Inferring Reaction Systems
from Ordinary Differential Equations

François Fages, Steven Gay, Sylvain Soliman

Inria Paris-Rocquencourt, France

Abstract

In Mathematical Biology, many dynamical models of biochemical reaction systems are presented with Ordinary Differential Equations (ODE). Once kinetic parameter values are fixed, this simple mathematical formalism completely defines the dynamical behavior of a system of biochemical reactions and provides powerful tools for deterministic simulations, parameter sensitivity analysis, bifurcation analysis, etc. However, without requiring any information on the reaction kinetics and parameter values, various qualitative analyses can be performed using the *structure* of the reactions, provided the reactants, products and modifiers of each reaction are precisely defined. In order to apply these structural methods to parametric ODE models, we study a mathematical condition for expressing the consistency between the structure and the kinetics of a reaction, without restricting to Mass Action law kinetics. This condition, satisfied in particular by standard kinetic laws, entails a remarkable property of independence of the influence graph from the kinetics of the reactions. We derive from this study a heuristic algorithm which, given a system of ODEs as input, computes a system of reactions with the same ODE semantics, by inferring well-formed reactions whenever possible. We show how this strategy is capable of automatically curating the writing of ODE models in SBML, and present some statistics obtained on the model repository biomodels.net.[1]

Keywords: systems biology, chemical reaction network theory, ordinary differential equations, SBML.

1. Introduction

In Mathematical Biology, many models are presented as a system of Ordinary Differential Equations (ODEs). Once the kinetic parameter values are fixed, this simple mathematical formalism completely defines the dynamical behavior of a system of biochemical

---

[1]This paper is an extended version of a communication presented at CMSB'12. The algorithms described in this paper are implemented in the open-source software modeling platform Biocham [1, 2] available at `http://lifeware.inria.fr/biocham/` release 3.5. The models used in the experiments are available from `http://www.biomodels.net/` release 24

reactions. It provides powerful tools for both transient and steady-state analysis via numerical integration, parameter sensitivity analysis, or bifurcation analysis, but only when kinetic information is available.

In absence of knowledge on the kinetics of each reaction, various qualitative analyses can nevertheless be performed using the *structure* of the reactions. This approach has rapidly developed in Systems Biology for reasoning on large interaction networks, with for instance, the analysis of qualitative attractors in a logical dynamics of gene networks *à la* Thomas [3, 4, 5], reachability and temporal logic properties in reaction networks [6, 7, 8, 9, 10], structural invariants in the Petri net representation of the reactions [11, 12, 13, 14, 15, 16], or model reductions using graph theory concepts [17, 18]. These qualitative analysis tools do not rely on kinetic information, but on the structure of the reaction network which has thus to be correctly written as a set of formal reactions, with well-identified reactants, products and modifiers (and in certain cases their stoichiometry) for each reaction.

For instance, in [19], it is elaborated that structural information hidden in kinetic laws may affect the results obtained from structural analyses, such as elementary mode analysis [20], flux balance analysis [21], chemical organization theory [22], deficiency analysis or chemical reaction network theory [23, 24].

It is worth noticing that these structural analyses may also directly support dynamic analyses. For instance, [25] applies network decomposition for a modular parameter estimation approach, [13] introduces a structural persistence criterion, Petri net place invariants reveal conservation laws in [26], while transition invariants can be used to identify fragile nodes and the core of a network [27], or to determine steady state solutions [28].

Furthermore, knowing the correct structure of each reaction is mandatory when a reaction network must be interpreted as a stochastic process (Continuous-Time Markov Chain, CTMC) *à la* Gillespie [29].

The question of the correct identification of a structured reaction model from a system of ODEs is thus important and is not new. Actually for the restricted case of models with only Mass Action kinetics a general solution is provided in [30]. This approach was evolved over the years, see for instance [31] for sparse/dense/core solutions when numerical values are provided for the parameters, or [32] for unicity conditions in the symbolic case, still in the restricted framework of mass action law kinetics. In [19], the authors present an algorithm that uncovers hidden structural information for some Systems Biology Markup Language (SBML) [33, 34] models of the `biomodels.net` repository [35], with restricting to reaction models without inhibitors.

