Structural simplification of chemical reaction networks in partial steady states

Guillaume Madelaine, Cédric Lhoussaine, Joachim Niehren, Elisa Tonello

Structural simplification of chemical reaction networks in partial steady states[☆]

Guillaume Madelaine[a,b,*], Cédric Lhoussaine[a,b], Joachim Niehren[a,c], Elisa Tonello[d]

[a] *CRIStAL, UMR 9189, 59650 Villeneuve d'Ascq, France*
[b] *University of Lille, France*
[c] *INRIA Lille, France*
[d] *University of Nottingham, United Kingdom*

Abstract

We study the structural simplification of chemical reaction networks with partial steady state semantics assuming that the concentrations of some but not all species are constant. We present a simplification rule that can eliminate intermediate species that are in partial steady state, while preserving the dynamics of all other species. Our simplification rule can be applied to general reaction networks with some but few restrictions on the possible kinetic laws. We can also simplify reaction networks subject to conservation laws. We prove that our simplification rule is correct when applied to a module of a reaction network, as long as the partial steady state is assumed with respect to the complete network. Michaelis-Menten's simplification rule for enzymatic reactions falls out as a special case. We have implemented an algorithm that applies our simplification rule repeatedly and applied it to reaction networks from systems biology.

Keywords: bioinformatics; systems biology; reaction network; contextual equivalence.

1. Introduction

Reaction networks [1] are systems of chemical reactions, which are widely used in systems biology [2] in order to model and analyze the dynamics of molecular biological systems [3, 4, 5]. Reaction networks have a formal semantics that describes the evolution of chemical solutions over time (without fixing the initial solution). This semantics lays the foundation for formal analysis of biological systems. In the present paper, we focus on the deterministic semantics which, in the thermodynamic limit [6], describes the average concentrations of species in chemical solutions, rather than considering the stochastic semantics which models their probability distributions.

Formal analysis methods for biological systems are based on their reaction networks. If all kinetic parameters are known, reaction networks can be used to simulate biological systems [7]. Otherwise, the missing parameters can be estimated so that they fit with experimental data [8]. Reaction networks can also be used to verify properties of biological systems [9], to control biological systems in real time [10], or to predict network changes such as gene knockouts leading to desired behavior [11, 12, 13]. The analysis techniques to solve these tasks may be either numeric or symbolic. Numeric techniques are usually based on the ordinary differential equations (ODEs) for the deterministic semantics only, while assuming the knowledge of all parameters. Symbolic analysis methods, in contrast, may be based on the structure of a reaction network and not only the ODEs. Structural methods are sometimes advantageous as argued in [14] for instance.

Given that biological systems are highly complex, their models by reaction networks may become huge (see e.g. [15]). Furthermore, a large number of kinetic parameters tend to be unknown, which limits the applicability of

[☆] This work has been funded by the French National Research Agency research grant Iceberg ANR-IABI-3096.
[*] Corresponding author: guillaume.madelaine@ed.univ-lille1.fr

formal analysis techniques. Possible workarounds are to analyze only smaller modules of reaction networks, or to simplify bigger reaction networks in order to reduce their size and the number of their unknown parameters. Therefore, various simplification methods for reaction networks were developed (see [16] for an overview), often motivated by Michaelis-Menten's seminal reduction of enzymatic reactions [17].

Structural simplification methods that rewrite reaction networks rather than Odes are particularly relevant for biological modeling. Indeed, biologists are spending much effort in designing reaction networks in practice, as for instance on the development of the metabolic and regulation networks for *B. Subtilis* in the Subtil Wiki [15]. Even though many of the reactions used there are motivated by structural simplification of larger reaction networks with more details, no such formal justification is given.

A purely structural simplification method for reaction networks without kinetic rates was developed in [18]. It consists of a set of rewrite rules that remove intermediate species in reaction networks. For instance, we can simplify the network with the following two reactions on the left into the single reaction on the right, when considering B as an intermediate species:

$$\left.\begin{array}{l} A_1 + \ldots + A_n \to B \\ B \to C_1 + \ldots + C_m \end{array}\right\} \Rrightarrow A_1 + \ldots + A_n \to C_1 + \ldots + C_m \quad (1)$$

These two networks perform the same transformations on chemical solutions that do not contain intermediate B molecules, if one ignores intermediate results. Indeed, this simplification rule is correct with respect to the attractor equivalence (also proposed in [18]). There, a reaction network is considered as non-deterministic program (operating on multisets), so that notions of observational program semantics from compiler construction [19] can be applied. A main weakness of this approach is that it can neither deal with kinetic rates nor with deterministic Ode semantics. Furthermore, some other rules for intermediate elimination, that one might want to have, turn out to be incorrect, as for instance:

$$\left.\begin{array}{l} A \to B \\ A' + B \to C \end{array}\right\} \quad \Rrightarrow \quad A + A' \to C \qquad (2)$$

To see the different input-output behavior, consider the multiset $nA + 0A' + 0B$. With the reaction network on the left, this chemical solution can be transformed to $0A + 0A' + nB$, while with the reaction on the right no transformation is possible.

An approach for the structural simplification of reaction networks with kinetic rates subject to Ode semantics was presented by Radulescu [20, 16]. The basic idea is also to remove intermediate species in reaction networks, but now assuming that the intermediate species are in quasi steady state. In a first phase, the graph structure of the reaction network is simplified. But rather than eliminating the intermediate species step by step, elementary flux modes [21] are computed in order to remove all intermediates simultaneously. The kinetic rates are then assigned to the simplified network, under the condition that the intermediates were in quasi steady state. Interestingly, the simplification rule in (2), which is wrong with respect to the nondeterministic attractor equivalence, becomes correct for the deterministic Ode semantics, when assuming that the concentrations of the intermediate species are steady. The point is that the concentration of B is not steady in the chemical solution $nA + 0A' + 0B$ on which the reaction network on the left could act but not the one on the right. Indeed a key idea of Radulescu's approach is to resolve the steady state equations for an intermediate species B after its concentration variable x_B. Since this is not always possible in an exact manner, an approximate method is proposed. The main weakness of the approach is that its correctness is not formally stated. The two main obstacles on the way to a correctness statement are the use of approximations, and the fact that kinetic rates are assigned only at the end to the reduced network.

The objective of the present article is to overcome the shortcomings of the two previous approaches. We search for a rewrite rule that can eliminate intermediate species in reaction networks with kinetic rates while exactly preserving the Ode semantics. Furthermore, our rewrite rule should remain correct when applied to a module of a reaction network, and should remain applicable when the parameters of the kinetics laws are unknown. For this, we are ready to assume that the concentration of intermediate species are steady, i.e. that the concentration of the intermediate B satisfies $x_B = c$ for some positive constant function c. Such partial steady state assumptions are more flexible than general steady state assumptions in that the concentrations of some species may still change over time. In contrast to quasi steady state assumptions no ap-

proximation is made.

The present article extends a conference paper at CMSB'2015 [22]. The contributions are as follows. We first present a rewrite rule for the structural simplification of reaction with kinetic rates, which removes one intermediate species at a time. In order to avoid approximations, we assume that all kinetic rates of the network are either linear in the concentration of the intermediate species, or else constructed in a uniform manner by applying some invertible function. More precisely, for any intermediate species B, there must be an invertible function f, common to all reactions consuming B, and such that their kinetic laws are some product of the form $f(x_B)e$, where e is an expression not containing x_B and possibly varying with the reactions. We second prove that our rewrite rule does preserve the Ode semantics of the reaction network, even if the network is plugged into any possible context. Compared to the conference version we permit networks with conservation laws. For instance, such an expression may impose that $x_A + x_B = c$ for some positive constant function c, i.e., that the sum of the concentrations of A and B is steady. Beside of being more general, the addition of conservation laws also simplifies the presentation of the rewrite rule and of the correctness proof.

An interesting question is whether repeated elimination of intermediates with our rewrite rule does always produce the same set of results as Radulescu's simplification method. As we worked out in detail in a follow up paper [23], this is *not* the case. The point is that by the computation of elementary modes one does also eliminate "dependent reactions" – that is linear combinations of others – that are generated by the elimination of intermediates.

rrrr Similar results as here were developed in parallel and independently by Saez et al. [24]. They also propose an exact method for the elimination of linear intermediate species in partial steady state, and prove that it preserves the deterministic Ode semantics. Their procedure eliminates several intermediate species in one big step. Clearly, the objective of simultaneous elimination subsumes that of single intermediate elimination, as in the present paper. Compared to Radulescu's method, which also supports simultaneous elimination, it is based on the computation of the spanning trees of the graph of intermediates, instead of using elementary modes. The kinetic functions are supposed to be linear in the intermediates, which is one of the two possible restrictions considered in the present paper. This restriction combined with the spanning tree method enables them to prove a formal correctness result for their simultaneous intermediate elimination. A major difference of Saez reduction method to the present article and Radulescu's approach is that it always produces a *unique* reduced network as a result. We believe that the same network can always be obtained by repeated elimination of intermediate species and dependent reactions, but many different results are possible thereby too. We refer to the confluence study in [23] for further discussions of this topic.

Outline. In Section 2, we illustrate our simplification method with two examples. First we revisit Michaelis-Menten's reduction rule, and then we illustrate how one can simplify a small gene expression network with inhibition. We recall the formal definitions of reaction networks in Section 3, and of conservation laws in Section 4. In Section 5, we contribute a contextual equivalence relation for reaction networks, and in Section 6 a set of simplification rules, that we prove correct with respect to this equivalence relation. In Section 7, we illustrate, with biological examples, how much reaction networks can be simplified in practice. We then discuss further related work in Section 8, and the relevance of our results in Section 9. We conclude and discuss future work in Section 10.

2. Motivating Examples

We first revisit Michaelis-Menten simplification rule for an enzymatic reaction network, and illustrate how the notions of modules and contexts of reaction networks intervene there. We then illustrate how the same ideas can be lifted to more general reaction networks by looking at the example of an inhibited gene expression.

Let $\mathbb{R}_+$ be the set of nonnegative real numbers $t \geq 0$. Kinetic expressions will denote positive real functions that map time points in $\mathbb{R}_+$ to concentrations in $\mathbb{R}_+$. We denote by $\mathbb{C}$ the set of all positive constant functions $\mathbb{C} = \{f : \mathbb{R}_+ \to \mathbb{R}_+ \mid f(0) = f(t) \geq 0 \text{ for all } t \in \mathbb{R}_+\}$.

2.1. Michaelis-Menten revisited

The Michaelis-Menten simplification rule applies to the following enzymatic reaction network with three reac-

tions that we name r_1, r_2, and r_{-1} here:

$$\begin{array}{rrcll} r_1: & S+E & \to & C & ; \; k_1 x_S x_E \\ r_2: & C & \to & P+E & ; \; k_2 x_C \\ r_{-1}: & C & \to & S+E & ; \; k_{-1} x_C. \end{array}$$

Substrate S and enzyme E must first build a complex C via r_1, before C can produce product P and free enzyme E by r_2. Alternatively, complex C can also be redecomposed into S and E by r_{-1}. All three reactions must satisfy the mass action law with different rate constants k_1, k_2, and k_{-1} respectively as as specified by the kinetic expressions. The variables x_S, x_E, and x_C in the kinetic expressions denote some positive real functions, which specify the evolution of the concentrations of the respective species over time.

See Figure 1 for an illustration of this reaction network as a Petri-net. The dotted arrows on S and P mean that these species can be produced or consumed by admissible contexts, while the other species C and E cannot. The Odes system of this network can be derived as usual and are presented in Fig. 2.

The idea of Michaelis-Menten simplification rule is to eliminate the intermediate C from the system, under the assumption that the concentration of C is steady, or in other words, when considering only those solutions of the Odes of the reaction network in which $x_C = c_C$ for some positive constant function $c_C \in \mathbb{C}$. Clearly, the value of the constant function c_C must then be equal to $x_C(0)$, i.e., to the initial concentration of C at time $t = 0$. For other initial concentrations, where this assumption is not satisfied, there is still hope for approximate correctness with some kind of quasi steady state assumptions, but this kind of arguments is out of the scope of the present paper.

Lemma 1. When considering this reaction network in isolation and not as a module of a larger network, any solution of its Odes where C is in partial steady state must satisfy:

- $\frac{dx_C}{dt} = 0$, and
- x_S, x_E, and x_P are constant, and
- if $x_E \neq 0$ then $x_S = 0$.

Proof. This can be seen as follows. If C is in partial steady state then $x_C = c_C$ for some positive constant function $c_C \in \mathbb{C}$. Hence $\frac{dx_C}{dt} = 0$. By observing that the two

$$\begin{aligned} \frac{dx_S}{dt} &= k_{-1} x_C - k_1 x_S x_E \\ \frac{dx_E}{dt} &= (k_2 + k_{-1}) x_C - k_1 x_S x_E \\ \frac{dx_C}{dt} &= -(k_2 + k_{-1}) x_C + k_1 x_S x_E \\ \frac{dx_P}{dt} &= k_2 x_C \end{aligned}$$

Figure 2: Odes system for Michaelis-Menten module with 3 enzymatic reactions without any context.

