research Fund through grant K83440. GSz acknowledges the support of the Hungarian Scientific Research Fund through grant NF104706.

References

1. Bailey, J.E.: Complex biology with no parameters. Nature Biotechnology **19**(6), 503–504 (2001)
2. Chellaboina, V.: Modeling and Analysis of Mass-Action Kinetics. IEEE Control System Magazine (August), 60–78 (2009)
3. Conradi, C., Flockerzi, D., Raisch, J., Stelling, J.: Subnetwork analysis reveals dynamic features of complex (bio)chemical networks. Proceedings of the National Academy of Sciences **104**(49), 19,175–19,180 (2007)
4. Craciun, G., Feinberg, M.: Multiple equilibria in complex chemical reaction networks : extensions to entrapped species models. IEE Proc.-Syst. Biol. **153**(4), 179–186 (2006).
5. Craciun, G., Pantea., C.: Identifiability of chemical reaction networks. Journal of Mathematical Chemistry **44**, 244–259 (2008)
6. 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)
7. Fages, F., Gay, S., Soliman, S.: Inferring reaction systems from ordinary differential equations. Theoretical Computer Science **1**, 1–15 (2014).
8. Feinberg, M.: Complex Balancing in General Kinetic Systems. Archive for Rational Mechanics and Analysis **49**(3), 187–194 (1972)
9. Feinberg, M.: Lectures on Chemical Reaction Networks. Tech. rep., Department of Chemical Engineering, University of Rochester (1980)
10. 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), 2229–2268 (1987)
11. Fleming, R.M.T., Thiele, I.: Mass conserved elementary kinetics is sufficient for the existence of a non-equilibrium steady state concentration. Journal of Theoretical Biology **314**(0), 173–181 (2012)
12. Gábor, A., Hangos, K.M., Szederkenyi, G., Banga, J.R.: On the Verification and Correction of Large-scale Kinetic Models in system Biology. Computational Methods in Systems Biology, Lecture Notes in Computer Science **8130**, 206–219 (2013)
13. Haddad, W.M., Chellaboina, V.: Stability and dissipativity theory for nonnegative dynamical systems: {A} unified analysis framework for biological and physiological systems. Nonlinear Analysis: Real World Applications **6**, 35–65 (2005)
14. Haddad, W.M., Chellaboina, V.S., Hui, Q.: Nonnegative and Compartmental Dynamical Systems. Princeton University Press (2010)
15. Hárs, V., Tóth, J.: On the inverse problem of reaction kinetics. In: M. Farkas, L. Hatvani (eds.) Qualitative Theory of Differential Equations, *Coll. Math. Soc. J. Bolyai*, vol. 30, pp. 363–379. North-Holland, Amsterdam (1981)
16. 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), 3067–3074 (2006)
17. Horn, F., Jackson, R.: General mass action kinetics. Archive for Rational Mechanics and Analysis **47**(2), 81–116 (1972)
18. Johnston, M.D., Siegel, D.: Linear conjugacy of chemical reaction networks. Journal of Mathematical Chemistry **49:7**(7), 1263–1282 (2011).
19. Johnston, M.D., Siegel, D., Szederkényi, G.: Computing weakly reversible linearly conjugate chemical reaction networks with minimal deficiency. Mathematical Biosciences **241:1**(1), 88–98 (2013).
20. Kaleta, C., Richter, S., Dittrich, P.: Using chemical organization theory for model checking. Bioinformatics (Oxford, England) **25**(15), 1915–22 (2009).
21. Karp, R.L., Millán, M.P., Dasgupta, T., Dickenstein, A., Gunawardena, J.: Complex-linear invariants of biochemical networks. Journal of Theoretical Biology **311**(0), 130–138 (2012)
22. Neigenfind, J., Grimbs, S., Nikoloski, Z.: On the relation between reactions and complexes of (bio)chemical reaction networks. Journal of Theoretical Biology **317**(0), 359–365 (2013)
23. Nemcová, J., Schuppen, J.H.: Realization Theory for Rational Systems: The existence of Rational Realizations. SIAM J. Control Optim. **48**(4), 2840–2856 (2009)
24. Otero-Muras, I., Banga, J.R., Alonso, A.A.: Exploring multiplicity conditions in enzymatic reaction networks. Biotechnology Progress **25**(3), 619–631 (2009)
25. Prill, R.J., Marbach, D., Saez-Rodriguez, J., Sorger, P.K., Alexopoulos, L.G., Xue, X., Clarke, N.D., Altan-Bonnet, G., Stolovitzky, G.: Towards a rigorous assessment of systems biology models: the DREAM3 challenges. PloS ONE **5**(2), e9202 (2010)

26. Rao, S., van der Schaft, A., van Eunen, K., Bakker, B., Jayawardhana, B.: A model reduction method for biochemical reaction networks. BMC Systems Biology **8**(52) (2014).

