



# Structural simplification of chemical reaction networks preserving deterministic semantics

Guillaume Madelaine, Cédric Lhoussaine, Joachim Niehren

## ► To cite this version:

Guillaume Madelaine, Cédric Lhoussaine, Joachim Niehren. Structural simplification of chemical reaction networks preserving deterministic semantics. Computational Methods in Systems Biology (CMSB), Jérémie BOURDON; Olivier ROUX, Sep 2015, Nantes, France. pp.133-144. hal-01168038

**HAL Id: hal-01168038**

**<https://hal.archives-ouvertes.fr/hal-01168038>**

Submitted on 25 Jun 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Structural simplification of chemical reaction networks preserving deterministic semantics<sup>\*</sup>

Guillaume MADELAINE<sup>1,2</sup>, Cédric LHOSSAINE<sup>1,2</sup>, and Joachim NIEHREN<sup>1,3</sup>

<sup>1</sup> CRISTAL, UMR 9189, 59650 Villeneuve d’Ascq, France

<sup>2</sup> University of Lille, France

<sup>3</sup> INRIA Lille, France

**Abstract.** We study the structural simplification of chemical reaction networks preserving the deterministic kinetics. We aim at finding simplification rules that can eliminate intermediate molecules while preserving the dynamics of all others. The rules should be valid even though the network is plugged into a bigger context. An example is Michaelis-Menten’s simplification rule for enzymatic reactions. In this paper, we present a large class of structural simplification rules for reaction networks that can eliminate intermediate molecules at equilibrium, without assuming that all molecules are at equilibrium, i.e. in a steady state. We prove the correctness of our simplification rules for all contexts that preserve the equilibrium of the eliminated molecules. Finally, we illustrate at a concrete example network from systems biology that our simplification rules may allow to drastically reduce the size of reaction networks in practice.

## 1 Introduction

In systems biology [18], reaction networks are used to represent biological systems. They enable formal analyses [9], simulations with several semantics [7], parameter estimations and identifications [1], etc. With bigger and bigger networks, in order to keep the analyses as simple as possible, or to have quick simulations (in particular in the context of real-time control [29]), we need to be able to simplify reaction networks. Indeed, the reactions of many metabolic reaction networks are often motivated by simplifications of concrete chemical reactions, see e.g. [21], but these simplifications are always done in informal manner without any semantical guarantees. An exception is Michaelis-Menten’s simplification rule of enzymatic reactions, which is properly justified under quasi-steady-state assumption [27].

One usual approach is to simplify the ordinary differential equation (ODE) systems, that describe the deterministic semantics of reaction networks, but not the reaction networks themselves. In [17], authors presented a method based on the structure of enzyme-catalysed reactions to compute a simplified ODE system at steady-state. In [6], authors used dependency analysis of rule-based

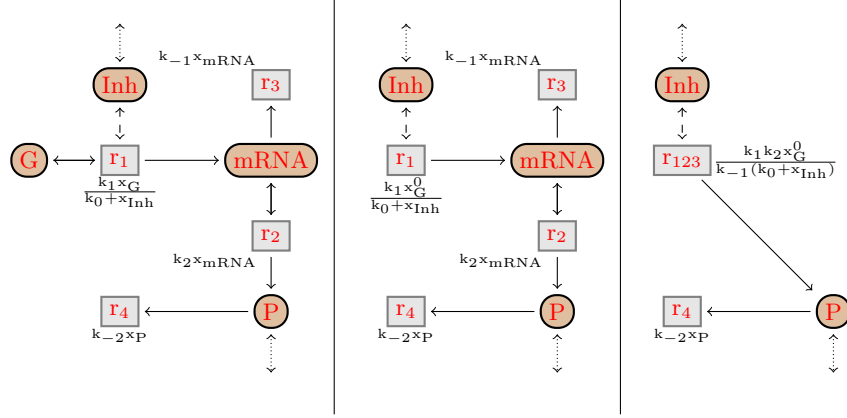
---

<sup>\*</sup> This work has been funded by the French National Research Agency research grant Iceberg ANR-IABI-3096

models to obtain a simplified ODE system. Many other simplification methods use the distinction between slow and fast reactions, as for instance methods based on invariant manifolds [11], quasi-steady state [3,27], quasi-equilibrium approximation [12] or tropicalization [28]. Other methods reduce the number of parameters, for instance by using Lie symmetries [19]. However, most of those methods require the parameter values, or at least their magnitudes, and those data are often unknown. Moreover, it is useful to preserve the reaction network and not just its ODE system, and transforming an ODE system back to a reaction network is a difficult issue, since not always possible, or not possible in a unique manner [8].