Odes for derivations of E and C are equal up to the sign, it follows that $\frac{dx_E}{dt} = -\frac{dx_C}{dt} = 0$ and so that $x_E = c_E$ for some positive constant function $c_E \in \mathbb{C}$. From the Ode for the derivation of C we obtain the linear equation:

$$(k_2 + k_{-1}) c_C = k_1 x_S c_E \tag{3}$$

If $c_E = 0$ then the proposition follows straightforwardly. Otherwise, $c_E \neq 0$ so that $x_S = \frac{(k_2+k_{-1})c_C}{k_1 c_E}$ is constant. The differential equation for S can now be reduced to following linear equation:

$$k_{-1} c_C = k_1 x_S c_E \tag{4}$$

In combination, equations (3) and (4) yield $c_C = 0$. From this, equation (4) again implies $x_S = 0$ since $c_E \neq 0$. Furthermore, $c_C = 0$ implies $\frac{dx_P}{dt} = 0$ so that x_P is constant. □

The situation becomes more interesting when the network is considered as a module of a larger network. For instance, we can consider the admissible context that adds one reaction for an inflow of substrate S with constant speed k_i and another reaction that models an outflow of product P with linear speed $k_o x_P$:

$$\begin{array}{lrcll} \text{substrate inflow:} & \emptyset & \to & S & ; \; k_i \\ \text{product outflow:} & P & \to & \emptyset & ; \; k_o x_P. \end{array}$$

Any partial steady state for C of the extended network, unifying the module with the three enzymatic reactions and the context with the two reactions modeling inflows

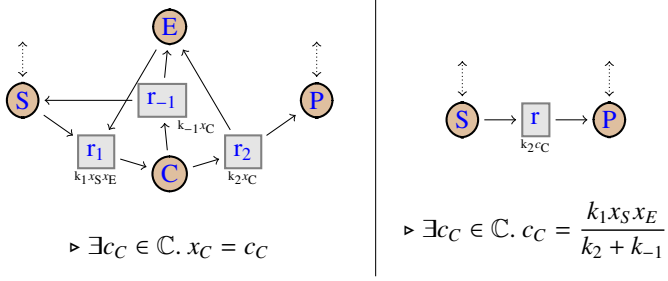


Figure 1: Basic step of Michaelis-Mentens simplification rule under the assumption that C is in partial steady state.

and outflows, must verify $x_C = c_C$ for some positive constant function $c_C \in \mathbb{C}$ satisfying:

$$c_C = \frac{k_1 x_S x_E}{k_2 + k_{-1}}. \tag{5}$$

Therefore, the ODE of the extended network can be simplified to:

$$\frac{dx_P}{dt} = k_2 c_C - k_o x_P$$

while keeping (5) as a conservation law. In analogy, but independently of any context, we can simplify the module with the three enzymatic reactions into the following reaction r:

$$r: \quad S \quad \rightarrow \quad P \quad ; \quad k_2 c_C$$

As we will see, this simplification of the module is correct for all its admissible contexts, when assuming that C is in partial steady state.

So far, the kinetics of the reduced network still look quite different from usual Michaelis-Menten kinetics. But if we assume in addition the conservation law $x_E + x_C = c_E^{total}$ for some constant function $c_E^{total} \in \mathbb{C}$ designing the total enzyme concentration, then it follows by elementary calculations, that:

$$c_C = c_E^{total} \frac{x_S}{x_S + \frac{k_{-1}+k_2}{k_1}}.$$

Hence, we do indeed obtain the classical Michaelis-Menten law by our simplification procedure.

It should be noticed that our assumption that c_C is constant is equivalent to that $\frac{x_S}{x_S + \frac{k_{-1}+k_2}{k_1}}$ is constant, and thus to the assumption that x_S is constant. The latter might look problematic with respect to the usual of Michaelis-Menten law. Since if x_S is not constant for some initial conditions, then this initial concentration of S may be at best close to a value for which C will be in partial steady state. Therefore, the dynamics of the simplified network may be at best an approximation of the original dynamics. Further arguments would be in order to establish such an approximation, but these are out of the scope of the present article.

2.2. Inhibited Gene Expression

In order to illustrate the main ideas of our proposal, we next consider the reaction network *Gene* of inhibited gene expression in Fig. 3.

This reaction network has four species: gene G, inhibitor *Inh*, messenger RNA *mRNA*, and protein P. It also has four reactions: reaction r_1 describes a transcription, the production of *mRNA* in presence of gene G. This reaction has also an inhibitor, *Inh*, that slows down the reaction. G and *Inh* are modifiers in r_1, indicated by a dashed arrow in the figure, meaning that they influence the speed rate of a reaction, but that the reaction does not modify their amounts. Reaction r_2 is the translation of *mRNA* into protein P, while reaction r_3 (resp. r_4) describes the degradation of *mRNA* (resp. P). Except for reaction r_1 with the inhibitor, all other reactions have mass-action kinetics.

Before attempting to simplify the network, we need to specify how the context may interact with it: this is indicated by pending dotted arrows in Fig. 3. We consider here that G and *mRNA* are internal species, that is, which can not be modified by the context. Then, the context can

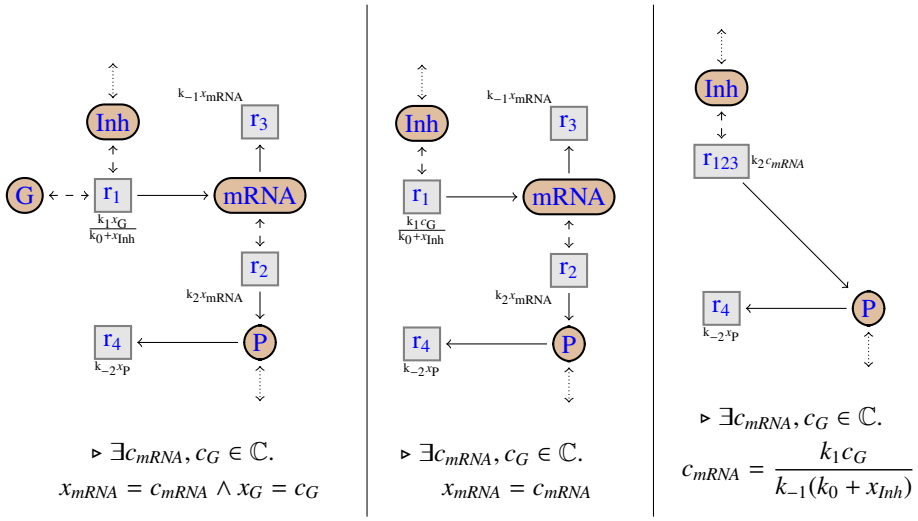


Figure 3: Graph of the reaction network *Gene* on the left, and its two simplifications. Molecular species are represented by circles, and reactions by squares. In the kinetic expressions near the reactions, the k_i are parameters while x_A is a variable representing the concentration of a species A. The variable c_G denotes the initial concentration of G which is assumed to remain constant, by imposing the conservation law $x_G = c_G$ for some positive constant function $c_G \in \mathbb{C}$. A dashed arrow means that the molecule acts as a modifier in the reaction, while a dot arrow means that the molecule can be modified by the context.

be any set of reactions that does not involve G and *mRNA*. It can for instance transform P into another protein, or produce something else in presence of *Inh*, etc.

In this network, we are especially interested in the dynamics of protein P. Therefore, we will say that P is an observable species, that we should not remove. On the other hand, we will assume that the intermediate *mRNA* is at steady state, i.e. that $x_{mRNA} = c_{mRNA}$ for some constant function $c_{mRNA} \in \mathbb{C}$. Hence, we want to eliminate *mRNA* in particular.

In order to simplify the network, we first notice that gene G is not modified by any reaction, and that it can not be modified by any context, since it is assumed internal. Therefore its concentration is constant over time, as described in the conservation law $x_G = c_G$. Hence, we can remove G from the network by replacing x_G by c_G, and removing the conservation law. This results in the simplified network in the middle of Fig. 3.

Now, consider the intermediate *mRNA*. It is an internal species, satisfying the following ODE independently of the context:

$$\frac{dx_{mRNA}}{dt} = \frac{k_1 c_G}{k_0 + x_{Inh}} - k_{-1} x_{mRNA}.$$

Since we assumed that *mRNA* is at steady state with $x_{mRNA} = c_{mRNA}$, we have $\frac{dx_{mRNA}}{dt} = 0$, so that we can deduce the conservation law:

$$c_{mRNA} = \frac{k_1 c_G}{k_{-1}(k_0 + x_{Inh})}.$$

Therefore we can remove *mRNA* from the network, and replace the variable x_{mRNA} by a constant c_{mRNA} in the kinetics of reaction r_2, while adding the above conservation law. We obtain the simplified network on the right of Fig. 3. The three reactions r_1, r_2 and r_3 that produce or consume the intermediate *mRNA* or use it as a modifier are merged into the new reaction r_{123}.

Note that sometimes, the steady state assumptions entail some other implicit conditions. For the Michaelis-Menten simplification rule, the simplification implied that x_S was constant, as discussed above. In this example, the fact that *mRNA* is at steady state entails that the concentration of the inhibitor *Inh* needs to be constant too. This implicit condition is still present in the simplified network, with the conservation law $c_{mRNA} = \frac{k_1 c_G}{k_{-1}(k_0 + x_{Inh})}$.

Therefore, the concentrations of *mRNA*, G and *Inh* need

to be constant to perform the simplification. However, we do not impose a total steady state. For instance, the concentration of the protein of interest P could still change over time. Its dynamics will be the same in both the initial and simplified network.

As we will see in this paper, the simplification rules used above preserve the deterministic semantics of reaction networks, in every context. Hence the simplified network is contextually equivalent to the first one. Note that we can not simplify the network anymore, since P is an observable species, and that both *Inh* and P can be modified by the context.

3. Syntax and semantics of reaction networks

Let *Spec* be a finite set of molecular species ranged over by A. For instance, we used $Spec = \{Inh, mRNA, G, P\}$ for the gene expression network.

We define a *(chemical) solution* $s \in Sol : Spec \to \mathbb{N}_0$ as a multiset of molecular species, i.e. a function from molecular species to natural numbers, with a finite support. Given natural numbers $n_1, \ldots, n_k$, we denote by $n_1A_1 + \ldots + n_kA_k$ the solution that contains n_i molecules of species A_i for all $1 \leq i \leq k$ and 0 molecules of all other species. n_i is called the *stoichiometric coefficient* of A_i in the solution. We define the intersection and difference of two solutions by $(s_1 \cap s_2)(A) = min(s_1(A), s_2(A))$ and $(s_1 - s_2)(A) = s_1(A) - s_2(A)$.

A *reaction* $r = (s_1 \rightarrow s_2;\ e) \in Reactions$ is a pair composed of two solutions s_1 and s_2, and a *kinetic expression* e. The reaction transforms the solution s_1, of so called *reactants*, into the solution s_2, of so called *products*. We denote by $sc_r(A) = s_2(A) - s_1(A) \in \mathbb{Z}$ the stoichiometric coefficient of A in the reaction r, and by $kin(r) = e$ the kinetic expression of r. We consider a countable set of variables *Vars* that contains, for any $A \in Spec$, a variable x_A for a concentration, and a variable c_A, that we will use later to define the steady state assumption. We also have a set of kinetic parameters $Param = \{k_0, k_1, \ldots\}$. Kinetic expressions are terms describing kinetic functions, that have the following abstract syntax:

$$e, f \quad ::= \quad x \mid k \mid \gamma \mid e + f \mid e - f \mid e * f \mid e/f \mid -e$$

where $x \in Vars$, $k \in Param$ and $\gamma \in \mathbb{R}$. As usual, we also simply denote ef for $e * f$, and use parenthesis (e) whenever the priority of the operators might not be clear. We denote by $Vars(e)$ the set of variables of e.

A *normalized reaction* is a reaction in which the reactants and the products do not have shared molecules, i.e. $s_1 \cap s_2 = \emptyset$. Given a reaction $r = (s_1 \rightarrow s_2;\ e)$ and $s = s_1 \cap s_2$, we denote the corresponding normalized reaction by $\widetilde{r} = (s_1 - s \rightarrow s_2 - s;\ e)$. Trivially, for any species A, we have $sc_r(A) = sc_{\widetilde{r}}(A)$. In the following, all reactions are assumed to be normalized.