27. van der Schaft, A., Rao, S., Jayawardhana, B.: On the mathematical structure of balanced chemical reaction networks governed by mass action kinetics. SIAM Journal on Applied Mathematics **73**(2), 953–973 (2013)

28. Shinar, G., Feinberg, M.: Structural sources of robustness in biochemical reaction networks. Science **327**(5971), 1389–1391 (2010).

29. Shinar, G., Feinberg, M.: Design principles for robust biochemical reaction networks: what works, what cannot work, and what might almost work. Mathematical Biosciences **231**(1), 39–48 (2011)

30. Szederkényi, G.: Computing sparse and dense realizations of reaction kinetic systems. Journal of Mathematical Chemistry **47**(2), 551–568 (2010).

31. Szederkényi, G., Banga, J.R., Alonso, A.A., Szederkenyi, G.: Inference of complex biological networks: distinguishability issues and optimization-based solutions. BMC systems biology **5**(1), 177 (2011).

32. Szederkényi, G., Hangos, K.M.: Finding complex balanced and detailed balanced realizations of chemical reaction networks. Journal of Mathematical Chemistry **49**(6), 1163–1179 (2011)

33. Szederkényi, G., Hangos, K.M., Péni, T.: Maximal and minimal realizations of reaction kinetic systems: computation and properties. MATCH Communications in Mathematical and in Computer Chemistry **65**(2), 309–332 (2011)

34. Wang, L., Sontag, E.D.: On the number of steady states in a multiple futile cycle. Journal of Mathematical Biology **57**(1), 29–52 (2008)

A Nomenclature

Notation	Description
$A_\mathrm{k} \in \mathbb{R}^{m\times m}$	Kirchhoff matrix containing the reaction rate coefficients
$\tilde{A}_\mathrm{k} \in \mathbb{R}^{m\times \kappa}$	Kirchhoff matrix containing the principal reaction rate coefficients
$\mathcal{C} = \{C_1, \dots C_m\}$	set of complexes
$d_1, \dots d_m$	number of different kinetics from complexes $C_1, \dots C_m$, respectively
$D : x \in \overline{\mathbb{R}}_+^n \to \mathbb{R}_+$	positive multivariate polynomial with leading 1
$\varphi : x \in \overline{\mathbb{R}}_+^n \to \overline{\mathbb{R}}_+^\kappa$	vector of kinetics
g_{il}	the lth kinetics of the reaction in complex C_i
k_{ijl}	reaction rate coefficient of the reaction $C_i \to C_j$ of the lth kinetics
κ	number of different kinetics in the network
E_d	edges of the weighted, directed reaction graph
$\eta^{(i)}$	complex composition vector of complex C_i
m	number of complexes
$M = YA_\mathrm{k}$	monomial coefficient matrix
$\tilde{M} = Y\tilde{A}_\mathrm{k}$	kinetic coefficient matrix
μ_{ij}	stoichiometric coefficient of specie X_i in (product) complex C_j
n	number of species
ν_{ij}	stoichiometric coefficient of specie X_i in (source) complex C_j
$P(x) \in \mathbb{R}^{\kappa\times m}$	rate weighting matrix
$\Psi(x) \in \overline{\mathbb{R}}_+^n \to \overline{\mathbb{R}}_+^m$	monomial vector function
$r(x) \in \overline{\mathbb{R}}_+^n \to \overline{\mathbb{R}}_+$	reaction rate function
$\mathbb{R}^n$	the space of n-dimensional real vectors
$\overline{\mathbb{R}}_+^n = [0,\infty)^n$	n-dimensional non-negative orthant
$\mathbb{R}_+^n = (0,\infty)^n$	n-dimensional positive orthant
$\mathbb{R}^{m\times n}$	the space of m × n dimensional real matrices
$\mathcal{R}$	set of reactions
$\rho^{(l,k)} = \eta^{(l)} - \eta^{(k)}$	reaction vector corresponding to the reaction $C_k \to C_l$
$\mathcal{S}$	set of species
S	stoichiometric subspace
V_d	vertices of the weighted, directed reaction graph
$X_1, \dots X_n$	species
$x \in \mathbb{R}^n$	species concentration
Y	complex composition (stoichiometric) matrix
$\tilde{Y}$	truncated complex composition (stoichiometric) matrix