In this paper, we describe an algorithm for finding a reaction models for a given system of ODEs, considering reaction with inhibitors and general kinetic expressions. The first contribution of this paper is to propose a mathematical condition for expressing the consistency between the kinetic expression and the reactant-product-inhibitor structure of a reaction. We introduce well-formedness (Def. 2.5) and strictness (Def. 2.6) conditions for reactions, and show that they are satisfied by standard kinetics such as Mass Action law, Michaelis-Menten, Hill and negative Hill kinetics. The well-formedness condition is also shown to entail a property of independence of the influence graph (or symbolic Jacobian matrix) from the kinetics of the reactions (Thm. 2.16). This result generalizes a previous result in [36, 37]

to reactions with inhibitors. It shows that the influence graph of a well-formed reaction system with inhibitors is essentially independent of the kinetics, can be computed in linear time in the number of reactions when the number of species per reaction is bounded, and can thus advantageously be used to perform multi-stationarity analyses by circuit analysis *à la* Thomas [38, 39, 40, 41, 42, 43, 5, 4, 3].

The second contribution of this paper is to use these well-formedness and strictness conditions to prove the completeness of a new general algorithm for inferring a reaction system equivalent to an ODE system. This algorithm, of time complexity in $O(n \times t)$ where n is the number of variables and t the number of terms in the ODE, is shown to preserve the ODE semantics of the reactions (soundness 3.10), as well as their well-formedness when applied to the ODE semantics of a non-decomposable well-formed reaction system (weak completeness 3.12).

Our third contribution is to show that our algorithm can be used to automatically curate the writing of ODE models with reactions, as required in SBML. The fact that SBML has become a standard for sharing and publishing models has helped in making modelers formalize the reaction structure of their models. Unfortunately, SBML does not enforce any strong coherence between the structure and the kinetics of a reaction. Therefore the structural interpretation of models transcribed in reaction-based formalisms such as SBML may vary according to different choices of representation of the original ODE model as a reaction system, and may invalidate some structural analyses. We compare our results to the one presented in [19], and provide some statistics obtained on the rewriting in SBML of the curated part of the `biomodels.net` repository, showing that our method is able to automatically decrease the number of non well-formed reaction systems from 65% to 29%.

2. A Theory of Well-formed Reactions and Kinetics

In this section, we consider a finite set $S = \{x_1, \ldots, x_s\}$ of s molecular species, and a finite set of n reactions over S which are formally represented as multiset rewriting rules with kinetic expressions.

Multisets are used for representing reactants, products and inhibitors in reactions. A multiset s of molecular species is a function $S \longrightarrow \mathbb{N}$ which gives the number (stoichiometric coefficient) $s(x)$ of each molecular species $x \in S$ in s. We have $s(x) = 0$ if x does not belong to s, and $s(x) \geq 1$ if x belongs to s, which is also written $x \in s$ by abuse of notation. The empty multiset is written $\emptyset$. Equivalently, a multiset s will also be denoted by the linear expression $\sum_{i=1}^{m} s(x_i) \times x_i$, which gives the stoichiometric coefficients of each molecular species x_i in s. This corresponds to the classical chemical notation $2H + O \longrightarrow H_2O$.

We shall now introduce the well-formedness and strictness conditions and describe some of their properties.

2.1. Well-formedness and Strictness Conditions

In the following definition, a reaction is composed of multisets for reactants, products and inhibitors that are not assumed to be disjoint.

Definition 2.1. *A reaction is a quadruple* (r, m, p, f), *where r is the multiset of reactants, m the multiset of inhibitors, p the multiset of products, and f, called kinetic expression, a mathematical function over molecular species concentrations, $f : \mathbb{R}^s \longrightarrow \mathbb{R}$. A reaction system is a finite set of reactions.*

The species that are both reactants and products in a reaction are called *catalysts*. For the sake of readability, a reaction (r, m, p, f) will also be written $r \ / \ m \xrightarrow{f} p$ or just $r \xrightarrow{f} p$ if it has no inhibitor, i.e. when $m = \emptyset$. The kinetic expression will also be omitted if it is not relevant.

Example 2.2. *For instance, the following reaction, transcribed from Kohn's map of the cell cycle [44],*

$$pMPF + Cdc25 \xrightarrow{k_1 \times pMPF \times Cdc25} MPF + Cdc25$$
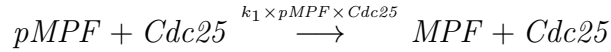
expresses the activation of the Mitosis Promoting Factor MPF by the kinase Cdc25. It has as rate law $f = k_1 \times pMPF \times Cdc25$, i.e. a Mass Action kinetics with parameter k_1. In this reaction, pMPF is a reactant, MPF a product, Cdc25 a reactant and a product at the same time, i.e. a catalyst in our terminology, and there is no inhibitor.