Given a set of reactions R, we also want to merge the reactions with the same reactants and the same products. We denote by $\widetilde{R}$ the set of reactions obtained by normalizing and then merging the reactions in R, that is:

$$\widetilde{R} = \{(s_1 \rightarrow s_2;\ \sum_{\substack{r \in R,\\ \widetilde{r} = (s_1 \rightarrow s_2;\ e)}} e) \mid \text{ where } s_1, s_2 \in Sol\}.$$

For now, a *(normalized) reaction network* R is a normalized set of reactions. We will later extend this definition with conservation laws.

Now, we define, as usual, the dynamics of reaction networks in terms of ordinary differential equations. Their solutions are the concentrations (depending on time) of the molecular species.

Let $\beta : Param \to \mathbb{R}_+$ be the interpretation of the parameters. We assume here that β is fixed to simplify the notations, but notice that our simplification is correct for any interpretation β.

The *(chemical) concentration* of a chemical species is a function from time to non negative numbers $\mathbb{R}_+ \to \mathbb{R}_+$. Kinetic expressions are interpreted as actual kinetic functions. For any $V \subseteq Vars$, a *V-assignment* $\alpha : V \to (\mathbb{R}_+ \to \mathbb{R}_+)$ maps concentration variables to concentrations.

Given this assignment, the interpretation $[\![e]\!]_\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ of a kinetic expression e is defined by induction on the structure of e as follows, where $t \in \mathbb{R}_+$ and $op \in \{+, -, *, /\}$:

$$[\![x]\!]_\alpha(t) = \alpha(x)(t) \qquad [\![k]\!]_\alpha(t) = \beta(k)$$
$$[\![\gamma]\!]_\alpha(t) = \gamma \qquad [\![-e]\!]_\alpha = -[\![e]\!]_\alpha$$
$$[\![e\ op\ f]\!]_\alpha(t) = [\![e]\!]_\alpha(t)\ op\ [\![f]\!]_\alpha(t)$$

Given a multiset of reactions, we only consider assignments α such that for any kinetic expression e occurring in

this network, its interpretation $[\![e]\!]_\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuously differentiable function from time to non negative real numbers, standing for the actual reaction rate.

Reactions $(s_1 \rightarrow s_2;\ e)$ also have to respect the following *coherence property*: if the concentration of some reactants is zero at some time point then the kinetic rate must be zero too, i.e., for any species A occuring in s_1, any time point $t \geq 0$, and variable assignment α, if $[\![x_A]\!]_\alpha(t) = 0$, then $[\![e]\!]_\alpha(t) = 0$. Note that a kinetic expression may contain variables x_A that are not reactants (neither in the product) of the reaction; such species, called *modifiers*, change the rate of the reaction, but the reaction does not modify their amounts.

An *ordinary differential equation* is an equation of the form $dx/dt = e$, with $x \in \mathit{Vars}$ a variable and e a kinetic expression. A *system of ordinary differential equations* is a conjunction of ordinary differential equations.

For any reaction network R, we can assign it the system of ordinary differential equations, denoted $\textsc{Ode}(R)$, defined as follows:

$$\textsc{Ode}(R) = \bigwedge_{A \in \mathit{Spec}} \frac{dx_A}{dt} = \sum_{r \in R} sc_r(A) kin(r).$$

As expected, a reaction network and its normalized form have the same deterministic interpretation.

Lemma 2. For any R, $\textsc{Ode}(R) = \textsc{Ode}(\widetilde{R})$.

Proof.

$$\begin{aligned}
\textsc{Ode}(\widetilde{R}) &= \bigwedge_{A \in \mathit{Spec}} \frac{dx_A}{dt} = \sum_{(s_1 \rightarrow s_2;\ e) \in \widetilde{R}} sc_r(A) e \\
&= \bigwedge_{A \in \mathit{Spec}} \frac{dx_A}{dt} = \sum_{(s_1 \rightarrow s_2;\ e) \in \widetilde{R}} sc_r(A) \sum_{\substack{r' \in R, \\ \widetilde{r'} = (s_1 \rightarrow s_2;\ e')}} e' \\
&= \bigwedge_{A \in \mathit{Spec}} \frac{dx_A}{dt} = \sum_{r' \in R} sc_{r'}(A) kin(r') \\
&= \textsc{Ode}(R)
\end{aligned}$$

□

A *solution* of an equation $dx/dt = e$ is a V-assignment α such that $[\![dx/dt]\!]_\alpha = [\![e]\!]_\alpha$, where V is the set of variables of the equation, and $[\![dx/dt]\!]_\alpha$ is the first derivative of $[\![x]\!]_\alpha$ with respect to the time t. More generally, for any formula φ, we denote by $\mathit{sol}(\varphi)$ the set of V-assignments that satisfy φ, with V the set of free variables of φ (in the following, some variable for non observable species will be existentially quantified).

4. Conservation law

A *partial steady state assumption* for a species A is a formula of the form $\exists c_A \in \mathbb{C}.\ x_A = c_A$, meaning that x_A is constant, and thus equivalent to the membership formula $x_A \in \mathbb{C}$. We consider consider more general formulas to express conservation laws.

Definition. A *conservation law*, denoted by L, is a formula of the form:

$$\exists \bar{x} \in \mathbb{C}.\ e_1 = e'_1 \wedge \ldots \wedge e_m = e'_m,$$

where $\bar{x} = \{x_1, \ldots, x_n\}$ and the notation $\exists \bar{x} \in \mathbb{C}.\ \varphi$ is a shortcut for $\exists x_1 \in \mathbb{C}. \ldots \exists x_n \in \mathbb{C}.\ \varphi$.

For instance, the formula $\exists c_E^{total} \in \mathbb{C}.\ x_E + x_C = c_E^{total}$ in the initial Michaelis-Menten network in Figure 1 states that the total concentration of the enzyme, free or bound, is constant over time.

As we saw in the preliminary examples, conservation laws need to be rewritten during the simplification. Therefore we add them to our notion of reaction networks.

Definition (Reaction network with conservation law)**.** A *reaction network with conservation law* $N = R \triangleright L$ is a pair of a multiset R of chemical reactions and a conservation law L.

5. Contextual equivalence

We now introduce two other fundamental notions for our equivalence, the notion of *observable species*, and the notion of contexts and *internal species*. We then define the contextual equivalence relation of reaction networks with conservation law.

An *observable species* is a particularly interesting molecule, that should not be removed by the simplification, in contrast to some other molecular species that may only be relevant for some models with "low-level" details. The equivalence will then only preserve the dynamics of

these observable species. For instance, in the *Gene* reaction network depicted in Figure 3, we are especially interested in the dynamics of the protein P, that will therefore be the only observable species. We denote by $O \subseteq Vars$ the set of observable species.

We can then define a first non-contextual equivalence. Two reaction networks with conservation law are equivalent if they have the same deterministic deterministic dynamics for the observable species.

Definition (Non-Contextual Equivalence)**.** Let O be a set of observable species, and $\{\tilde{x}\} = \{x_A \mid A \notin O\}$ the set of variables of unobserved species. Then the reaction networks $N = R \triangleright L$ and $N' = R' \triangleright L'$ are *non-contextually equivalent* if they have the same solutions restricted to the observable species, that is:

$$N \sim_O N' \text{ iff } sol(\exists \tilde{x}.\ \textsc{Ode}(R) \wedge L) = sol(\exists \tilde{x}.\ \textsc{Ode}(R') \wedge L').$$

In the gene expression network, the equivalence only needs to preserve the dynamics of P, and can neglect the others.

We now extend the definition to a contextual equivalence. The notion of *context* allows us to establish equivalences from the equivalence of smaller and isolated subnetworks. This makes proofs of equivalences considerably easier. Indeed, take two networks N_1 and N_2 that differ only by sub-networks, that is $N_1 = N_1' \cup M$ and $N_1 = N_2' \cup M$ for some N_1', N_2' and M. With an equivalence that is preserved by contexts, to prove that N_1 and N_2 are equivalent it is sufficient to prove the equivalence of N_1' and N_2'. To obtain such a desired property, one usually extends an equivalence as a congruence by closing it over all contexts. However, here, this approach would result in a too strong equivalence. In order to make significant simplifications, we will instead close our equivalence $\sim_O$ over *some* contexts said to be *compatible*. Those are defined from a set $I \subseteq Vars$ of molecular species called *internal species*. A context, actually a reaction network N, is *compatible* with I, if $\forall A \in I$, A has no occurrence in N. In other words, a context compatible with I is a reaction network that does not depend and does not act on species in I. We denote by $Context(I)$ the set of compatible contexts with I. In the gene expression network, the internal species are $I = \{G, mRNA\}$. Then the context can interact only with the inhibitor *Inh* and the protein P.

Definition (Equivalence)**.** Let O be a set of observable species and I a set of internal species, the reaction networks with conservation law N_1 and N_2 are (contextually) *equivalent* for O and I, denoted $N_1 \simeq_{O,I} N_2$, iff $\forall M \in Context(I).\ N_1 \cup M \sim_O N_2 \cup M$.

Note that the sets of observable species, internal species and species at steady state are generally not the same and are independent. In the gene network, $O = \{P\}$, while $I = \{mRNA, G\}$. The species *mRNA* and G are also the species at steady state. In the following, we will see that we want to remove *intermediate species*, that are species that are non observable, internal and at steady state, like *mRNA* in the gene network.

6. Simplification

In this section, we present the simplification method. It is composed of a set of simplification axioms that transforms a reaction network into another one, smaller and equivalent.

We first present how the simplification rules work, and how to compare the size of two networks. Then we introduce some simple simplification rules and their properties, followed by a more complex axiom, that removes an intermediate species. Finally, we show how, under some specific conditions, we can use this axiom and the conservation laws to remove a set of linear intermediate species.

6.1. Simplification

A simplification rule transforms a network into another one, under some conditions. The general schema for a rule is depicted in Fig. 4 (left), where a network N is simplified into the network M, under some particular conditions, and when O (resp. I) is the set of observable (resp. internal) species. Note that, after applying a simplification rule, we always directly normalize the simplified network.

If we can simplify a network into another one, then we can do it in any compatible context, as depicted in the axiom (Context) in Fig. 4 (right).

For any reaction network N, we denote $Spec(N)$ the set of species A such that either A or x_A occurs in N (i.e. in a reaction, a kinetics or the conservation law).

We say that a network $N = R \triangleright L$ is *smaller* than a network $M = R' \triangleright L'$ if fewer molecules occur in N than in M, or N has fewer reactions than N, i.e. $N < M$ iff:

$$\frac{\textit{Conditions}}{N \quad \overset{\triangle}{\simeq}_{O,I} \quad M}\ (\textsc{Name}) \qquad \frac{N_0 \in \textit{Context}(I) \qquad N \overset{\triangle}{\simeq}_{O,I} M}{N_0 \cup N \quad \overset{\triangle}{\simeq}_{O,I} \quad N_0 \cup M}\ (\textsc{Context})$$

Figure 4: On the left, simplification rule sketch. The network N, under the *Conditions*, and with the observable species O and the internal species I, can be simplified into M. On the right, the context simplification rule. If a network N can be simplified into a network M, then it is also possible in any context N_0 compatible with the internal species I.

- either $|Spec(N)| < |Spec(M)|$
- or $|Spec(N)| = |Spec(M)|$ and $|R| < |R'|$.

As we will see later, the simplification rules reduce the size of networks.

We define the *simplification relation* $\overset{\triangle}{\simeq}_{O,I}$ for the internal species I and the observable species O as the smallest equivalence relation between networks that contains the simplifications described in Fig. 4 (right), Fig. 5 and Fig. 6. We will show that $\overset{\triangle}{\simeq}_{O,I} \subseteq \simeq_{O,I}$, i.e. a simplified network is contextually equivalent to the initial network.

6.2. Simple simplification rules

The first three simplification rules are given in Fig. 5. The first one, (Useless), deletes a reaction $(\emptyset \to \emptyset;\ e)$ that does not impact the network dynamics. The rule (Modifier) removes an internal species A only used as a modifier in the reactions (i.e. it appears in the kinetics, but not in any reactants or products). It uses a (simple) conservation law on A to compute and remove its concentration from the kinetic expression. In the simplified network, the notation $e[x_A/e_A]$ stands for the substitution of x_A by e_A in the expression e, i.e. any occurrence of x_A is replaced by e_A. This axiom is for instance used on the gene G in the gene expression network of Section 2.2.

These rules are sound for the equivalence, and reduce the size of the network.

Proposition 1. The rules (Useless) and (Modifier) reduce the size of the network.

Proof. Trivially, (Useless) reduces the number of reactions while (Modifier) reduces the number of involved species. □

Theorem 1. The rules (Useless) and (Modifier) are sound for the contextual equivalence, i.e. if we can simplify a network N into M by applying one of these rules, then $N \simeq_{O,I} M$.