B Further examples and decompositions of biochemical reaction rate functions

In this section we show for a set of biochemical reaction rate functions –mostly from the well-known modeling software COPASI [16,]–, how to formulate them using elementary biochemical reaction rate functions according to (4). We have to note that not every reaction rate function can be formulated as (4).

Whenever a reaction rate function is reversible, two irreversible elementary reaction rate functions are formulated, r_f for the forward and r_r for the reverse reaction. In what follows S, P, I and A stand for the concentration of substrates, products, inhibitors and activators, respectively. The notations used for the constant parameters of the reaction rate functions are adapted from the COPASI software.

– Mass Action (reversible)

$$k_1 \prod_i S_j - k_2 \prod_j P_j \quad\Longrightarrow\quad r_f = k_1 \prod_i S_j; \quad r_r = k_2 \prod_i P_j,$$

– Michaelis-Menten (reversible)

$$\frac{V_f \frac{S}{K_{ms}} - V_r \frac{P}{K_{mp}}}{1 + \frac{S}{K_{ms}} + \frac{P}{K_{mp}}} \quad\Longrightarrow\quad r_f = k_1 \frac{S}{1 + k_3 S + k_4 P}; \quad r_b = k_2 \frac{P}{1 + k_3 S + k_4 P}$$

where $k_1 = \frac{V_f}{K_{ms}}$, $k_2 = \frac{V_r}{K_{mp}}$, $k_3 = 1/K_{ms}$ and $k_4 = 1/K_{mp}$.

– Hill Cooperativity (irreversible)

$$\frac{VS^h}{K^h + S^h} = \frac{V(\frac{S}{K})^h}{1 + (\frac{S}{K})^h} \implies r = k_1 \frac{S^h}{1 + k_2 S^h},$$

where $k_1 = V/K^h$ and $k_2 = 1/K^h$.

– Ordered Bi Uni

$$\frac{V_f(S_a S_b - \frac{P}{K_{eq}})}{S_a S_b + K_{ma} S_b + K_{mb} S_a + \frac{V_f}{V_r K_{eq}} \left(K_{mp} + P\left(1 + \frac{S_a}{K_{ia}}\right)\right)} \implies$$

$$r_f = k_f \frac{S_a S_b}{1 + k_1 S_a S_b + k_3 S_b + k_4 S_a + k_5 P + k_6 P S_a};$$

$$r_b = k_r \frac{P}{1 + k_1 S_a S_b + k_3 S_b + k_4 S_a + k_5 P + k_6 P S_a},$$

where $k_1 = \frac{V_r K_{eq}}{V_f K_{mp}}$, $k_f = V_f k_1$, $k_r = V_f k_1 / K_{eq}$, $k_3 = K_{ma} k_1$, $k_4 = K_{mb} k_1$, $k_5 = k_1 / K_{mp}$ and $k_6 = k_1/(K_{mp} K_{ia})$.

– Allosteric inhibition (reversible)

$$\frac{V_f \frac{S}{K_{ms}} - V_r \frac{P}{K_{mp}}}{1 + \frac{S}{K_{ms}} + \frac{P}{K_{mp}} + (\frac{I}{K_i})^n} \implies$$

$$r_f = k_f \frac{S}{1 + k_1 S + k_2 P + k_3 I^n}$$

$$r_r = k_r \frac{P}{1 + k_1 S + k_2 P + k_3 I^n},$$

where $k_f = V_f/K_{ms}$, $k_r = V_r/K_{mp}$, $k_1 = 1/K_{ms}$, $k_2 = 1/K_{mp}$ and $k_3 = 1/K_i^n$.

– Mixed Inhibition (irreversible)

$$\frac{V \frac{S}{K_m}}{1 + \frac{I}{k_{is}} + \frac{S}{K_m} + \frac{S}{K_m} \frac{I}{k_{ic}}} \implies r = k_1 \frac{S}{1 + k_2 I + k_3 S + k_4 S I},$$

where $k_1 = V/K_m$, $k_2 = 1/k_{is}$, $k_3 = 1/K_m$ and $k_4 = 1/(K_m k_{ic})$.