Another approach is to consider reaction networks as programs [5,25,16], and to apply simplification rules directly to such programs, similarly to what is done in compiler construction [26,24]. This means to directly simplify the reaction network and not the corresponding ODE system, or even while ignoring the kinetics all over. Such structural simplification methods are usually based on a small-step semantics, saying how chemical solutions may evolve non-deterministically. They are often contextual, i.e. the simplification rules remain correct when the network is plugged into a bigger context. In our own previous work [20], we proposed to simplify reaction networks while preserving the reachability of final components, called attractors. However, those methods do not fit well with the deterministic semantics, even though the simplification rules obtained seem sensible for biological systems. Previous structural simplification methods were presented in [10], where subgraph epimorphisms are used to reduce reaction networks. Similar works had been done in Petri Nets [2,23], preserving its usual properties (liveness, deadlock, termination, etc). In [4], Cardelli presented morphisms that preserve the deterministic semantics, but does not give simplification rules for them.

In this article, we aim at finding a new approach for simplifying reaction networks that preserves the deterministic semantics, i.e. the evolution of concentrations of molecular species over time. The approach should be structural in that it applies to reaction networks directly without computing the ODE system. It should be contextual, so that we can easily simplify modules or subnetworks in a larger context while preserving the overall dynamics. Therefore, we propose a collection of simplification rules that eliminate intermediate molecules while preserving the dynamics of all others. Some simplification rules are based on partial equilibrium conditions on the intermediate molecules (but a general steady-state is not assumed). Such conditions were already assumed to justify Michaelis-Menten’s exact simplification for enzymatic reactions [22] which is widely accepted. There the intermediate complex needs to be at equilibrium; when it is only close to the equilibrium, then a small error is made which can be estimated. A network obtained by applying a simplification rule has the same deterministic semantics than the original one, in all contexts that preserve the equilibrium conditions on intermediate molecules. For applying a simplification rule, the corresponding ODE system is not needed, and the kinetic parameters may be unknown. We illustrate the usefulness of the simplification by applying



**Fig. 1.** Reaction graphs of the *Gene* network on the left, and its two simplifications. Molecules are represented by circles, and reactions by squares. In the kinetic expressions near the reactions, the  $k_i$  are parameters while  $x_A$  is a variable representing the concentration of a molecule  $A$ .  $x_G^0$  denotes the initial concentration of  $G$ . A dash arrow means that the molecule acts as a modulator in the reaction, while a dot arrow means that the molecule can be modified by the context.

it to biological examples, where it allows to drastically reduce the size of reaction networks.

**Outline.** We first illustrate the basic ideas and motivations at an example in Section 2. We recall the formal definitions of reaction networks with their deterministic semantics in Section 3. In Section 4, we contribute a contextual equivalence relation for reaction networks, and in Section 5 a set of simplification axioms, that we prove correct with respect to this equivalence relation. In Section 6, we illustrate at a biological example, how much reaction networks can be simplified in practice. We finally conclude and discuss future work in Section 7.

## 2 Preliminary example

We first present a preliminary example, to illustrate our simplification.

Consider the reaction network *Gene* in Fig. 1 on the left. It has four species: a gene  $G$ , an inhibitor  $Inh$ , some  $mRNA$ , and a protein  $P$ . The reaction  $r_1$  describes a transcription, the production of  $mRNA$  in presence of a gene  $G$ . This gene is required to apply the reaction, but its amount is not modified by it. This reaction has also a modulator,  $Inh$ , indicated by a dashed arrow. A modulator influences the speed rate of a reaction, but is not required to apply it. Here,  $Inh$  slows down the reaction  $r_1$ . The reaction  $r_2$  is the translation of  $mRNA$  into the protein  $P$ , while the reaction  $r_3$  (resp.  $r_4$ ) describes the degradation of  $mRNA$  (resp.  $P$ ). Aside from the first one, every reaction has a simple mass-action kinetic.

In order to simplify the network, we first need to specify how the environment interacts with the network: this is indicated by pending dotted arrows in Fig. 1.

We consider here that  $G$  and  $mRNA$  are internal molecules, that is, they can not be modified by the context. Then, the context can be any set of reactions that does not contain  $G$  and  $mRNA$ . It can for instance transform  $P$  into another protein, or produces something else in presence of  $Inh$ , etc.

In this network, we are especially interested in the protein  $P$ , and on the contrary we want to eliminate the intermediate  $mRNA$ . To do that, we will assume that  $mRNA$  is at equilibrium, i.e. its concentration,  $x_{mRNA}$ , is constant over time.

Let us simplify our network. First, notice that the gene  $G$  is not modified by any reaction. It is used in reaction  $r_1$ , but only as an activator, i.e. on both sides of the reaction. Moreover,  $G$  is an internal molecule, that can not be modified by the context. Therefore its concentration,  $x_G$ , is constant over time:  $x_G = x_G^0$ . Then we make this modification in the kinetic expression of reaction  $r_1$ , and remove completely  $G$  from the network. The new network is pictured in Fig. 1 (middle).

Now, consider the intermediate  $mRNA$ . It is an internal molecule, and its (complete) ordinary differential equation is:

$$\frac{dx_{mRNA}}{dt} = \frac{k_1 x_G^0}{k_0 + x_{Inh}} - k_{-1} x_{mRNA}$$

Since we assumed that  $mRNA$  is at equilibrium, i.e.  $\frac{dx_{mRNA}}{dt} = 0$ , we deduce:

$$x_{mRNA} = \frac{k_1 x_G^0}{k_{-1}(k_0 + x_{Inh})}$$

Therefore we remove  $mRNA$  from the network, and replace the variable  $x_{mRNA}$  in the kinetics of reaction  $r_2$ , by the expression computed above. We obtain the simplified network in Fig. 1 (right) where  $r_1$ ,  $r_2$  and  $r_3$  are merged into the new reaction  $r_{123}$ .