Proof. The axiom (Useless) removes a reaction that can not modify the concentrations of the molecules. So the Odes systems of both networks are the same, their solutions are equal, and the networks are equivalent.

For (Modifier), A is an internal species, and is never used as reactant or product of any reaction. Therefore its concentration is constant, and can be computed by the conservation law $x_A = e_A$ in the initial network. Consequently, if we have a solution α of the initial network (in some context N_0), that satisfies the conservation laws, then the Odes systems for both networks (in the context N_0) are the same, and α is also a solution for the simplified network (the additional conservation law $\exists c_A \in \mathbb{C}.\ c_A = e_A$ is directly satisfied, since in the initial network we have $e_A = x_A$). Reciprocally, if α is solution for the simplified network, we build the solution α' that is equal to α, except for the species A, where $\alpha'_c(A) = e_A$. Since e_A is constant (by the conservation law) and x_A is too (since it is internal and not involved in any reaction), α' will still be a solution of the simplified network that satisfies the conservation laws. Then once again the Odes are the same, and α' is a solution for the initial network too. And since α and α' are equal modulo the intermediate species, then the networks are contextually equivalent. □

6.3. Intermediate

We now present the (Intermediate) axiom, depicted in Fig. 6, that aims at eliminating an internal, non-observable species at steady state, by merging two-by-two the reactions that produce and consume it.

Let A be the intermediate (i.e. internal, non-observable and at steady state) species that we want to remove. In order to be able to compute systematically the kinetic rates of the simplified networks, all reactions that consume A need to have the same kind of dependency on x_A. So there needs to be an expression f (with $x_A \in Vars(f)$), such

$$\frac{}{\{(\emptyset \to \emptyset;\ e)\} \triangleright L \quad \overset{\triangle}{\simeq}_{O,I} \quad \emptyset \triangleright L} \text{ (Useless)}$$

$$\frac{\forall r \in R, sc_r(A) = 0 \qquad A \in I \qquad x_A \notin Vars(e_A)}{R \triangleright (\exists \bar{x} \in \mathbb{C}.\ \varphi \wedge x_A = e_A) \quad \overset{\triangle}{\simeq}_{O,I} \quad R[x_A/e_A] \triangleright (\exists \bar{x}, c_A \in \mathbb{C}.\ \varphi[x_A/e_A] \wedge c_A = e_A)} \text{ (Modifier)}$$

Figure 5: Simple simplification axioms. (Useless) removes a useless reaction, while (Modifier) removes an internal species, only used as modifier and present in a simple conservation law.

$$\frac{\begin{array}{c} N = R_N \triangleright \exists \bar{x} \in \mathbb{C}.\ \varphi \wedge x_A = c_A \\ A \in I \qquad A \notin O \qquad c_A \in \bar{x} \\ split(R_N, \varphi, A, f) = (R, R', R_{mod1}, R_{mod2}, \varphi_1, \varphi_2) \\ f \text{ is a term with } x_A \in Vars(f) \\ f^{-1} \text{ is an inverse term of } f \text{ for } x_A \text{ or } (R_{mod2} = \emptyset \text{ and } \varphi_2 = \top) \\ T = \sum_{(s_1 \to s_2 + nA;\ e) \in R} ne \qquad T' = \sum_{(s'_1 + n'A \to s_2;\ fe') \in R'} n'e' \qquad F = T/T' \end{array}}{\begin{array}{l} N \quad \overset{\triangle}{\simeq}_{O,I} \quad \{(n's_1 + ns'_1 \to n's_2 + ns'_2;\ \frac{ee'}{T'}) \mid \\ \qquad (s_1 \to s_2 + nA;\ e) \in R, \quad (s'_1 + n'A \to s'_2;\ fe') \in R'\} \cup \\ \{(s_1 \to s_2;\ e[x/F]) \mid (s_1 \to s_2;\ e[x/f]) \in R_{mod1}\} \cup \\ \{(s_1 \to s_2;\ e[x_A/f^{-1}[x_A/F]]) \mid (s_1 \to s_2;\ e) \in R_{mod2}\} \\ \triangleright \exists \bar{x} \in \mathbb{C}.\ \ \varphi_1[f/F] \ \wedge\ \varphi_2[x_A/f^{-1}[x_A/F]] \ \wedge\ f[x_A/c_A] = F \end{array}} \text{ (Intermediate)}$$

Figure 6: General intermediate rule that removes the intermediate species A from reaction network N.

that the kinetic expression of any reaction with A in the reactants is of the form fe', with $x_A \notin Vars(e')$.

Given an expression f with $x \in Vars(f)$, we say that f is *invertible* in x if there exists an expression, called the *inverse* of f for x and denoted f^{-1}, such that:

$$\forall \alpha.\ [\![f[x/f^{-1}]]\!]_\alpha = [\![x]\!]_\alpha.$$

Then, we will partition the reactions involving A into the ones that produce it, the ones that consume it and the ones that use A as modifier, that are themselves divided into the ones with a kinetic expression that depends on f, and the others. The same kind of partition is also made for the conservation laws. Formally, given a network of the form $N = R_N \triangleright \exists \bar{x} \in \mathbb{C}.\ \varphi \wedge x_A = c_A$, with $c_A \in \bar{x}$, an intermediate species A and an expression f such that $x_A \in Vars(f)$, we define the partition $split(R_N, \varphi, A, f) = (R, R', R_{mod1}, R_{mod2}, \varphi_1, \varphi_2)$ such that $R_N = R \cup R' \cup R_{mod1} \cup R_{mod2}$, $\varphi = \varphi_1 \wedge \varphi_2$, and the following conditions are satisfied:

- The (non-empty) set R contains the reactions that produce A. The kinetics of these reactions should not depend on the concentration of A. Formally, a production reaction is a reaction of the form $r = (s_1 \to s_2 + nA;\ e)$, with $A \notin s_1, s_2$, any stoichiometric coefficient $n > 0$, and $x_A \notin Vars(e)$.
- R' is the (non-empty) set of reactions that consume A. The kinetics of these reactions depend on the concentration of A, but always according to the expression f. They do not have to be necessarily linear in x_A. Therefore these reactions are of the form $r' = (s'_1 + n'A \to s'_2;\ fe')$, with $A \notin s'_1, s'_2$, some coefficient $n' > 0$, and such that $x_A \notin Vars(e')$.
- We divide the reactions with A as modifier into two (potentially empty) sets. The first set, R_{mod1}, regroups the reactions with kinetic expressions dependent on f. I.e. any reaction $r \in R_{mod1}$ has some expression e with a variable x such that $r = (s_1 \to s_2;\ e[x/f])$. We have $x_A \notin Vars(e)$, and $A \notin s_1, s_2$ since it is a modifier.
- The second set R_{mod2} contains all the other reactions,

without any condition. However, in order to do the simplification, we need an additional condition: either f is invertible in x_A, or the set R_{mod2} is empty.

- Similarly to the reactions with A as modifier, we partition the conservation laws into φ_1, the ones that depend on f, and φ_2, the others. Formally, $\varphi_1 = \bigwedge_i (e_i[x/f] = e'_i[x/f])$ with $x_A \notin Vars(e_i), Vars(e'_i)$. As before, if f is not invertible in x_A, then there is no φ_2.

Now that we have defined the partitions, we can describe the simplification of the network. This consists in combining two-by-two the production reactions from R with the consumption reactions from R'. First, we define two expressions, representing the sums of the weighted kinetics of the production (resp. consumption) reactions by:

$$T = \sum_{(s_1 \to s_2 + nA;\ e) \in R} ne,$$

$$T' = \sum_{(s'_1 + n'A \to s_2;\ fe') \in R'} n'e'$$

$$F = \frac{T}{T'}.$$

The conditions on the kinetic expressions of the consumption reactions, and the fact that A is an intermediate species, and at steady state, imply that we can easily compute the value: $[\![f]\!]_\alpha = [\![F]\!]_\alpha$.

Then, the combination of a production and a consumption reaction will be the normalization of the (general) reaction:

$$(n's_1 + ns'_1 \to n's_2 + ns'_2;\ \frac{ee'}{T'}).$$

For the reactions with A as modifier, and for the conservation laws, we just replace f by F, and, when f is invertible, x_A by $f^{-1}[x_A/F]$, where f^{-1} is the inverse of f for x_A.

We also replace the conservation law $\exists c_A \in \mathbb{C}.\ x_A = c_A$ by $\exists c_A \in \mathbb{C}.\ f[x_A/c_A] = F$. This is necessary to keep track of the fact that x_A is constant, and its consequences.

Consequently, we obtain a new simplified network, not involving the species A, and composed of the combining reactions, the modified reactions with A as modifier, the modified conservation laws, and the new additional conservation law.

This axiom can simplified an intermediate species with a non-linear kinetic rate, and stoichiometric coefficients grater than one. Note that if the kinetic rates are linear in the concentration of the intermediates, then the function f is the identity, and so if trivially invertible. The two following examples present networks with non-linearity, where we can still do the simplification.

Example 1 (Quadratic kinetic rate)**.** Consider the following reaction network, with two reactions:

$$r_1 = (A \to X;\ k_1 x_A) \quad r_2 = (2X \to B;\ k_2 x_X^2).$$

We want to simplify the intermediate species X. The reaction r_2 has a non-linear kinetic rates $k_2 x_X^2$. It is not invertible, but since X is not used as modifier, we can still do the simplification, and obtain the following reaction:

$$r'_1 = (2A \to B;\ \frac{k_1 x_A}{2}).$$

Example 2 (Michaelis-Menten kinetic rate)**.** Consider now the following reaction network, with three reactions:

$$r_1 = (A \to X;\ k_1 x_A) \quad r_2 = (B \to C;\ k_2 x_B x_X)$$

$$r_3 = (X \to D;\ V\frac{x_X}{K + x_X}).$$

The only reaction using the intermediate X as reactant is r_3, and has a non-linear kinetic expression $f = V\frac{x_X}{K + x_X}$. We assume here that $K > 0$ and $V > 0$. The function f is invertible, with $f^{-1} = \frac{Kx_X}{V - x_X}$, so we can apply the simplification and obtain the reduced network:

$$r'_1 = (A \to D;\ k_1 x_A) \quad r'_2 = (B \to C;\ k_1 k_2 K \frac{x_A x_B}{V - k_1 x_A}).$$

Note that in the simplified network, we always have $V - k_1 x_A \neq 0$. Otherwise, we would have $V = k_1 x_A$. But the steady state assumption on X implies $k_1 x_A - V\frac{x_X}{K + x_X} = 0$, and therefore $V = V\frac{x_X}{K + x_X}$. Since $V \neq 0$, this would mean that $K + x_X = x_X$, i.e. $K = 0$, that contradicts our hypothesis.

Proposition 2. The axiom (Intermediate) reduces the size of the network.

Proof. It removes a molecular species. □

Theorem 2. The axiom (Intermediate) is sound for the contextual partially steady state equivalence, i.e. if we can simplify a network N into M by applying the axiom, then $N \simeq_{O,I} M$.

Proof. Let N_1 be the initial network, and N_2 be the simplified network obtained by applying the axiom (Intermediate) on the intermediate species A. We have to prove that $N_1 \simeq_{O,I} N_2$.

Let M be a context compatible with I, i.e. for any species $B \in I$, B has no occurrence in M. As in the axiom, we partition N_1 into $(R, R', R_{mod1}, R_{mod2}, \varphi_1, \varphi_2)$. We will only consider here the case where the function f is invertible in x_A, with f^{-1} as inverse. The other case, without R_{mod2} and φ_2, is similar.

The proof follows the following steps. First, we will show that, if we assume that $x_A = f^{-1}[x_A/F]$, then both networks $N_1 \cup M$ and $N_2 \cup M$ generate the same Odes system. Then we will prove that any solution for the network $N_1 \cup M$ also satisfies the above hypothesis, as well as the conservation law of $N_2 \cup M$. Reciprocally, for any solution α of $N_2 \cup M$, we will find an equivalent solution (modulo O) that satisfies both the above hypothesis and that is solution of $N_1 \cup M$.