– Catalytic Activation (irreversible)

$$\frac{V_{max} \frac{S}{K_{mS}} \frac{A}{K_{mA}}}{1 + \frac{S}{K_{mS}} + \frac{A}{K_{mA}} + \frac{S}{K_{mS}} \frac{A}{K_{mA}}} \implies r = k \frac{SA}{1 + k_1 S + k_2 A + k_3 SA},$$

where $k = V_{max}/(K_{mS} K_{mA})$, $k_1 = 1/K_{mS}$, $k_2 = 1/K_{mA}$ and , $k_3 = 1/(K_{mS} K_{mA})$.

– Substrate inhibition (irreversible)

$$\frac{V \frac{S}{K_m}}{1 + \frac{S}{K_m} + (\frac{S}{K_{si}})^2} \implies r = k \frac{S}{1 + k_1 S + k_2 S^2},$$

where $k = V/K_m$, $k_1 = 1/K_m$ and $k_2 = 1/K_{si}^2$.

– Substrate activation (irreversible)

$$\frac{V(\frac{S}{K_{ms}})^2}{1 + \frac{S}{K_{sc}} + \frac{S}{K_{sa}} + (\frac{S}{K_{sa}})^2} \implies r = k \frac{S^2}{1 + k_1 S + k_2 S^2},$$

where $k = V/K_{ms}^2$, $k_1 = \frac{K_{sa} + K_{sc}}{K_{sa} K_{sc}}$ and $k_2 = 1/K_{sa}^2$.

C Non-negativity of the solutions of bio-CRNs

In order to prove the non-negativity of the solution, one need to check the Lipschitz condition and the essential non-negativity of the right hand side of (7). It is easy to see, that the right hand sides of the ODEs are continuously differentiable, therefore they are locally Lipschitz.

To show the essential non-negativity of the right hand side functions, insert (11) into (7), for $p = 1, \dots n$, then the p-th equation reads as

$$f_p(x) = \sum_{l=1}^{m} Y_{pl} \cdot \sum_{j=1}^{\kappa} \tilde{A}_{\mathrm{k},lj} \cdot \varphi_j(x) \tag{34}$$

Rewriting the sum over all the κ kinetics into two sums: over the reactant complexes and over the kinetics in each of these complexes, one arrives to

$$f_p(x) = \sum_{l=1}^{m} Y_{pl} \sum_{j=1}^{m} \sum_{i=1}^{d_i} \tilde{A}_{\mathrm{k},l\, z_j+i} \cdot \varphi_{z_j+i}(x), \tag{35}$$

where $z_j = \sum_{k=1}^{j-1} d_k$. Using (11) and (4)

$$f_p(x) = \sum_{l=1}^{m} Y_{pl} \sum_{j=1}^{m} \sum_{i=1}^{d_i} \tilde{A}_{\mathrm{k},l\, z_j+i} \cdot g_{ji}(x) = \sum_{l=1}^{m} Y_{pl} \sum_{j=1}^{m} \sum_{i=1}^{d_i} \tilde{A}_{\mathrm{k},l\, z_j+i} \cdot \frac{\prod_{o=1}^{n} x_o^{Y_{oj}}}{D_{ji}(x)} \ . \tag{36}$$

From the definition of $\tilde{A}_k$ we know that the coefficients $\tilde{A}_{\mathrm{k},l\, z_j+i}$ are negative when $l = j$ and non-negative otherwise. So we decompose the summation over j into the two cases

$$f_p(x) = \sum_{l=1}^{m} Y_{pl} \sum_{j=1, j\neq l}^{m} \sum_{i=1}^{d_i} \tilde{A}_{\mathrm{k},l\, z_j+i} \cdot \frac{\prod_{o=1}^{n} x_o^{Y_o,j}}{D_{ji}(x)} - \sum_{l=1}^{m} Y_{pl} \sum_{i=1}^{d_i} |\tilde{A}_{\mathrm{k},l\, z_j+i}| \cdot \frac{\prod_{o=1}^{n} x_o^{Y_{ol}}}{D_{li}(x)} \ . \tag{37}$$

Notice that the first term is always non-negative and the second term contains the factor $Y_{pl} x_p^{Y_{pl}}$. If $Y_{pl} = 0$, then $\lim_{x_p \to 0} \frac{0 \cdot x_p^0}{D(x)} = 0$. If $Y_{pl} > 0$, since the denominator term (6) cannot approach zero, $\lim_{x_p \to 0} \frac{Y_{pl} x_p^{Y_{pl}}}{D(x)} = 0$ and f_p is indeed essentially non-negative.