As we will see in this paper, the simplification rules used above preserve the deterministic semantics of reaction networks, in every context. So the simplified network is contextual equilibrium-equivalent to the first one. Note that we can not simplify the network anymore, since both  $Inh$  and  $P$  can be modified by the context.

### 3 Reaction networks

We introduce reaction networks and define their deterministic semantics in terms of ordinary differential equations.

Let  $Spec$  be a set of molecular species ranging over by  $A, B, C$ . We define a (chemical) solution  $s \in Sol : Spec \rightarrow \mathbb{N}_0$  as a function from molecular species to natural numbers. Given natural numbers  $n_1, \dots, n_k$ , we denote by  $n_1 A_1 + \dots + n_k A_k$  the solution that contains  $n_i$  molecules of species  $A_i$  for all  $1 \leq i \leq k$  and 0 molecule of all other species.

A *kinetic reaction*  $r = (s_1 \rightarrow s_2 ; e)$  is a pair composed of a *reaction*  $s_1 \rightarrow s_2$  and a *kinetic expression*  $e$ . The reaction transforms the solution  $s_1$ , called *reactants*, into the solution  $s_2$ , called *products*. The molecules present in the same amount in both reactants and products are called *activators*. They are not modified by the reaction, but are required to apply it. Kinetic expressions are symbolic functions defined from *concentration variables*,  $Vars_{Spec} = \{x_A \mid A \in Spec\}$ , symbols of *initial concentrations*,  $Const_i = \{x_A^0 \mid A \in Spec\}$ , and symbols of *kinetic parameters*,  $Const_k = \{k_0, k_1, \dots\}$ :

$$e, f, \dots ::= x \mid x^0 \mid k \mid e + f \mid e - f \mid e \times f \mid e/f \mid -e \mid (e)$$

where  $x \in Vars_{Spec}$ ,  $x^0 \in Const_i$  and  $k \in Const_k$ . As usual, we also simply denote  $ef$  for  $e \times f$ .

The (*chemical*) *concentration* of a chemical species is a function from time to positive numbers  $\mathbb{R}_+ \rightarrow \mathbb{R}_+$ . Kinetic expressions are interpreted as actual kinetic functions by means of an *assignment*  $\alpha$  that maps concentration variables to concentrations ( $\alpha_c$ ), initial concentrations to non negative real values ( $\alpha_0$ ) and kinetic parameters to non negative real values ( $\alpha_k$ ):

$$\alpha_c : Vars_{Spec} \rightarrow (\mathbb{R}_+ \rightarrow \mathbb{R}_+) \quad \alpha_0 : Const_i \rightarrow \mathbb{R}_+ \quad \alpha_k : Const_k \rightarrow \mathbb{R}_+$$

We only consider assignments  $\alpha$  consistent on initial concentrations, that is for any species  $A$ ,  $\alpha_c(x_A)(0) = \alpha_0(x_A^0)$ . Given an assignment  $\alpha$ , the interpretation  $[e]_\alpha$  of a kinetic expression  $e$  is thus defined as follows:

$$\begin{aligned} [x]_\alpha(t) &= \alpha_c(x)(t) & [x^0]_\alpha(t) &= \alpha_0(x^0) & [k]_\alpha(t) &= \alpha_k(k) & [(e)]_\alpha(t) &= [e]_\alpha(t) \\ [-e]_\alpha &= -[e]_\alpha & [e \text{ op } f]_\alpha(t) &= [e]_\alpha(t) \text{ op } [f]_\alpha(t) \text{ where } op \in \{+, -, \times, /\} \end{aligned}$$

Given a set of kinetic reactions, we only consider assignments  $\alpha$  such that for any kinetic expression  $e$  occurring in this network, its interpretation  $[e]_\alpha : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a continuously differentiable function from time to non negative real numbers, standing for the actual reaction rate. Kinetic reactions  $(s_1 \rightarrow s_2 ; e)$  also have to respect the following *coherence property*: the actual rate given by any assignment  $\alpha$  is equal to zero if and only if one of the reactants is not present:  $\forall \alpha. [e]_\alpha(t) = 0$  iff  $\exists A \in s_1. [x_A]_\alpha(t) = 0$ . Note that a kinetic expression can contain concentration variables of molecules that are not present in the reactants of the reaction; such molecules, called *modulators*, are not required to apply the reaction, but modify its rate.

**Definition 1.** A *reaction network* is a pair  $\langle I, R \rangle$ , composed of a set of internal molecules  $I$ , which specifies that some molecules can not interact with the context, and a set of kinetic reactions  $R$ .