We start by assuming the following hypothesis, with f^{-1}, T, T' and F as defined in the axiom:

$$x_A = f^{-1}[x_A/F]. \tag{6}$$

Consider the Odes system of $N_1 \cup M$. Since A is at steady state, any solution will verify $\frac{dx_A}{dt} = 0$. For any other species $X \neq A$, we denote by $F_X(M)$ the part of the differential equation on X corresponding to the reactions of the context M. Therefore, the differential equations are:

$$\begin{aligned}\frac{dx_X}{dt} = &\sum_{(s_1 \to s_2+nA;\ e)\in R} (s_2(X) - s_1(X))e \\ &+ \sum_{(s'_1+n'A \to s'_2;\ fe')\in R'} (s'_2(X) - s'_1(X))fe' \\ &+ \sum_{(s_1 \to s_2;\ e[x/f])\in R_{mod1}} (s_2(X) - s_1(X))e[x/f] \\ &+ \sum_{(s_1 \to s_2;\ e)\in R_{mod2}} (s_2(X) - s_1(X))e \\ &+ F_X(M).\end{aligned}$$

We can replace f and x_A by their values, according to the hypothesis (6), and obtain the equation:

$$\begin{aligned}\frac{dx_X}{dt} = &\sum_{(s_1 \to s_2+nA;\ e)\in R} (s_2(X) - s_1(X))e \\ &+ \sum_{(s'_1+n'A \to s'_2;\ fe')\in R'} (s'_2(X) - s'_1(X))Fe' \\ &+ \sum_{(s_1 \to s_2;\ e[x/f])\in R_{mod1}} (s_2(X) - s_1(X))e[x/F] \\ &+ \sum_{(s_1 \to s_2;\ e)\in R_{mod2}} (s_2(X) - s_1(X))e[x_A/f^{-1}[x_A/F]] \\ &+ F_X(M).\end{aligned} \tag{7}$$

Let us consider now the Odes system for $N_2 \cup M$. Since we totally remove the species A, we have $\frac{dx_A}{dt} = 0$. The context M is the same in both networks, so we have the same function $F_X(M)$. Then, according to the description of the axiom, the differential equation for $X \neq A$ is the following:

$$\begin{aligned}\frac{dx_X}{dt} = &\sum_{\substack{(s_1 \to s_2 + nA;\ e) \in R,\\ (s'_1 + n'A \to s'_2;\ fe') \in R'}} (n's_2(X) + ns'_2(X) - n's_1(X) - ns'_1(X))\frac{ee'}{T'} \\ &+ \sum_{(s_1 \to s_2;\ e[x/f])\in R_{mod1}} (s_2(X) - s_1(X))e[x/F] \\ &+ \sum_{(s_1 \to s_2;\ e)\in R_{mod2}} (s_2(X) - s_1(X))e[x_A/f^{-1}[x_A/F]] \\ &+ F_X(M).\end{aligned} \tag{8}$$

We notice that the three last lines of equations (7) and (8) are the same. Moreover, we can divide the first sum of

(8) into two parts, and obtain the remaining sums in (7):

$$\sum_{\substack{(s_1 \to s_2 + nA;\ e) \in R,\\ (s_1' + n'A \to s_2';\ fe') \in R'}} (n's_2(X) + ns_2'(X) - n's_1(X) - ns_1'(X))\frac{ee'}{T'}$$

$$= \sum_{(s_1 \to s_2+nA;\ e)\in R} \ \sum_{(s_1'+n'A \to s_2';\ fe')\in R'} (n's_2(X) - n's_1(X))\frac{ee'}{T'}$$

$$+ \sum_{(s_1'+n'A \to s_2';\ fe')\in R'} \ \sum_{(s_1 \to s_2+nA;\ e)\in R} (ns_2'(X) - ns_1'(X))\frac{ee'}{T'}$$

$$= \sum_{(s_1 \to s_2+nA;\ e)\in R} (s_2(X) - s_1(X))\frac{eT'}{T'}$$

$$+ \sum_{(s_1'+n'A \to s_2';\ fe')\in R'} (s_2'(X) - s_1'(X))\frac{Te'}{T'}$$

$$= \sum_{(s_1 \to s_2+nA;\ e)\in R} (s_2(X) - s_1(X))e$$

$$+ \sum_{(s_1'+n'A \to s_2';\ fe')\in R'} (s_2'(X) - s_1'(X))Fe'.$$

Therefore, if the hypothesis is satisfied, the two networks have the same Odes systems, and the same solutions.

Let α be a solution of for the network $N_1 \cup M$, that satisfies its conservation laws. Since $A \in I$, we know that it does not appear in M. Then we can fully express the differential equation for its concentration in $N_1 \cup M$ by:

$$\frac{dx_A}{dt} = \sum_{(s_1 \to s_2+nA;\ e)\in R} ne \ - \sum_{(s_1'+n'A \to s_2';\ fe')\in R'} n'fe'.$$

The conservation law $\exists c_A \in \mathbb{C}.\ x_A = c_A$ implies $\frac{dx_A}{dt} = 0$. We can therefore rewrite the equation above, using the value T and T' defined in the axiom, as:

$$f = \frac{\displaystyle\sum_{(s_1 \to s_2+nA;\ e)\in R} ne}{\displaystyle\sum_{(s_1'+n'A \to s_2';\ fe')\in R'} n'e'} = \frac{T}{T'} = F.$$

Since f is invertible, we can compute the value of the concentration of A, that is:

$$x_A = f^{-1}[x_A/F]$$

Consequently, if α is a solution for $N_1 \cup M$, then the hypothesis is satisfied, and the Odes systems are the same. We need to verify that α also satisfies the conservation laws of the network $N_2 \cup M$. This can be found simply by substituting x_A with $f^{-1}[x_A/F]$ in the initial conservation laws. So α is also a solution of $N_1 \cup M$ that satisfies its conservation laws.

Now let α be a solution for $N_2 \cup M$. We define α' such that $\alpha_0'(c_A) = \alpha_c'(x_A) = f^{-1}[x_A/F]$, and $\alpha'(X) = \alpha(X)$ otherwise. Since A is a non-observable species, $\alpha' =_O \alpha$. We will show that α' is a solution of both $N_2 \cup M$ and $N_1 \cup M$ and satisfies their conservation laws. α satisfies $\exists c_A \in \mathbb{C}.\ f[x_A/c_A] = F$, therefore $f^{-1}[x_A/F]$ is constant over time. Since x_A does not appear in F, it is also constant with α'. So the hypothesis (6) is satisfied by α'. Consequently, the Odes systems for both networks are the same. Since α and α' are equal except for the value of x_A, and since in $N_2 \cup M$ we have $\frac{dx_A}{dt} = 0$ and x_A does not appear in the other equations for $N_2 \cup M$, the fact that α is solution of $N_2 \cup M$ implies that α' is too. Then it is also solution for $N_1 \cup M$. And the conservation laws are directly satisfied.

Hence the two networks have the same solutions, modulo the intermediate species. Therefore they are contextually equivalent. □

6.4. *Intermediate with conservation law*

In this section, we study in more details the scenario with a set of intermediate molecules that only appear with stoichiometry one, with linear kinetics, and satisfy a particular conservation law. We will see that we can iteratively apply the axiom (Intermediate) on those species, until there is only one, and then apply the axiom (Modifier) on this last species.

We first define the conditions on these species. Let $\mathcal{U}$ be a set of intermediate species. A reaction $r = (s_1 \to s_2;\ e)$ is *$\mathcal{U}$-linear* if:

- either r does not involve the species of $\mathcal{U}$ (except potentially in the kinetic expression), i.e. $\forall A \in \mathcal{U}$, $A \notin s_1$ and $A \notin s_2$;
- or it transforms exactly one species of $\mathcal{U}$ into another one, with a linear kinetic expression, i.e. $\exists A_1 \neq A_2 \in \mathcal{U}.\ r = (A_1 + s_1' \to A_2 + s_2';\ x_{A_1}e')$, with for any $B \in \mathcal{U}$

(including A_1 and A_2), $B \notin s'_1$, $B \notin s'_2$ and $x_B \notin Vars(e')$.

Then, $\mathcal{U}$ is a *linear intermediate set* for a network $N = R \triangleright L$, for the internal species I and for the observable species O if:

- these species are internal and non-observable, i.e. $\mathcal{U} \subseteq I$ and $\mathcal{U} \cap O = \emptyset$;
- they are at steady state, and there is a conservation law linear in $\mathcal{U}$:

$$L = \quad \exists \bar{x} \in \mathbb{C}.\ \varphi \wedge \left(\sum_{A \in \mathcal{U}} x_A e_A = e \right) \bigwedge_{A \in \mathcal{U}} \exists c_A \in \mathbb{C}.\ x_A = c_A$$

  with for any $A \in \mathcal{U}$, $x_A \notin Vars(e_A)$, $x_A \notin Vars(e)$, $e_A \neq 0$, and for any other species $B \in \mathcal{U}$, $x_B \notin Vars(e_A)$;
- any reaction $r \in R$ is $\mathcal{U}$-linear.

Note that a linear conservation law of the form described in the second condition of the definition is always satisfied, as a consequence of the structure of the reactions described in the third condition. In the Michaelis-Menten simplification in Section 2.1, the set $\{C, E\}$ is a linear intermediate set.

We will first show that we can apply the axiom (Intermediate) on a linear set, and obtain another linear set.

Lemma 3. If $\mathcal{U}$ is a linear intermediate set for N, with $|\mathcal{U}| > 1$, then we can use the axiom (Intermediate) on any $A \in \mathcal{U}$.

Proof. A is an intermediate species. The reactions that consume A are all linear in x_A, i.e. have the same function $f(x_A) = x_A = f^{-1}(x_A)$. So we can apply the axiom and remove A. □

Lemma 4. If $\mathcal{U}$ is a linear intermediate set for N, with $|\mathcal{U}| > 1$ and $A \in l$, and if we use the axiom (Intermediate) on A and obtain the simplified network M, then $\mathcal{U}\backslash A$ is a linear intermediate set for M.

Proof. First, we trivially have that $(\mathcal{U}\backslash A) \subseteq I$ and $\mathcal{U}\backslash A \cap O = \emptyset$.

Let us consider the structure of the reactions. Since the stoichiometric coefficient is always 1, and the species of $\mathcal{U}$ never appear in the same reactants or products of a reaction, the production (resp. consumption) reactions will be of form $(s_1 + A_i \rightarrow s_2 + A;\ x_{A_i} e)$ (resp. $(s'_1 + A \rightarrow s'_2 + A'_i;\ x_A e')$). Therefore, the combined reactions will be of form $r = (s_1 + A_i \rightarrow s'_2 + A'_i;\ \frac{x_{A_i} e e'}{T'})$, with $T' = \sum_{(s'_1 + A \rightarrow s'_2 + A'_i;\ x_A e') \in R'} e'$. So this reaction transforms a species of $\mathcal{U}\backslash A$ into another one. And since e and e' (and so T' too) do not depend on the concentrations of the species of $\mathcal{U}$, r has a linear kinetics.

For the other reactions, with A as modifier, only their kinetic expressions are modified. Since x_A appears in the kinetics of these reactions, then no species of $\mathcal{U}$ can appear as reactant or product, and therefore the simplified reactions still verify the conditions of the linear intermediate set.

Consider the linear conservation law $\sum_{A \in \mathcal{U}} x_A e_A = e$ from the initial network. It will be rewritten into $\frac{T}{T'} e_A + \sum_{A_i \in \mathcal{U}\backslash A} x_{A_i} e_{A_i} = e$, with $T = \sum_{(s_1 + A_i \rightarrow s_2 + A;\ x_{A_i} e) \in R} x_{A_i} e$, and T' as defined above. The expressions e, e_A and e_{A_i} above do not depend on any concentrations of $\mathcal{U}$. As a consequence T, and consequently the new conservation law, are linear in the species of $\mathcal{U}\backslash A$.

In conclusion, $\mathcal{U}\backslash A$ is a linear intermediate set for the simplified network. □

So we can apply iteratively the axiom (Intermediate) on the species of the linear intermediate set, until we have only one species. Then we have a simple linear conservation law on the last remaining species, and we can apply the rule (Modifier).

Lemma 5. If $\mathcal{U}$ is a linear intermediate set for N, with $\mathcal{U} = \{A\}$, then we can use the axiom (Modifier) on A.

Proof. A is an intermediate species, and its concentration appears in a linear conservation law of the form $x_A e = k$, equivalent to $x_A = \frac{e}{k}$. By definition of linear intermediate set, A cannot appear in the reactants or the products of any reaction (since a reaction needs to transform A into another species of the set, and there is no other species).

So we can apply the axiom, using the linear conservation law. □

Theorem 3. If $\mathcal{U}$ is a linear intermediate set for N, then we can remove all the species of $\mathcal{U}$ from the network, by iteratively applying (Intermediate), followed by (Modifier) on the last species.

7. Examples

In this section, we present three examples of network simplification. First, we use the axiom (Intermediate) to reduce a network with non-linear kinetics. Then we show a bigger example of simplification, on the reaction network of the *Tet-On* system [25, 26, 27]. We finally consider the simplification of a NF-κB network [20].

7.1. Non linear intermediate species

We start by considering the reaction network depicted in Fig. 7 (left), that can represent for instance the double enzymatic-phosphorilation of a protein P. The species P can bind to a first enzyme E and form the complex C. Then C can either dissociate back, or produce P' while releasing E. The protein P' can now bind to another enzyme E', to create the complex C'. Once again, C' can dissociate back to P' and E', or can release E' and produce the final species P''.