A simplified version of that reaction can be written by omitting the kinase Cdc25, as follows:

$$pMPF \xrightarrow{V \times pMPF/(K+pMPF)} MPF$$
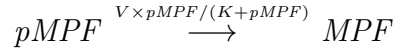
That form typically derives from three reactions describing the reversible association of pMPF and Cdc25 and the dissociation to MPF, by making a quasi steady state approximation on Cdc25, which results in a Michaelis-Menten kinetics with parameters K and V.

It is worth noting that in a reaction, a reactant or a product can also be an inhibitor if it appears in m.

Example 2.3. *For instance, the Botts-Morales general modifier mechanism accounts for a modifier M that can enhance and slow down a reaction $A \longrightarrow B$, depending on its concentration [45]. This can be represented in our setting by a reaction of the form*

$$A + M \ / \ M \longrightarrow B + M.$$
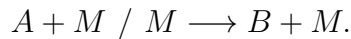
SBML does not distinguish between catalysts and inhibitors which are just considered as "modifiers" in SBML annotations. However we find it useful for the theory developed here to distinguish between the activation or inhibitory effects of a modifier, and mark it syntactically as such in the structure of the reaction. If a modifier has both activation and inhibitory effects, it will just appear in r, m and p in our setting, without loss of generality.

It is also worth noting that we consider only irreversible reactions, as in Feinberg's Chemical Reaction Network theory [23]. A reversible reaction is thus represented by two reactions, one for each direction. This is one important difference with the Systems Biology

Markup Language (SBML) that permits the declaration of a reversible reaction with only one single kinetic expression which can be negative.

These distinctions do not affect the system of ODEs that is classically associated to a reaction system by the Reaction Rate Equation as follows:

**Definition 2.4.** *The ODE semantics of a reaction system*

$$R = \{ \; r_i \; / \; m_i \xrightarrow{f_i} p_i \; \}_{i=1,\ldots,n}$$

over molecules $\{x_1, \ldots, x_s\}$, is the system of ordinary differential equations

$$\dot{x}_j = \sum_{i=1}^{n} (p_i(x_j) - r_i(x_j)) \times f_i$$

for $1 \leq j \leq s$.

Our aim is to go in the reverse direction, that is to infer from any ODE system a reaction system with the same ODE semantics. Let us first remark that any ODE system $\dot{x}_j = g_j$ can be trivially transcribed in a reaction system using artificial synthesis reactions for each molecular species, with the terms of the differential equation as kinetic expressions, as follows:

$$\emptyset \xrightarrow{g_j} x_j$$
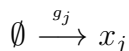
Since the ODE semantics is identical to the original ODE system, this is correct as far as numerical simulations are concerned, but prevents the use of structural analysis methods or stochastic simulations as the structures of the reactions are totally meaningless. It is worth remarking that some ODE models have nevertheless been transcribed in SBML using that scheme, since it does not affect simulations. This is the case for instance of model `BIOMD0000000008.xml` in biomodels.net for the ODE model of [46]. We will use that example in Section 3.1 to illustrate our reaction inference algorithm and its capability of automatically curating the writing in SBML of ODE models as reaction systems.

In order to try to infer meaningful reactions from ODEs, we are interested in mathematical conditions for expressing the consistency of the kinetic expression f with the structure (r, m, p) of a reaction. Furthermore, since it is common practice to aggregate a system of elementary reactions in one abstract reaction with more complex kinetics (the simplest example of which are Michaelis-Menten and Hill kinetics for enzymatic reactions), we do not content ourselves with elementary kinetic expressions such as mass action law kinetics, but seek abstract consistency properties that can be applied to any mathematical expression given as kinetics. This is in contrast to most work on chemical reaction network theory [23, 24, 32], but in accordance with the use in SBML of MathML for writing the kinetic expressions without any limitation on the use of mathematical symbols.

Definition 2.5. *A reaction (r, m, p, f) over molecular species $\{x_1, \ldots, x_s\}$ is* well-formed *if the following conditions hold:*

5

1. $f(x_1, \ldots, x_s)$ *is a partially differentiable function, non-negative on* $\mathbb{R}_+^s$;
2. $x_i \in r$ *if and only if* $\partial f / \partial x_i(\vec{x}) > 0$ *for some value* $\vec{x} \in \mathbb{R}_+^s$;
3. $x_i \in m$ *if and only if* $\partial f / \partial x_i(\vec{x}) < 0$ *for some value* $\vec{x} \in \mathbb{R}_+^s$.

The first condition expresses that the kinetic expression must be a differentiable and non-negative function for all non-negative values of the variables. The second (resp. third) condition states that the partial derivative of f w.r.t. a reactant (resp. an inhibitor) must be positive (resp. negative) for some (not necessarily all) non-negative values of the variables.