From any network  $N = \langle I, R \rangle$  and from its kinetic expressions, we can infer a *system of ordinary differential equations* defined by

$$ODE(N) = \left[ \frac{dx_A}{dt} = \sum_{(s_1 \rightarrow s_2 ; e) \in R} (s_2(A) - s_1(A))e \right]_{A \in Spec}$$

Given any assignment  $\alpha_0$  of the initial concentrations and any assignment  $\alpha_k$  of the kinetic parameters, by the Cauchy-Lipschitz theorem, the system  $ODE(N)$  has a unique differentiable solution  $\alpha_c$ , defined on a maximal interval including 0. Moreover, we only consider solutions  $\alpha_c$  defined on (at least)  $[0, +\infty[$ . Otherwise, we say that  $N$  has no *valid solution* for these assignments.

An *equilibrium condition*  $e$  is defined similarly to kinetic expressions and interpreted as function from time to positive numbers. It is satisfied by an assignment  $\alpha$  iff  $\alpha_c$  satisfies  $\frac{de}{dt} = 0$  given the initial concentration and parameter assignments  $\alpha_0$  and  $\alpha_k$ . An equilibrium condition can for instance impose the equilibrium of a particular molecule (for instance  $e = x_A$ ), a solution ( $e = \sum_{A \in s} s(A)x_A$ ), or a reaction ( $e = f$  for the reaction  $(r ; f)$ ). We denote by  $\mathbf{E}$  a set of equilibrium conditions. Given a network  $N$  and equilibrium conditions  $\mathbf{E}$ , the deterministic dynamics of  $N$  that satisfies  $\mathbf{E}$  is defined as

$$\begin{aligned} sol(N, \mathbf{E}) = \{ \alpha \mid \alpha_c \text{ satisfies } \mathbf{E} \text{ and is a valid solution of } ODE(N) \\ \text{for initial concentrations } \alpha_0 \text{ and parameter assignments } \alpha_k \} \end{aligned}$$

Since we are particularly interested in the molecules that are not at equilibrium, we say that two assignments  $\alpha$  and  $\alpha'$  are *equal modulo equilibrium conditions*, denoted  $\alpha_{\mathbf{E}} \alpha'$ , if they are equal on those molecules.

## 4 Contextual Equilibrium-Equivalence

We present here a notion of weak equilibrium-equivalence between reaction networks, then the definition of contexts, and finally the contextual equilibrium-equivalence.

**Definition 2 (Weak Equilibrium-Equivalence).** *Two networks  $N$  and  $M$  are weakly equilibrium-equivalent for  $\mathbf{E}$ , denoted  $N \sim^{\mathbf{E}} M$ , if they have the same solutions modulo equilibrium conditions  $sol(N, \mathbf{E}) =_{\mathbf{E}} sol(M, \mathbf{E})$ .*

A *context*  $\mathcal{C}$  is itself a reaction network. Given a set of internal molecules  $I$ , we say that a context  $\mathcal{C}$  is *compatible* with  $I$  if  $\forall A \in I, A$  has no occurrence in  $\mathcal{C}$ . We denote by  $Context(I)$  the set of compatible contexts with  $I$ . Given a network  $N = \langle I, R \rangle$  and a compatible context  $\mathcal{C} = \langle I', R' \rangle \in Context(I)$ , we denote by  $\mathcal{C}[N] = \langle I \cup I', R \cup R' \rangle$  the network placed into the context.

**Definition 3 (Contextual Equilibrium-Equivalence).** *Let  $\mathbf{E}$  be an equilibrium, the reaction networks  $N = \langle I, R \rangle$  and  $M = \langle I', R' \rangle$  are contextually equilibrium-equivalent for  $\mathbf{E}$ , denoted  $N \equiv^{\mathbf{E}} M$ , if they are weakly equilibrium-equivalent in any compatible context, i.e.  $\forall \mathcal{C} \in Context(I \cup I'). \mathcal{C}[N] \sim^{\mathbf{E}} \mathcal{C}[M]$ .*

## 5 Simplification axioms

In this section, we present some simplification axioms, that transform a network into a contextually equilibrium-equivalent network. The soundness proofs

of those axioms are given in the annex<sup>4</sup>. These simplification axioms reduce the size of a reaction network, either by completely removing a molecule from the set of reactions, by decreasing the number of reactions, or by simplifying a reaction.

We first present 2 simple simplification axioms, followed by 4 instances of a more general axiom, based on the presence of an intermediate molecule. Finally, we present this general axiom. Notice that the axioms are quite similar to the ones we presented for the attractor equivalence with a qualitative and observational semantics in [20].

The first 2 simplification axioms are given in Fig. 2. The first one, **(USELESS)**, deletes a reaction  $s \rightarrow s$  that does not impact the network dynamics. The axiom **(ACTIVATOR)** removes an internal molecule  $A$  only used as an activator in the reactions (i.e. is always present in the same amount in both sides of the reaction). It is for instance the case for the gene  $G$  in the *Gene* network in Section 2.

$$\frac{}{\langle \emptyset, \{(s \rightarrow s; e)\} \rangle \equiv^{\mathbf{E}} \langle \emptyset, \emptyset \rangle} \text{ (USELESS)}$$

$$\frac{\forall (s_1 \rightarrow s_2; e) \in R, s_1(A) = s_2(A)}{\langle \{A\}, R \rangle \equiv^{\mathbf{E}} \langle \{A\}, \{(r \setminus A; e[x_A^0/x_A]) \mid (r; e) \in R\} \rangle} \text{ (ACTIVATOR)}$$

**Fig. 2.** Simple simplification axioms.

The next four axioms in Fig. 3 are instances of the more general axiom **(INTERMEDIATE)**. These axioms aim at eliminating an internal and intermediate molecule which is at equilibrium.