The protein P is an input, modifiable by the context, while P'' is the output. The other species are all internal, non-observable, and we assume them at steady state. So we have $I = O = \{E, C, P', E', C'\}$. In addition to the partial steady state, there are two conservation laws, one for each enzyme: $x_E + x_C = c_E^{total}$ and $x_{E'} + x_{C'} = c_{E'}^{total}$. The network is formally described in Fig. 8.

We can first apply (Intermediate) on the complexes C and C', followed by (Modifier), using the conservation laws on E and E', as in the Michaelis-Menten example described in Section 2. We obtain the network depicted in the middle of Fig. 7. Note that by convenience, we also simplify the conservation laws, by removing redundant and useless ones. The network is formally described in Fig. 9

We then want to remove the intermediate P'. It is involved in the reaction (12), with a non-linear kinetic expression. However, since this is the only reaction with P' as reactant, we can still apply the axiom (Intermediate). We obtain the equivalent network on the right of Fig. 7, with the only reaction:

$$P \to P'' \; ; \; \frac{V x_P}{x_P + K}$$

and the conservation law $\exists c_P \in \mathbb{C}.\ c_P = x_P$.

7.2. The Tet-On system

We present here the simplification of the *Tet-On* system [25, 26, 27] using our axioms. The initial *Tet-On* reaction network, depicted in Fig. 10 (left), has 10 reactions and 11 parameters. We simplify it into the equivalent $\textit{Tet-On}_{simple}$ network, depicted on Fig. 10 (right), with only 2 reactions and 3 parameters.

The *Tet-On* system [25, 26, 27] 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 $\textit{Tet-On} = R \triangleright L$ where R is the set of reactions from Fig. 11, inspired by the *Tet-On* model from [27], and with the following unique conservation laws, that preserves the total amount of *rtTA* (bound or free), as well as the concentration of the gene P_{TRE3G}, and where φ represents the partial steady state described below:

$$\begin{aligned} L = \quad & \exists c_{rtTA}^{total}, c_{P_{TRE3G}}^{total} \in \mathbb{C}.\ x_{rtTA} + x_{rtTADox} = c_{rtTA}^{total} \\ & \wedge x_{P_{TRE3G}} = c_{P_{TRE3G}}^{total} \wedge \varphi. \end{aligned}$$

In the network, the doxycycline *Dox* moves into the cell and becomes Dox_i by reaction (13). We assume here that the amount of *Dox* is controlled by the environment (for instance by a microfluidics device [28]), and therefore the network can not modify its concentration. Then Dox_i is either degraded by reaction (14), or binds to the artificial transcription factor *rtTA* by reaction (15). The complex *rtTADox* either dissociates (16), or activates the transcription of the gene P_{TRE3G}, producing *mRNA* (17). *mRNA* either degrades (19) or is translated into *GFP* (20). Finally, *GFP* needs to be activated into GFP_a (22) in order to become fluorescent and thus observable by a microscope. Both *GFP* and GFP_a can also be degraded (21, 23).

We are particularly interested in GFP_a, since it is the only experimentally observable species. Therefore we assume that all other species are at steady state, and

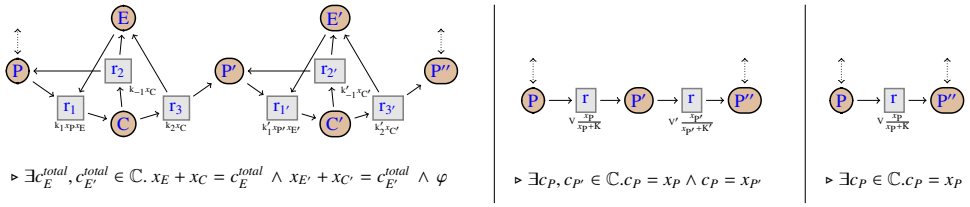


Figure 7: Reaction graphs of the detailed (left), intermediate (middle) and simplified (right) non-linear networks. P and P'' can be modified by the context. In the left network, φ is the conservation law representing the partial steady state. In the middle and right networks, the parameters are $V = k_2 c_E^{total}$, $V' = k_2' c_{E'}^{total}$, $K = \frac{k_{-1}+k_2}{k_1}$ and $K' = \frac{k_{-1}'+k_2'}{k_1'}$.

$$P + E \to C \;;\; k_1 x_P x_E \qquad P' + E' \to C' \;;\; k_1' x_{P'} x_{E'}$$
$$C \to P + E \;;\; k_{-1} x_C \qquad C' \to P' + E' \;;\; k_{-1}' x_{C'}$$
$$C \to P' + E \;;\; k_2 x_C \qquad C' \to P'' + E' \;;\; k_2' x_{C'}$$

$$\triangleright\ \exists c_E, c_C, c_{P'}, c_{E'}, c_{C'} \in \mathbb{C}.\ x_E = c_E \wedge x_C = c_C \wedge x_{P'} = c_{P'} \wedge x_{E'} = c_{E'} \wedge x_{C'} = c_{C'}$$
$$\wedge\ \exists c_E^{total}, c_{E'}^{total} \in \mathbb{C}.\ x_E + x_C = c_E^{total} \wedge x_{E'} + x_{C'} = c_{E'}^{total}$$

Figure 8: The detailed non-linear network.

non-observable. Therefore, $O = \{GFP_a\}$ and $\varphi = \bigwedge_{A \in Spec \setminus GFP_a} \exists c_A \in \mathbb{C}.\ x_A = c_A$. Since we assume that *Dox* is controlled by the environment, while any other species is inside the cell, it will be the only species modifiable by the context. Therefore all species except *Dox* are internal: $I = Spec \setminus Dox$. The simplification follows the axioms from Fig. 5, and Fig. 6, and as a consequence the two networks are equivalent. Note that in the following simplification, for the sake of readability, some kinetic expressions and conservation laws are sometimes slightly rewritten into equivalent expressions.

One can start by observing that the gene P_{TRE3G} is only used as a modifier, in reaction (17). We apply the axiom (Modifier), removing P_{TRE3G} from this reaction, while replacing $x_{P_{TRE3G}}$ by $c_{P_{TRE3G}}^{total}$ in its kinetic function. The conservation law on $x_{P_{TRE3G}}$ becomes $\exists c_{P_{TRE3G}} \in \mathbb{C}.\ c_{P_{TRE3G}} = c_{P_{TRE3G}}^{total}$, and can directly be removed.

Then consider the internal, non-observable species *rtTADox*. It is at steady state, and involved in three reactions: one that produces it (15), one that consumes it (16), and one that uses it as a modifier (17). Then we use the axiom (Intermediate) on it, and compute its value at steady state: $x_{rtTADox} = \frac{k_2}{k_{-2}} x_{rtTA} x_{Dox_i}$. The three reactions are merged into:

$$\emptyset \to \emptyset \quad ; \quad k_2 x_{rtTA} x_{Dox_i} \tag{24}$$
$$\emptyset \to mRNA \quad ; \quad c_{P_{TRE3G}}^{total} V_1 \frac{x_{rtTA} x_{Dox_i}}{x_{rtTA} x_{Dox_i} + k_{-2} K_1 / k_2} \tag{25}$$

and the conservation law becomes:

$$\exists c_{rtTADox} \in \mathbb{C}.\ c_{rtTADox} = \frac{k_2}{k_{-2}} x_{rtTA} x_{Dox_i} \tag{26}$$
$$\wedge \exists c_{rtTA}^{total} \in \mathbb{C}.\ x_{rtTA} + \frac{k_2}{k_{-2}} x_{rtTA} x_{Dox_i} = c_{rtTA}^{total} \tag{27}$$
$$\bigwedge_{A \in \{Dox, Dox_i, rtTA, mRNA, GFP\}} \exists c_A.\ x_A = c_A. \tag{28}$$

We can delete the reaction (24) with the axiom (Useless). *rtTA* is now only used as modifier. The conservation law (27) can be rewritten into $\exists c_{rtTA}^{total} \in \mathbb{C}.\ x_{rtTA} =$

$$P \to P' \; ; \; \frac{V x_P}{x_P + K} \qquad (10) \qquad\qquad P' \to P'' \; ; \; \frac{V' x_{P'}}{x_{P'} + K'} \qquad (12)$$

$$\triangleright \; \exists c_P, c_{P'} \in \mathbb{C}.\; c_P = x_P \wedge x_{P'} = c_{P'}$$

Figure 9: Intermediate non-linear network, with $V = k_2 c_E^{total}$, $V' = k'_2 c_{E'}^{total}$, $K = \frac{k_{-1} + k_2}{k_1}$ and $K' = \frac{k'_{-1} + k'_2}{k'_1}$.

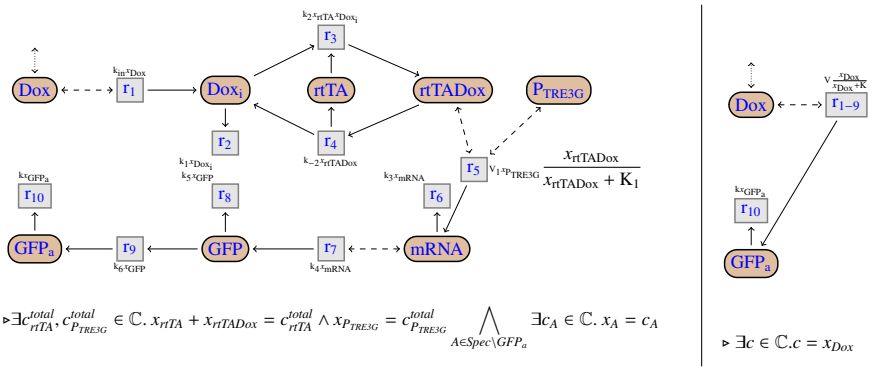


Figure 10: Reaction graphs of the detailed (left) and simplified (right) *Tet-On* networks. Species are represented by circles, and reactions by squares. A dash arrow means that the species acts as a modifier in the reaction, while a dot arrow means that the species can be modified by the context. In the right network, the parameters are $V = c_{P_{TRE3G}} V_1 k_4 k_6 / k_3 (k_5 + k_6)$ and $K = k_1 k_{-2} K_1 / c_{rtTA} k_{in} k_2$.

$\frac{k_{-2} c_{rtTA}^{total}}{k_{-2} + k_2 x_{Dox_i}}$. At this point we can use the axiom (Modifier) and replace x_{rtTA} in (25):

$$\emptyset \to mRNA \quad ; \quad \frac{V_2 x_{Dox_i}}{K_2 + x_{Dox_i}} \qquad (29)$$

with $V_2 = \frac{c_{P_{TRE3G}}^{total} V_1 c_{rtTA}^{total}}{K_1 + c_{rtTA}^{total}}$ and $K_2 = \frac{k_{-2} K_1}{k_2 (K_1 + c_{rtTA}^{total})}$.

We then apply axiom (Intermediate) on *GFP*, replacing the reactions (20), (21) and (22) by:

$$\emptyset \to GFP_a \quad ; \quad \frac{k_4 k_6}{k_5 + k_6} x_{mRNA}. \qquad (30)$$

Also, we can apply (Intermediate) on Dox_i, and replace reactions (13), (14), and (29) with:

$$\emptyset \to mRNA \quad ; \quad \frac{V_2 x_{Dox}}{\frac{k_1 K_2}{k_{in}} + x_{Dox}}. \qquad (31)$$

Finally, we use the axiom (Intermediate) followed by (Useless) on *mRNA*, and merge the reactions (19), (30) and (31) into:

$$\emptyset \to GFP_a \quad ; \quad V \frac{x_{Dox}}{K + x_{Dox}} \qquad (32)$$

with

$$V = \frac{c_{P_{TRE3G}}^{total} V_1 c_{rtTA}^{total} k_4 k_6}{k_3 (k_5 + k_6)(K_1 + c_{rtTA}^{total})}$$

$$K = \frac{k_1 k_{-2} K_1}{k_{in} k_2 (K_1 + c_{rtTA}^{total})}.$$

The conservation law becomes, after simplification of redundant equations and useless parameters:

$$\exists c \in \mathbb{C}.\; c = x_{Dox}$$

So we obtain the reaction network $\textit{Tet-On}_{simple} = R' \triangleright (\exists c.\; c = x_{Dox})$, with R' given by

$$\emptyset \to GFP_a \; ; \; V \frac{x_{Dox}}{x_{Dox} + K} \qquad\qquad GFP_a \to \emptyset \; ; \; k x_{GFP_a}.$$

$$\emptyset \to Dox_i \ ; \ k_{in} x_{Dox} \quad (13)$$
$$Dox_i \to \emptyset \ ; \ k_1 x_{Dox_i} \quad (14)$$
$$rtTA + Dox_i \to rtTADox \ ; \ k_2 x_{rtTA} x_{Dox_i} \quad (15)$$
$$rtTADox \to rtTA + Dox_i \ ; \ k_{-2} x_{rtTADox} \quad (16)$$
$$\emptyset \to mRNA \ ; \quad V_1 x_{P_{TRE3G}} \frac{x_{rtTADox}}{x_{rtTADox} + K_1} \quad (17)$$
$$mRNA \to \emptyset \ ; \ k_3 x_{mRNA} \quad (19)$$
$$\emptyset \to GFP \ ; \ k_4 x_{mRNA} \quad (20)$$
$$GFP \to \emptyset \ ; \ k_5 x_{GFP} \quad (21)$$
$$GFP \to GFP_a \ ; \ k_6 x_{GFP} \quad (22)$$
$$GFP_a \to \emptyset \ ; \ k x_{GFP_a} \quad (23)$$

Figure 11: Reactions of the detailed *Tet-On* network.