It is worth noting that we do not impose the *monotonicity* condition that for any variable $x_i \in V$, $\partial f / \partial x_i$ should be either non-negative on the positive orthant, or non-positive on the positive orthant. In our setting, a molecular species can thus be both a reactant and an inhibitor in a well-formed reaction, depending on the values of the concentrations. On the other hand we shall make use of the following:

Definition 2.6. *A reaction* (r, m, p, f) *is* strict *if its kinetics* $f(x_1, \ldots, x_s) = 0$ *whenever* $x_j = 0$ *for any* x_j *such that* $r(x_j) > 0$.

This condition expresses that the kinetics must be zero if the concentration of one of the reactants is zero. If the kinetics is a rational expression, that strictness condition implies that the kinetic expression is a product of the reactants with a fractional expression defined for all non-negative values of the variables. More generally it enforces the positivity of the system:

Definition 2.7 (Positive System). *A dynamical system over* $\mathbb{R}^k$ *is called positive if* $\mathbb{R}_+^k$ *is an invariant set for the system, i.e.,* $\forall x_0 \geq 0, t \geq 0 \ x(t, x_0) \geq 0$

Proposition 2.8 (Positivity). *The ODE semantics of a well-formed and strict reaction system defines a positive system.*

Proof. In Definition 2.4 we have $\dot{x}_j = \sum_{i=1}^{n} (p_i(x_j) - r_i(x_j)) \times f_i$ and since the system is well-formed, the f_i are all non-negative. The only negative terms thus have $r_i(x_j) > 0$ and from the strictness condition this entails that $f_i = 0$ when $x_j = 0$. Hence $\dot{x}_j \geq 0$ whenever $x_j = 0$ since it is a sum of non-negative terms. Therefore x_j cannot become negative when its initial value is non-negative, and since this holds for all j, the system is positive. $\qquad\square$

The strictness condition excludes the writing of a reversible reaction with one single reaction by summing the kinetic expressions of each direction, as allowed in SBML, when the reactants differ from the products.

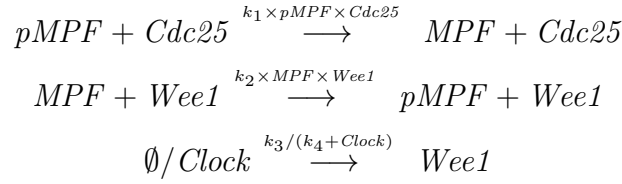It also excludes the existence of a strict well-formed reaction system for any ODE system, as shown by

Example 2.9. *The equation* $\dot{x} = -k$ *is not the ODE semantics of any strict well-formed reaction system, since that ODE defines a non-positive system (Prop. 2.8). That ODE can be associated to the non-strict well-formed reaction system*

$$x \xrightarrow{l \times x} 2 \times x$$

$$x \xrightarrow{k+l \times x} \emptyset$$
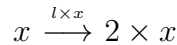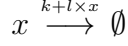
(where the kinetic expression is not null when $x = 0$). This is the result computed in that case by Algorithm 3.6 described later.

Example 2.10. *As for an example with inhibitors, let us consider the following three reactions representing the core of the action of the Circadian clock on the Cell Cycle, as described by Matsuo et al. [47]:*
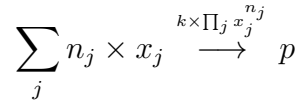
$$pMPF + Cdc25 \xrightarrow{k_1 \times pMPF \times Cdc25} MPF + Cdc25$$
$$MPF + Wee1 \xrightarrow{k_2 \times MPF \times Wee1} pMPF + Wee1$$
$$\emptyset / Clock \xrightarrow{k_3/(k_4 + Clock)} Wee1$$

where k_1, k_2, k_3 are parameters. The first reaction is the one of Example 2.2. Those reactions are well-formed and strict. In particular, we have $\partial Wee1/\partial Clock = \partial\big(k_3/(k_4 + Clock)\big)/\partial Clock$ $= -k_3/(k_4 + Clock)^2 < 0$ for showing the inhibitory effect of Clock in the synthesis reaction of Wee1. Their ODE semantics is
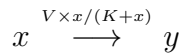
$$
\begin{aligned}
p\dot{M}PF &= k_2 \times MPF \times Wee1 - k_1 \times pMPF \times Cdc25 \\
\dot{M}PF &= k_1 \times pMPF \times Cdc25 - k_2 \times MPF \times Wee1 \\
\dot{W}ee1 &= k_3/(k_4 + Clock) \\
\dot{C}dc25 &= 0 \\
\dot{C}lock &= 0
\end{aligned}
$$