In the first one, **(INTER)**, the intermediate molecule  $A$  is only used in two reactions, one time as the unique product, and the other as the unique reactant. Since  $A$  is at equilibrium, the kinetic expressions of these reactions have to be equal, i.e.  $e = k_2 x_A$ . The axiom removes  $A$  and merges both reactions into one, keeping only the kinetic expression  $e$ . The parameter  $k_2$  is eliminated.

The second axiom, **(MICHAELIS-MENTEN)**, simplifies a three-steps enzyme-catalyzed transformation. A substrate  $S$  binds to an enzyme  $E$  to form the complex  $C$ . Then the complex either transforms back to  $S + E$ , or produces the product  $P$  while releasing  $E$ . Assuming that the enzyme  $E$  and the complex  $C$  are at equilibrium, the axiom merges the reactions into a unique one, that directly transforms  $S$  into  $P$ . The equilibrium of  $C$  imposes that the simplified reaction has a Michaelis-Menten kinetics of the form  $V \frac{x_S}{x_S + K}$  [22].

The last two, **(CASCADE<sub>1</sub>)** and **(CASCADE<sub>2</sub>)**, concern a cascade of reactions, where the intermediate molecule  $A$ , at equilibrium, is produced in presence of some activators  $s$ , and then is either degraded or used to produce some  $s'$ . The axioms eliminate  $A$ , so the simplified networks directly produced  $s'$  in presence of  $s$ . The simplified kinetic expressions are obtained by computing the value of  $x_A$  at equilibrium, and by replacing it in the third kinetic reaction.

<sup>4</sup> [www.cristal.univ-lille.fr/~guillaume.madelaine/doc/2015\\_structural\\_simplification.pdf](http://www.cristal.univ-lille.fr/~guillaume.madelaine/doc/2015_structural_simplification.pdf)



$$\begin{array}{c}
\frac{x_A \in \mathbf{E} \quad A \notin s, s'}{\langle \{A\}, \{(s \rightarrow A; e), (A \rightarrow s'; k_2 x_A)\} \rangle \equiv^{\mathbf{E}} \langle \{A\}, \{(s \rightarrow s'; e)\} \rangle} \text{ (INTER)} \\
\\
\frac{x_C, x_E \in \mathbf{E}}{\begin{array}{c} \langle \{E, C\}, \\ \{(E + S \rightarrow C; k_1 x_E x_S), \\ (C \rightarrow E + S; k_{-1} x_C), \\ (C \rightarrow E + P; k_2 x_C)\} \rangle \\ \equiv^{\mathbf{E}} \langle \{E, C\}, \\ \{(S \rightarrow P; k_2(x_E^0 + x_C^0) \frac{x_S}{x_S + \frac{k_2 + k_{-1}}{k_1}})\} \rangle \end{array}} \text{ (MICHAELIS-MENTEN)} \\
\\
\frac{x_A \in \mathbf{E} \quad A \notin s, s'}{\begin{array}{c} \langle \{A\}, \{(s \rightarrow s + A; e), \\ (A \rightarrow \emptyset; k_{-1} x_A), (A \rightarrow s'; k_2 x_A)\} \rangle \\ \equiv^{\mathbf{E}} \langle \{A\}, \{(s \rightarrow s + s'; \frac{k_2}{k_{-1} + k_2} e)\} \rangle \end{array}} \text{ (CASCADE}_1\text{)} \\
\\
\frac{x_A \in \mathbf{E} \quad A \notin s, s'}{\begin{array}{c} \langle \{A\}, \{(s \rightarrow s + A; e), \\ (A \rightarrow \emptyset; k_{-1} x_A), (A \rightarrow A + s'; e')\} \rangle \\ \equiv^{\mathbf{E}} \langle \{A\}, \{(s \rightarrow s + s'; e'[\frac{e}{k_{-1}}/x_A])\} \rangle \end{array}} \text{ (CASCADE}_2\text{)}
\end{array}$$