$Tet\text{-}On_{simple}$ is equivalent to the initial network:

Proposition 3. The initial and simplified networks are contextually equivalent:

$$Tet\text{-}On \simeq_{O,I} Tet\text{-}On_{simple}.$$

Notice that, aside from the kinetics and the conservation law, the simplified network is equal to the one we obtained with our qualitative simplification in [18].

7.3. NF-κB signaling

Finally, we present the simplification of a sub-module of a NF-κB signaling model from [20]. The nuclear factor κB (NF-κB) is involved in the regulation of several important genes for immune and stress response, cytokine production, and cell survival.

The initial NF-κB network, a sub-module of the model from [20], is depicted in Fig. 12 (left) and has 10 species and 13 reactions[1]. We simplify it into an equivalent network, depicted in Fig. 12 (right), with only 4 species and 2 reactions. It models the cytoplasmic part of the NF-κB signaling.

The initial network NF-κB $= R \triangleright L$, as in [20], contains six intermediates and internal species $I = \{$Ikk, Ikk|act, Ikk|inac, Ikk|act:IkBa, Ikk|act:IkBa:p50:p65, p50:p65@csl$\}$. The conservation law is:

$$L \quad = \quad \bigwedge_{X \in I} \exists c_X.\, x_X = c_X.$$

We only observe the non-intermediate molecules, i.e:

$$O = \{\text{A20}, \text{IkBa@csl}, \text{IkBa:p50:p65@csl}, \text{p50:p65@nsl}\}.$$

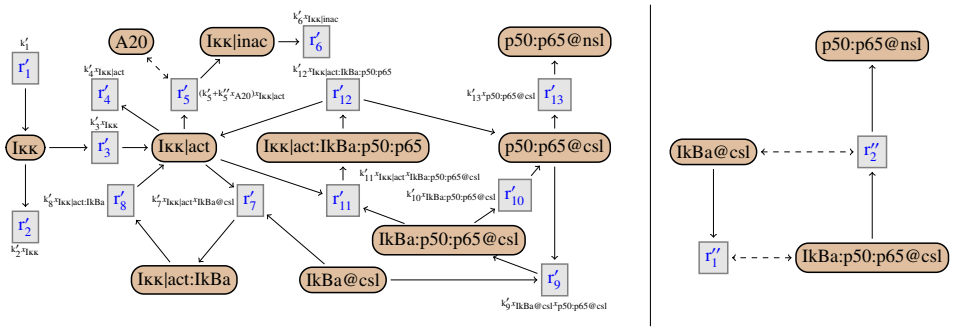
The reaction r'_1 represents the production of the transcription factor Ikk. This can go into an activated state Ikk|act with the reaction r'_3. The activated transcription factor can go into an inactivate state Ikk|inac, with reaction r'_5. Note that this reaction is the normalization of two reactions of the initial model. It merges reaction (Ikk|act $\to$ Ikk|inac; $k'_5 x_{\text{Ikk|act}}$) and reaction (Ikk|act + A20 $\to$ Ikk|inac + A20; $k''_5 x_{\text{A20}} x_{\text{Ikk|act}}$), with the modifier A20. The transcription factor can be degraded, in any of its three states, with reactions r'_2, r'_4, r'_6. In its activated form, it can bind to the protein IkBa@csl and form the complex Ikk|act:IkBa (r'_7). IkBa@csl can also bind first to p50:p65@csl, forming the complex IkBa:p50:p65@csl (r'_9), and then bind to the transcription factor, forming the complex Ikk|act:IkBa:p50:p65 (r'_{11}). These three complexes can dissociate, while consuming the protein IkBa@csl, in reactions r'_8, r'_{10} and r'_{12}. Finally, the protein p50:p65@csl can go into the nucleus, forming p50:p65@nsl, via reaction r'_{13}.

For the simplification, we can first remark that A20 is only used as a modifier, and therefore can be removed with the rule (Modifier), that replaces x_{A20} with c_{A20}. Then the intermediate Ikk|inac is only produced by reaction r'_5, and consumed by r'_6. We can remove it with the axiom (Intermediate). The resulting reaction is automatically merged with r'_4 by the normalization, forming the reaction:

$$r'_{14} = (\text{Ikk|act} \to \emptyset;\ (k'_4 + k'_5 + k'_5 c_{\text{A20}}) x_{\text{Ikk|act}}).$$

The intermediate Ikk is produced by reaction r'_1, and consumed by r'_2 and r'_3. We can use (Intermediate) to remove

[1] Note that in the original model from [20], reactions r'_9 and r'_{13} are reversible. We choose to not put the reverse reactions here, since they are ignored in their simplification.



$$r_1'' = \text{IkBa@csl} \to \emptyset \; ; \; \frac{\left[\frac{k_1'k_3'k_7'}{(k_2'+k_3')k_5''k_{10}}\left(\frac{k_{13}'}{k_9'}+x_{\text{IkBa@csl}}\right)+[K+x_{\text{A20}}]x_{\text{IkBa:p50:p65@csl}}\right]k_{10}'x_{\text{IkBa@csl}}}{\left(\frac{k_{13}'}{k_9'}+x_{\text{IkBa@csl}}\right)\left(\frac{k_4'+k_5'}{k_5''}+x_{\text{A20}}\right)}$$

$$r_2'' = \text{IkBa:p50:p65@csl} \to \text{p50:p65@nsl} \; ; \; \frac{\frac{k_{10}'k_{13}'}{k_9'}[K+x_{\text{A20}}]x_{\text{IkBa:p50:p65@csl}}}{\left(\frac{k_{13}'}{k_9'}+x_{\text{IkBa@csl}}\right)\left(\frac{k_4'+k_5'}{k_5''}+x_{\text{A20}}\right)}$$

Figure 12: Reaction graphs of the detailed (left) and simplified (right) NF-κB networks. Species are represented by circles, and reactions by squares. A dash arrow means that the species acts as a modifier in the reaction, while a dot arrow means that the species can be modified by the context. In the kinetic rates of the simplified reactions r_1'' and r_2'', we have $K = \dfrac{k_1'k_3'k_{11}' + (k_2'+k_3')(k_4'+k_5')k_{10}'}{(k_2'+k_3')k_5''k_{10}'}$.

it. The fusion of r_1' and r_2' gives a useless reaction, that we directly remove with the rule (Useless), while the fusion of r_1' and r_3' forms the reaction:

$$r_{15}' = (\emptyset \to \text{Iκκ|act};\ \frac{k_1'k_3'}{k_2'+k_3'}).$$

The simplification proceeds similarly, with the removal of the intermediates Iκκ|act:IkBa, Iκκ|act:IkBa:p50:p65, p50:p65@csl and Iκκ|act, while normalizing and discarding the useless reactions and merging reactions, until we obtain only the two reactions presented in Figure 12 (right), with the parameter:

$$K = \frac{k_1'k_3'k_{11}' + (k_2'+k_3')(k_4'+k_5')k_{10}'}{(k_2'+k_3')k_5''k_{10}'}.$$

In [20], Radulescu et al. applied their reduction method to the same sub-module. Their method is also based on eliminating the intermediate molecules, but does not compute the exact rates of the simplified reactions. Instead, an approximation is applied based on the dominant parameters. The authors find two reactions with the same structure than us, but not with the same rates, as we can see in Figure 13. Note that if we only keep the dominant rates of our simplified reactions, we find the same kinetic rates as in [20].

8. Further related work

Much effort has been spent on the simplification of biological systems. One simplification approach – as developed in [18] and described in the introduction – consists in carrying over structural simplification methods from compiler construction [29, 19], by looking at reaction networks as programs [30, 31, 32] with a non-deterministic semantics. As stated already, it preserves the final non-deterministic behavior of the network, in term of attractors, that are final connected components. Simplification methods preserving the attractors were also developed for boolean networks in [33]. Other structural meth-

$$\text{IkBa@csl} \to \emptyset \; ; \; \frac{\frac{k'_1 k'_{11}}{k''_5} x_{\text{IkBa@csl}} x_{\text{IkBa:p50:p65@csl}}}{(\frac{k'_{13}}{k'_9} + x_{\text{IkBa@csl}})(\frac{k'_5}{k''_5} + x_{\text{A20}})}$$

$$\text{IkBa:p50:p65@csl} \to \text{p50:p65@nsl} \; ; \; \frac{\frac{k'_1 k'_{11} k'_{13}}{k''_5 k'_9} x_{\text{IkBa:p50:p65@csl}}}{(\frac{k'_{13}}{k'_9} + x_{\text{IkBa@csl}})(\frac{k'_5}{k''_5} + x_{\text{A20}})}$$

Figure 13: Reactions of the NF-κB network after applying the simplification from [20].

ods aims at preserving the usual properties of Petri Nets [34, 35, 36]. In [37], subgraph epimorphisms are used to reduce reaction networks.

Approaches aiming at simplifying reaction networks while preserving the deterministic semantics are often motivated by a simplification method for the underlying Odes. Symmetries between species can be used to reduce the network [38, 39]. Many other simplification methods use equilibrium and the distinction between slow and fast reactions, as for instance methods based on invariant manifolds [40], quasi steady state [41, 42], quasi-equilibrium approximation [43] or tropicalization [44].

Some of those methods eliminate *intermediate species*, with concentration assumed at steady state. The prime example is Michaelis-Menten simplification of enzymatic reaction networks [17], which is justified by a quasi steady state approximation [42]. King and Altman [45] presented a simplification of a more general sequence of enzyme-catalyzed reactions into a single reaction, when the intermediate complexes are at steady state. It is based on a graphic procedure to compute the kinetic rate of the reduced reaction. In [46], a similar procedure is proposed, using the spanning trees of the graph of intermediates to compute the simplified Odes system of more general reaction networks. It should however be noticed that it is tedious to transform Odes back into reaction networks, since some Odes system do not stem from any reaction networks, while other do but not in a unique manner (see e.g. [47]).

9. Discussion

In this section, we discuss the different conditions and limitations of our method.

Most of the simplification methods for chemical reaction networks with deterministic semantics are based on Odes simplification, i.e. aim at producing a reduced Odes system for the initial reaction network [45, 48, 49, 46, 44, 45]. In contrast, in our work, we aim at obtaining a structural simplification, to derive a new simplified reaction network. Having the reaction network, rather than only the Odes, provides more details about the biological system and its structure, and makes it easier to understand. Since it is easy to compute the Odes system from the reaction network, while rebuilding a well-formed reaction network from an Odes system is difficult [47], it can be useful to simplify the network and not only the Odes system. Moreover, our simplification axioms are actually really close to the simplification axioms we presented in [18], for a non- deterministic semantics, that aims at preserving the final non-deterministic behavior of a reaction network. Therefore, we could build a simplification method that preserves both semantics, that was not possible when considering only the Odes.

The partial steady state assumption is one important necessary condition in our work to remove an intermediate species. Numerous simplification methods are based on quasi steady state and similar equilibrium conditions [50, 40, 20, 46, 44]. An exact equilibrium is necessary to have an exact simplification, i.e. to obtain a deterministic semantics of the reduced network that is per-

fectly equal to the initial one. Note however that in practice, reaching this exact equilibrium is not possible, the solutions will only be close to the partial steady state, and therefore the simplified network behavior will be slightly different than the original one. It would be interesting to try to estimate the error made in the simplification based on how close the network is to the partial steady state.

The impact of order of application of our simplification rules is an important question. Our simplification method is currently not confluent, meaning that different orders can produce different simplified networks. One source of non-confluence comes from the condition on the kinetic expressions of the consuming reactions in the axiom (Intermediate), that are not preserved when we apply the axiom on different intermediate species. Those conditions are necessary to be able to compute the steady state value, and to build the simplified network, and cannot be generalized further. In a follow-up paper [23], we showed that even if we restrain the axiom (Intermediate) to species with linear kinetics, the simplification is still not confluent. Indeed, some simplification orders can generate "dependent reactions", that are combination of other reactions of the network. By adding an axiom removing these dependent reactions, we proved that, independently of the order of application of the simplification axioms, we can always obtain the same reaction structure (without the kinetic rates) and the same Odes. However, the distribution of the kinetic rates between the reactions may be different.