D Kinetic realizability

D.1 Dynamic equations with reaction vector formalism

The balance equations can also be written using the reaction rate functions and the reaction vectors. Let denote the reaction rate corresponding to the reaction $(C_i, C_j, G_l) \in \mathcal{R}$ by r_{ijl} and the set of all reaction rates by $\nabla = \{r_{ijl}\}$ for $i = 1, \dots m$; $i = 1, \dots m$; $l = 1, \dots d_i$. Then, the balance equation reads as

$$\dot{x} = \sum_{r_{ijl} \in \nabla} r_{ijl}(x) \left(\eta^{(j)} - \eta^{(i)} \right), \tag{38}$$

where the reaction vector η^i is the i-th column of the complex composition matrix Y. Let $\{\omega_i\}$ be the standard basis on $\mathbb{R}^m$, thus $\eta^{(i)} = Y\omega_i$. Inserting above yields

$$\dot{x} = Y \sum_{r_{ijl} \in \nabla} r_{ijl}(x) \left(\omega_j - \omega_i \right).$$

Rewrite the sum for each source complex, product complex and each kinetics

$$\dot{x} = Y \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{l=1}^{d_i} r_{ijl}(x) \left(\omega_j - \omega_i \right).$$

Here we used the convention, that if the reaction $(C_i, C_j, G_l) \notin \mathcal{R}$, then the corresponding principal reaction rate coefficient $k_{ijl} = 0$, thus the reaction rate $r_{ijl}(x) = 0$. Let also use the definition of the bio-chemical reaction rate functions from (4)

$$\dot{x} = Y \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{l=1}^{d_i} k_{ijl} \frac{x^{\eta^{(i)}}}{D_{il}(x)} (\omega_j - \omega_i) \tag{39}$$

D.2 The proof of kinetic realizability

Consider the following autonomous ordinary differential equations

$$\dot{x}(t) = f(x(t)), \quad x(0) = x_0, \quad t \in [0, t_f),$$

where $f(x) : \mathbb{R}^n \to \mathbb{R}^n$ and $x \in \mathbb{R}^n$. Assume that the right hand side function f is composed by the linear combination of biochemical reaction rate functions (4). Then, there exists a bio-CRN with n species included if and only if for each $o = 1, \ldots n$ the $f_o([x_1, x_2, \ldots, x_{o-1}, 0, x_{o+1} \ldots x_n])$ is a *non-negative* linear combination of the biochemical reaction rate functions.

Proof: *(Sufficiency)* In the ith equation each term in $f_i(x)$ has the form

$$a_i \frac{x_1^{p_1} x_2^{p_2} \ldots x_i^{p_i} \ldots x_n^{p_n}}{D(x)}, \tag{40}$$

where the exponents $p_j \geq 0$ for $j = 1 \ldots n$ and specially $p_i > 0$, or

$$b_i \frac{x_1^{q_1} x_2^{q_2} \ldots x_{i-1}^{q_{i-i}} x_{i+1}^{q_{i+i}} \ldots x_n^{q_n}}{D(x)}, \tag{41}$$

where the $b_i > 0$ and the x_i is missing from the nominator, i.e. $q_i \equiv 0$. Let

$$\operatorname{sign}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}.$$

The following reaction

$$\sum_{j=1}^{n} p_j X_j \xrightarrow{\left(|a_i|, \frac{\sum_{j=1}^{n} x_j^{p_j}}{D(x)}\right)} (p_i + \operatorname{sign}(a_i)) X_i + \sum_{j=1, j\neq i}^{n} p_j X_j,$$

results the term (40) in f_i, but does not contribute to any other f_j, $j = 1, \ldots n$, $j \neq i$, since the stoichiometric coefficients for any other X_j are the same in the reactant and in the product side. Further, the reaction

$$\sum_{j=1}^{n} q_j X_j \xrightarrow{\left(b_i, \frac{\sum_{j=1}^{n} x_j^{q_j}}{D(x)}\right)} X_i + \sum_{j=1, j\neq i}^{n} q_j X_j,$$

contributes (41) in f_i, but nothing for the other f_j.