**Fig. 3.** Instances of intermediate molecule axiom.

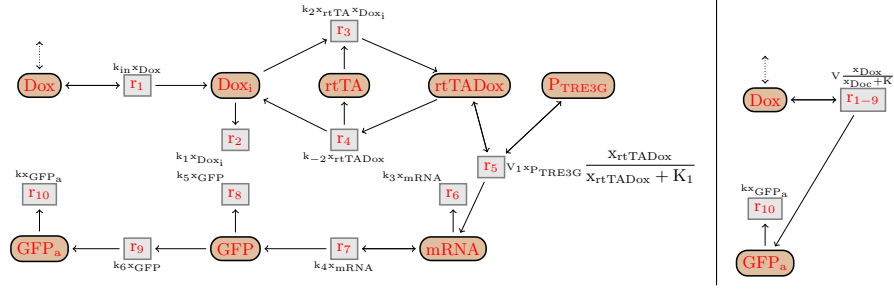
We finally present in Fig. 4 the general axiom (**INTERMEDIATE**). In this axiom, we consider an intermediate internal molecule  $A$ , at equilibrium. It simplifies a model with one reaction that can produce  $A$ , with a (non-empty) set of reactions that has only  $A$  as reactant and whose kinetic expressions are linear in  $x_A$ , and possibly a set of reactions with  $A$  as activator. Then the axiom eliminates  $A$ , and merges two-by-two the reactions. The linearity of the kinetic expression of some reactions is necessary to easily compute the expression of  $x_A$  at equilibrium, that is in this case  $x_A = \sum_j e_j / \sum_l e'_l$ .

$$\begin{array}{c}
\text{(INTERMEDIATE)} \\
\frac{x_A \in \mathbf{E} \quad \forall t. (\sum_l e'_l)(t) \neq 0 \quad A \notin s^{(1)}, s^{(2)}, s_l^{(2')}, s_m^{(1'')}, s_m^{(2'')} \quad x_A \notin e'_l}{\begin{array}{c} \langle \{A\}, \{(s^{(1)} \rightarrow s^{(2)} + A; e) \cup \\ \{(A \rightarrow s_l^{(2')}; x_A e'_l)\}_l \cup \\ \{(s_m^{(1'')} + A \rightarrow s_m^{(2'')} + A; e''_m)\}_m \rangle \\ \equiv^{\mathbf{E}} \langle \{A\}, \{(s^{(1)} \rightarrow s^{(2)} + s_l^{(2')}; e'_l \frac{e}{\sum_l e'_l})\}_l \cup \\ \{(s_m^{(1'')} + s^{(1)} \rightarrow s_m^{(2'')} + s^{(1)}; e''_m [\frac{e}{\sum_l e'_l}/x_A])\}_m \rangle \end{array}}
\end{array}$$

**Fig. 4.** General intermediate molecule axiom.

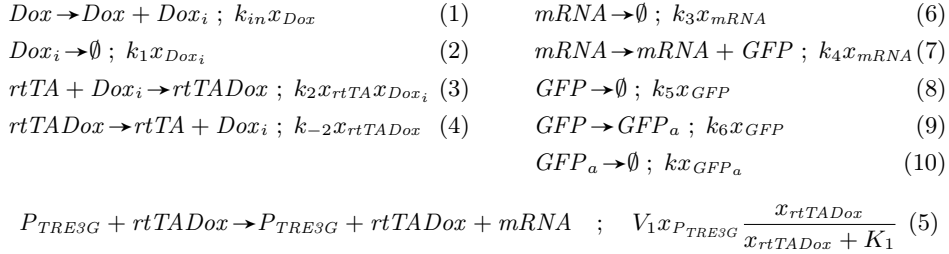
## 6 Simplification of the *Tet-On* reaction network

We present here the simplification of the *Tet-On* system [13,14,15] using our axioms. The initial *Tet-On<sub>detailed</sub>* reaction network, depicted in Fig. 5 (left), has 10 reactions and 11 parameters. We simplify it into the contextually equilibrium-equivalent *Tet-On<sub>simple</sub>* network, depicted on Fig. 5 (right), with only two reactions and 3 parameters.



**Fig. 5.** Reaction graphs of the detailed (left) and simplified (right) *Tet-On* networks. Molecules are represented by circles, and reactions by squares. In the kinetic expressions near the reactions, the  $k_i$  are parameters while  $x_A$  is a variable representing the concentration of a molecule  $A$ . A dash arrow means that the molecule acts as a modulator in the reaction, while a dot arrow means that the molecule can be modified by the context. In the right network, the parameters are  $V = x_{P_{TRE3G}}^0 V_1 k_4 k_6 / k_3 (k_5 + k_6)$  and  $K = k_1 k_{-2} K_1 / x_{rtTA}^0 k_{in} k_2$ .

The *Tet-On* system [13,14,15] describes how the production of activated green fluorescent proteins ( $GFP_a$ ) in a cell can be stimulated by the presence of doxycycline ( $Dox$ ) outside the cell. The detailed network is  $Tet-On_{detailed} = \langle I, R \rangle$  where every molecule is internal except for  $Dox$  (i.e.  $I = Spec \setminus Dox$ ), and  $R$  is the set of reactions from Fig. 6, inspired by the *Tet-On* model from [15].



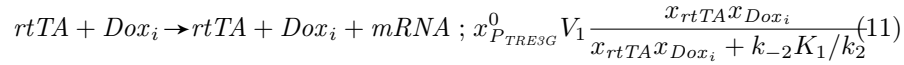
**Fig. 6.** Reactions of the detailed *Tet-On<sub>detailed</sub>* network.