We presented three axioms in this paper: the main axiom (Intermediate), and two related axioms. The axiom (Useless) is necessary to remove the useless reactions created by (Intermediate), while the axiom (Modifier) allows us to simplify all the species of a linear intermediate set. It could be interesting to have more simplification axioms that preserve the equivalence. We could for instance factorize a reaction of the form $(n * s_1 \rightarrow n * s_2;\ e)$ into the reaction $(s_1 \rightarrow s_2;\ n * e)$, or consider axioms based on the symmetries of the network. Note that it seems however difficult to have a complete set of axioms, that totally characterizes the equivalence.

Our simplification method has been implemented. The implementation includes the axioms presented in this article, as well as some of the others mentioned in the previous paragraph. Networks can be imported from SBML or text format. We have tested the methods on some networks from http://www.biomodels.net/. Theoretically the number of reactions may increase with the simplification, while the number of species decreases. In practice, the number of reactions decreases on every tested model. Note that when testing the simplification, the intermediate species were chosen arbitrarily by hand, but we do not check whether the system verifies appropriate conditions on parameters and initial conditions to justify the steady state assumption on these intermediate species.

10. Conclusion

We presented a new structural and contextual simplification of reaction networks, that preserves the deterministic semantics, under partial steady state hypothesis. It removes intermediate species, and can take into account conservation laws. It requires some conditions on the kinetic expressions, but is not restrained to linear kinetics. We have shown the usefulness of the simplification by applying it to biological networks. The simplification algorithm has been implemented and tested on biological models.

On the theoretical side, as future work, we want to investigate an approximated equivalence, when the species are only close to a steady state, and to compute the maximal error made by a simplification. A similar simplification method with a stochastic semantics will also be considered.

Acknowledgment: The authors would like to thank Michel Petitot, François Fages and Ovidiu Radulescu for their useful discussions, as well as members of the *PalBioSys* research network.

References

[1] M. Feinberg, Chemical reaction network structure and the stability of complex isothermal reactors—i. the deficiency zero and deficiency one theorems, Chemical Engineering Science 42 (10) (1987) 2229 – 2268. doi:http://dx.doi.org/10.1016/0009-2509(87)80099-4.
URL http://www.sciencedirect.com/science/article/pii/0009250987800994

[2] H. Kitano, Systems biology: a brief overview, Science 295 (5560) (2002) 1662–1664.

[3] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, U. Kummer, Copasi—a complex pathway simulator, Bioinformatics 22 (24) (2006) 3067–3074.

[4] L. Calzone, F. Fages, S. Soliman, BIOCHAM: an environment for modeling biological systems and formalizing experimental knowledge, Bioinformatics 22 (14) (2006) 1805–1807. doi:10.1093/bioinformatics/btl172.
URL http://dx.doi.org/10.1093/bioinformatics/btl172

[5] C. Kuttler, C. Lhoussaine, M. Nebut, Rule-based modeling of transcriptional attenuation at the tryptophan operon, Transactions on Computational Systems Biology XII (2010) 199–228.

[6] D. T. Gillespie, Deterministic limit of stochastic chemical kinetics, The Journal of Physical Chemistry B 113 (6) (2009) 1640–1644.

[7] H. De Jong, Modeling and simulation of genetic regulatory systems: A literature review, Journal of Computational Biology 9 (1).

[8] M. Ashyraliyev, Y. Fomekong-Nanfack, J. A. Kaandorp, J. G. Blom, Systems biology: parameter estimation for biochemical models, Febs Journal 276 (4) (2009) 886–902.

[9] N. Chabrier, F. Fages, Symbolic model checking of biochemical networks, in: C. Priami (Ed.), Computational Methods in Systems Biology, First International Workshop, CMSB 2003, Roverto, Italy, February 24-26, 2003, Proceedings, Vol. 2602 of Lecture Notes in Computer Science, Springer, 2003, pp. 149–162. doi:10.1007/3-540-36481-1_13.
URL http://dx.doi.org/10.1007/3-540-36481-1_13

[10] J. Uhlendorf, S. Bottani, F. Fages, P. Hersen, G. Batt, Towards real-time control of gene expression: controlling the hog signaling cascade., in: Pacific Symposium On Biocomputing, World Scientific, 2011, pp. 338–349.

[11] J. Kim, J. Reed, OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains, BMC Systems Biology 4 (1) (2010) 53+. doi:10.1186/1752-0509-4-53.
URL http://dx.doi.org/10.1186/1752-0509-4-53

[12] J. Niehren, M. John, C. Versari, F. Coutte, P. Jacques, Qualitative Reasoning about Reaction Networks with Partial Kinetic Information, in: Computational Methods for Systems Biology, Vol. 9308, Nantes, France, 2015, pp. 157–169.
URL https://hal.inria.fr/hal-01163391

[13] F. Coutte, J. Niehren, D. Dhali, M. John, C. Versari, P. Jacques, Modeling Leucine's Metabolic Pathway and Knockout Prediction Improving the Production of Surfactin, a Biosurfactant from Bacillus Subtilis, Biotechnology Journal 10 (8) (2015) 1216–34.
URL https://hal.inria.fr/hal-01153704

[14] F. Fages, S. Gay, S. Soliman, Inferring reaction systems from ordinary differential equations, Theor. Comput. Sci. 599 (2015) 64–78. doi:10.1016/j.tcs.2014.07.032.
URL http://dx.doi.org/10.1016/j.tcs.2014.07.032

[15] U. Mäder, A. G. Schmeisky, L. A. Flórez, J. Stülke, SubtiWiki–a comprehensive community resource for the model organism *Bacillus subtilis*., Nucleic Acids Res. 40 (2012) 1278–1287.

[16] O. Radulescu, A. N. Gorban, A. Zinovyev, V. Noel, Reduction of dynamical biochemical reaction networks in computational biology, Frontiers in Genetics 3.

[17] L. Michaelis, M. L. Menten, Die kinetik der invertinwirkung, Biochem. z 49 (333-369) (1913) 352.

[18] G. Madelaine, C. Lhoussaine, J. Niehren, Attractor equivalence: An observational semantics for reaction networks, in: Formal Methods in Macro-Biology, Springer, 2014, pp. 82–101.

[19] M. Schmidt-Schauss, D. Sabel, J. Niehren, J. Schwinghammer, Observational Program Calculi and the Correctness of Translations, Journal of

Theoretical Computer Science (TCS) 577 (2015) 98–124. doi:10.1016/j.tcs.2015.02.027.
URL https://hal.inria.fr/hal-00824349

[20] O. Radulescu, A. N. Gorban, A. Zinovyev, A. Lilienbaum, Robust simplifications of multiscale biochemical networks, BMC systems biology 2. doi:10.1186/1752-0509-2-86.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2654786&tool=pmcentrez&rendertype=abstract

[21] J. Gagneur, S. Klamt, Computation of elementary modes: a unifying framework and the new binary approach, BMC bioinformatics 5 (1) (2004) 1.

[22] G. Madelaine, C. Lhoussaine, J. Niehren, Structural simplification of chemical reaction networks preserving deterministic semantics, in: Computational Methods in Systems Biology, Springer, 2015, pp. 133–144.

[23] G. Madelaine, C. Lhoussaine, J. Niehren, Structural simplification of chemical reaction networks preserving deterministic semantics, in: Computational Methods in Systems Biology, Springer, 2015.

[24] M. Sáez, C. Wiuf, E. Feliu, Graphical reduction of reaction networks by linear elimination of species, Journal of Mathematical Biology (2016) 1–43.

[25] M. Gossen, 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).
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=49329&tool=pmcentrez&rendertype=abstract

[26] M. Gossen, S. Freundlieb, G. Bender, G. Müller, W. Hillen, H. Bujard, Transcriptional activation by tetracyclines in mammalian cells., Science (New York, N.Y.) 268 (5218).
URL http://www.ncbi.nlm.nih.gov/pubmed/7792603

[27] Z. Huang, C. Moya, A. Jayaraman, J. Hahn, Using the Tet-On system to develop a procedure for extracting transcription factor activation dynamics., Molecular bioSystems 6 (10). doi:10.1039/c003229h.
URL http://www.ncbi.nlm.nih.gov/pubmed/20552111

[28] J. Uhlendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt, 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).

[29] A. M. Pitts, Operational semantics and program equivalence, in: Applied Semantics, Vol. 2395 of LNCS, Springer, 2000. doi:10.1007/3-540-45699-6_8.
URL http://dx.doi.org/10.1007/3-540-45699-6_8

[30] A. Regev, E. Shapiro, Cellular abstractions: Cells as computation, Nature 419 (6905) (2002) 343–343.

[31] L. Cardelli, G. Zavattaro, On the computational power of biochemistry, Springer, 2008.

[32] M. John, C. Lhoussaine, J. Niehren, C. Versari, Biochemical reaction rules with constraints, in: 20th European Symposium on Programming Languages (ESOP), Springer, 2011, pp. 338–357.

[33] A. Naldi, E. Remy, D. Thieffry, C. Chaouiya, Dynamically consistent reduction of logical regulatory graphs, Theoretical Computer Science 412 (21). doi:10.1016/j.tcs.2010.10.021.
URL http://linkinghub.elsevier.com/retrieve/pii/S0304397510005839

[34] G. Berthelot, G. Roucairol, Reduction of petri-nets, in: Mathematical Foundations of Computer Science 1976, LNCS, 1976.
URL http://dx.doi.org/10.1007/3-540-07854-1_175

[35] T. Murata, J. Koh, Reduction and expansion of live and safe marked graphs, Circuits and Systems, IEEE Transactions on 27 (1).
URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1084711

[36] S. Haddad, J.-F. Pradat-Peyre, New efficient petri nets reductions for parallel programs verification, Parallel Processing Letters 16 (01) (2006) 101–116.

[37] S. Gay, S. Soliman, F. Fages, A graphical method for reducing and relating models in systems biology., Bioinformatics (Oxford, England) 26 (18). doi:10.1093/bioinformatics/btq388.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2935413&tool=pmcentrez&rendertype=abstract

[38] K. Batmanov, C. Kuttler, F. Lemaire, C. Lhoussaine, C. Versari, Symmetry-based model reduction for approximate stochastic analysis, in: Computational Methods in Systems Biology 2012 (CMSB 2012), Vol. 7605, London, United Kingdom, 2012, pp. 49–68.
URL http://hal.archives-ouvertes.fr/hal-00713386

[39] L. Cardelli, M. Tribastone, M. Tschaikowski, A. Vandin, Comparing chemical reaction networks: A categorical and algorithmic perspective.

[40] A. N. Gorban, I. V. Karlin, Method of invariant manifold for chemical kinetics, Chemical Engineering Science 58 (21) (2003) 4751–4768.

[41] M. Bodenstein, Eine theorie der photochemischen reaktionsgeschwindigkeiten, Z. phys. Chem 85 (329) (1913) 0022–3654.

[42] L. A. Segel, M. Slemrod, The quasi-steady-state assumption: a case study in perturbation, SIAM review 31 (3) (1989) 446–477.

[43] A. N. Gorban, I. V. Karlin, P. Ilg, H. C. Öttinger, Corrections and enhancements of quasi-equilibrium states, Journal of non-newtonian fluid mechanics 96 (1) (2001) 203–219.

[44] 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, 2014, pp. 27–35.

[45] E. L. King, C. Altman, A schematic method of deriving the rate laws for enzyme-catalyzed reactions, The Journal of physical chemistry 60 (10) (1956) 1375–1378.

[46] J. Gunawardena, A linear framework for time-scale separation in nonlinear biochemical systems, PLoS One 7 (5) (2012) e36321.

[47] F. Fages, S. Gay, S. Soliman, Inferring reaction models from odes, in: Computational Methods in Systems Biology, Springer, 2012, pp. 370–373.

[48] D. Degenring, C. Froemel, G. Dikta, R. Takors, Sensitivity analysis for the reduction of complex metabolism models, Journal of Process Control 14 (7) (2004) 729–745.

[49] V. Danos, J. Feret, W. Fontana, R. Harmer, J. Krivine, Abstracting the differential semantics of rule-based models: exact and automated model reduction, in: Logic In Computer Science (LICS), IEEE, 2010, pp. 362–381.

[50] B. L. Clarke, General method for simplifying chemical networks while preserving overall stoichiometry in reduced mechanisms, The Journal of chemical physics 97 (6) (1992) 4066–4071.