Since all terms in f can be realized by a biochemical reaction with either the biochemical rate $|a_i| \frac{\sum_{j=1}^{n} x_j^{p_j}}{D(x)}$ or $b_i \frac{\sum_{j=1}^{n} x_j^{q_j}}{D(x)}$, we proved the sufficiency.

Necessity. Let $o = 1, \dots n$ and recall (38)

$$f_o(x) = \sum_{r_{ijl} \in \nabla} r_{ijl}(x) \left(\eta_o^{(j)} - \eta_o^{(i)} \right).$$

Inserting the reaction rate functions (4) yields

$$f_o(x) = \sum_{\nabla} \frac{k_{ijl}}{D_{il}(x)} x^{\eta^{(i)}} \left(\eta_o^{(j)} - \eta_o^{(i)} \right).$$

Let $x_o = 0$. If $\eta_o^{(i)} > 0$, then the monomial $x^{\eta^{(i)}}$ is zero and therefore, the corresponding terms in the sum disappear. On the other hand, if $\eta_o^{(i)} = 0$, then the 0^0 is the problem, which is treated as

$$\frac{k_{ijl}}{D_{il}(x)} x^{\eta^{(i)}} \left(\eta_o^{(j)} - \eta_o^{(i)} \right) = \lim_{x_o \to 0^+} \frac{k_{ijl}}{D_{il}(x)} x^{\eta_1^{(i)}} x^{\eta_2^{(i)}} \dots x^{\eta_{o-1}^{(i)}} x_o^0 x^{\eta_{o+1}^{(i)}} \dots x^{\eta_n^{(i)}} \eta_o^{(j)} =$$

$$= \frac{k_{ijl}}{D_{il}([x_1, \dots, x_{o-1}, 0, x_{o-1} \dots x_n])} x^{\eta_1^{(i)}} x^{\eta_2^{(i)}} \dots x^{\eta_{o-1}^{(i)}} 1 x^{\eta_{o+1}^{(i)}} \dots x^{\eta_n^{(i)}} \eta_o^{(j)}.$$

Let $\overline{\mathcal{R}} = \{$ terms in which $\eta_o^{(i)} = 0\}$ and so

$$f_o([x_1, x_2, \dots, x_{o-1}, 0, x_{o+1} \dots x_n]) = \sum_{\overline{\mathcal{R}}} \frac{\eta_o^{(j)} k_{ijl}}{D_{il}(x)} x^{\eta_1^{(i)}} x^{\eta_2^{(i)}} \dots x^{\eta_{o-1}^{(i)}} x^{\eta_{o+1}^{(i)}} \dots x^{\eta_n^{(i)}},$$

which is indeed a *non-negative* linear combination of elementary biochemical reaction rate functions.

E Properties of the realizations

In this section we prove the three claims from Section 4.3.

Indirect proof of P1 and P2. Let $M = Y\tilde{A}_k$ and assume that $(Y, \tilde{A}_\mathrm{k}, P)$ is a dense realization of Σ and thus according to (13), $\tilde{A}_\mathrm{k}$ has the most number of positive entries among the possible solutions of $M = Y\tilde{A}_\mathrm{k}$. Further assume that $\tilde{A}'_\mathrm{k}$ is also a valid modified Kirchhoff matrix solution of $M = Y\tilde{A}'_\mathrm{k}$, but there is i, j, l, $i \neq j$ such that $\tilde{A}_{\mathrm{k},i,z_j+l} = 0$ but $\tilde{A}'_{\mathrm{k},i,z_j+l} > 0$. Then, it follows from (23) that $\tilde{A}''_\mathrm{k} = \frac{1}{2}\tilde{A}'_\mathrm{k} + \frac{1}{2}\tilde{A}_\mathrm{k}$ is also a valid dynamically equivalent realization of Σ, but $\tilde{A}''_\mathrm{k}$ has more positive elements than $\tilde{A}_\mathrm{k}$, which is a contradiction.

Proof of P3. ($\Rightarrow$) If the sparse and the dense realizations are structurally identical, then all the realizations are structurally the same, since the dense realisation is structurally unique. Thus any realisation is structurally unique. ($\Leftarrow$) If the structure of the realization is unique then the dense and the sparse realizations are trivially identical.