In the network, the doxycycline  $Dox$  moves into the cell and becomes  $Dox_i$  by reaction (1). We assume here that the amount of  $Dox$  is controlled by the environment (for instance by a microfluidics device [30]), and therefore the network can not modify its concentration. Then  $Dox_i$  is either degraded by reaction (2), or binds to the artificial transcription factor  $rtTA$  by reaction (3). The complex  $rtTADox$  either dissociates (4), or activates the transcription of the gene  $P_{TRE3G}$ , producing  $mRNA$  (5).  $mRNA$  either degrades (6) or is translated into

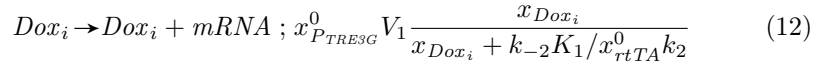
*GFP* (7). Finally, *GFP* needs to be activated into *GFP<sub>a</sub>* (9) in order to become fluorescent and thus observable by a microscope. Both *GFP* and *GFP<sub>a</sub>* can also be degraded (8, 10).

We are particularly interested by *GFP<sub>a</sub>*, since it is the only experimentally observable molecule. Therefore we assume that all other molecules are at equilibrium, i.e.  $\mathbf{E} = \{x_X \mid X \in \text{Spec} \setminus \text{GFP}_a\}$ . The simplification follows the axioms from Fig. 2, Fig. 3 and Fig. 4, so that will prove that the two networks are contextually equilibrium-equivalent for  $\mathbf{E}$ . Note that in the following simplification, for the sake of readability, some kinetic expressions were sometimes slightly rewritten into equivalent expressions.

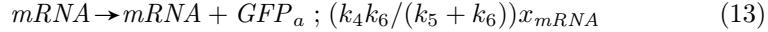
Let us first remark that the gene *P<sub>TRE3G</sub>* is only used as an activator, in the reaction 5. So we apply the axiom (ACTIVATOR), removing *P<sub>TRE3G</sub>* from this reaction, while replacing  $x_{P_{TRE3G}}$  by  $x_{P_{TRE3G}}^0$  in its kinetic function. Then *rtTADox* is an internal molecule at equilibrium, present in three reactions: one that produces it (3), one that consumes it (4), and one that uses it as an activator (5). Then we use the axiom (INTERMEDIATE) on it, followed directly by (USELESS), and merge the three reactions into:



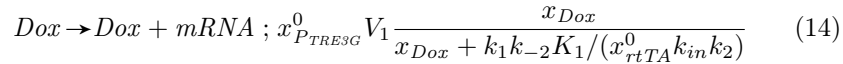
*rtTA* is only used as activator, so we apply (ACTIVATOR) and simplify (11) into:



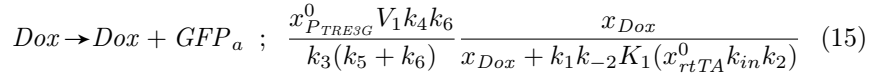
Apply axiom (CASCADE<sub>1</sub>) on *GFP*, replacing the reactions (7), (8) and (9) by:



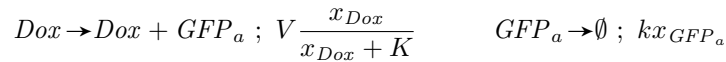
Also, apply (CASCADE<sub>2</sub>) on *Dox<sub>i</sub>*, and replace reactions (1), (2), and (12) by:



Finally we use the axiom (INTERMEDIATE) followed by (USELESS) on *mRNA*, and merge the reactions (6), (13) and (14) into:



Defining two new parameters  $V = x_{P_{TRE3G}}^0 V_1 k_4 k_6 / (k_3 (k_5 + k_6))$  and  $K = k_1 k_{-2} K_1 / (x_{rtTA}^0 k_{in} k_2)$ , we eventually obtain the following reaction network:



Notice that, aside from the kinetics, the simplified network is equal to the one we obtained with our qualitative simplification in [20].

## 7 Conclusion

We presented a new structural simplification of reaction networks, that preserved the deterministic semantics. The simplification is contextual, and is based on equilibrium conditions on intermediate molecules. We shown the usefulness of the simplification by applying it on two biological networks.

We are currently implementing the simplification algorithm, with a more complete set of axioms and compatible with the SBML format. This axioms include variants of the axioms presented here, for instance with different equilibrium conditions, but also other types of axioms, using for instance symmetries in the network. We plan to apply the simplification more systematically to biological systems. It would also be interesting to compare in depth the power of our structural simplification rules to that of the King-Altman method on ODE system [17]. On the theoretical side, as future work, we want to investigate an approximated equivalence, with approximated equilibrium conditions, and to compute the maximal error of a simplification. A similar simplification method with a stochastic semantics will also be considered.

**Acknowledgment:** The authors would like to thank Michel Petitot for its useful discussions as well as members of the *PalBioSys* research network.

## References

1. M. Ashyraliyev, Y. Fomekong-Nanfack, J. A. Kaandorp, and J. G. Blom. Systems biology: parameter estimation for biochemical models. *Febs Journal*, 276(4):886–902, 2009.
2. G. Berthelot and G. Roucairol. Reduction of petri-nets. In *Mathematical Foundations of Computer Science 1976*, LNCS. 1976.
3. M. Bodenstein. Eine theorie der photochemischen reaktionsgeschwindigkeiten. *Z. phys. Chem*, 85(329):0022–3654, 1913.
4. L. Cardelli. Morphisms of reaction networks that couple structure to function. *BMC systems biology*, 8(1):84, 2014.
5. L. Cardelli and G. Zavattaro. *On the computational power of biochemistry*. Springer, 2008.
6. V. Danos, J. Feret, W. Fontana, R. Harmer, and J. Krivine. Abstracting the differential semantics of rule-based models: exact and automated model reduction. In *Logic In Computer Science (LICS)*, pages 362–381. IEEE, 2010.
7. H. De Jong. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9(1), 2002.
8. F. Fages, S. Gay, and S. Soliman. Inferring reaction models from odes. In *Computational Methods in Systems Biology*, pages 370–373. Springer, 2012.
9. F. Fages and S. Soliman. Formal cell biology in biocham. In *Formal Methods for Computational Systems Biology*, pages 54–80. Springer, 2008.
10. S. Gay, S. Soliman, and F. Fages. A graphical method for reducing and relating models in systems biology. *Bioinformatics (Oxford, England)*, 26(18), Sept. 2010.
11. A. N. Gorban and I. V. Karlin. Method of invariant manifold for chemical kinetics. *Chemical Engineering Science*, 58(21):4751–4768, 2003.

12. A. N. Gorban, I. V. Karlin, P. Ilg, and H. C. Öttinger. Corrections and enhancements of quasi-equilibrium states. *Journal of non-newtonian fluid mechanics*, 96(1):203–219, 2001.
13. M. Gossen and H. Bujard. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 89(12), June 1992.
14. M. Gossen, S. Freundlieb, G. Bender, G. Müller, W. Hillen, and H. Bujard. Transcriptional activation by tetracyclines in mammalian cells. *Science (New York, N.Y.)*, 268(5218), June 1995.
15. Z. Huang, C. Moya, A. Jayaraman, and J. Hahn. Using the Tet-On system to develop a procedure for extracting transcription factor activation dynamics. *Molecular bioSystems*, 6(10), Oct. 2010.
16. M. John, C. Lhoussaine, J. Niehren, and C. Versari. Biochemical reaction rules with constraints. In *20th European Symposium on Programming Languages (ESOP)*, pages 338–357. Springer, 2011.
17. E. L. King and C. Altman. A schematic method of deriving the rate laws for enzyme-catalyzed reactions. *The Journal of physical chemistry*, 60(10):1375–1378, 1956.
18. H. Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, 2002.
19. F. Lemaire, A. Sedoglavic, and A. Urguplu. Moving frame based strategies for reduction of ordinary differential/recurrence systems using their expanded lie point symmetries. 2008.
20. G. Madelaine, C. Lhoussaine, and J. Niehren. Attractor equivalence: An observational semantics for reaction networks. In *Formal Methods in Macro-Biology*, pages 82–101. Springer, 2014.
21. U. Mäder, A. G. Schmeisky, L. A. Flórez, and J. Stülke. SubtiWiki—a comprehensive community resource for the model organism *Bacillus subtilis*. *Nucleic Acids Res.*, 40:1278–1287, Jan. 2012.
22. L. Michaelis and M. L. Menten. Die kinetik der invertinwirkung. *Biochem. z.*, 49(333-369):352, 1913.
23. T. Murata and J. Koh. Reduction and expansion of live and safe marked graphs. *Circuits and Systems, IEEE Transactions on*, 27(1), 1980.
24. A. M. Pitts. Operational semantics and program equivalence. In *Applied Semantics*, volume 2395 of *LNCS*. Springer, 2000.
25. A. Regev and E. Shapiro. Cellular abstractions: Cells as computation. *Nature*, 419(6905):343–343, 2002.
26. M. Schmidt-Schauss, D. Sabel, J. Niehren, and J. Schwinghammer. Observational Program Calculi and the Correctness of Translations. *Journal of Theoretical Computer Science (TCS)*, 577:98–124, Apr. 2015.
27. L. A. Segel and M. Slemrod. The quasi-steady-state assumption: a case study in perturbation. *SIAM review*, 31(3):446–477, 1989.
28. S. Soliman, F. Fages, O. Radulescu, et al. A constraint solving approach to tropical equilibration and model reduction. In *WCB-ninth Workshop on Constraint Based Methods for Bioinformatics, colocated with CP 2013*, 2013.
29. J. Uhlenendorf, S. Bottani, F. Fages, P. Hersen, and G. Batt. Towards real-time control of gene expression: controlling the hog signaling cascade. In *Pacific Symposium On Biocomputing*, pages 338–349. World Scientific, 2011.
30. J. Uhlenendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt, and P. Hersen. Long-term model predictive control of gene expression at the population and single-cell levels. *Proceedings of the National Academy of Sciences*, 109(35), 2012.