The well-formedness and strictness conditions are satisfied by standard kinetic laws. One can easily check
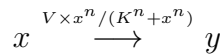
Proposition 2.11. *Reactions with*
mass action law kinetics:

$$\sum_j n_j \times x_j \xrightarrow{k \times \Pi_j x_j^{n_j}} p$$

Michaelis-Menten kinetics:

$$x \xrightarrow{V \times x/(K+x)} y$$

Hill kinetics:

$$x \xrightarrow{V \times x^n/(K^n + x^n)} y$$

or negative Hill kinetics:

$$\emptyset / x \xrightarrow{V/K^n + x^n} y$$
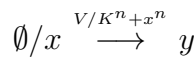
with rate constants $k, V, K > 0$ and exponent $n \geq 1$, are well-formed and strict.

We shall see in Section 4.2 that these conditions are currently violated in a majority of reaction systems of the biomodels.net repository, but that most of them can be automatically corrected by modifying their structure and writing in SBML, without changing their ODE semantics.

2.2. Influence Graph associated to a Well-formed Reaction System

The influence graph between molecular species induced by the ODE semantics of a well-formed reaction system enjoys a remarkable property of independence from the kinetics, which we present in this section. Influence graphs have been initially introduced in the setting of gene regulatory networks [3] as a simple abstraction enabling reasoning about complex regulation mechanisms. These graphs completely abstract from the precise interactions, especially at post-transcriptional level, and retain only the activation and inhibitory effects on gene transcription. As conjectured in [4], the existence of a positive circuit (resp. a negative circuit) in an influence graph has been proved to be a necessary condition for multi-stationarity, e.g. for cell differentiation, (resp. for oscillations, e.g. for homeostasis) in different formalisms, and in particular for ODE systems in [43, 39, 40, 41, 42] and recently in [38] for the ODE semantics of non-linear reaction systems.

Here, we show that in a well-formed reaction system, and under a very general assumption, the influence graph of the reactions is identical to the influence graph of the ODE semantics of the reactions.

On the one hand, in an ODE system, the influence graph is mathematically defined by the signs of the coefficients in the Jacobian matrix of the system, $(\partial \dot{x}_i / \partial x_j)$, as follows:

Definition 2.12. *The* differential influence graph (DIG) *associated to (the ODE semantics of) a reaction system is the graph that has for vertices the molecular species, and for labeled edges the following set of signed edges:*

$$\{x_i \longrightarrow^+ x_j \mid \partial \dot{x}_j / \partial x_i(\vec{x}) > 0 \text{ for some value } \vec{x} \in \mathbb{R}_+^s\}$$
$$\cup \{x \longrightarrow^- x_j \mid \partial \dot{x}_j / \partial x_i(\vec{x}) < 0 \text{ for some value } \vec{x} \in \mathbb{R}_+^s\}$$

Example 2.13. *The DIG of Example 2.10 can be depicted by the following graph:*
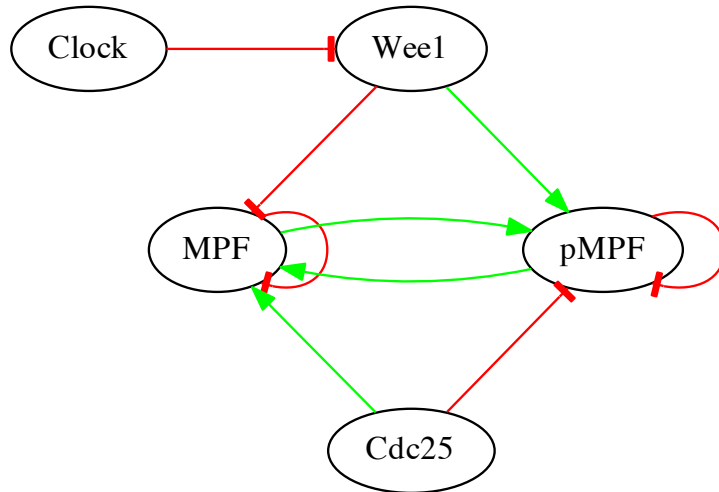
where the positive influences are represented by green arrows with triangular tips, and nega- tive influences are represented by red arrows with blunt tips. For instance, the negative influ- ence of Clock on Wee1 comes from the negative sign of $\partial \dot{Wee1}/\partial Clock$ as detailed in 2.10. There are negative loops on MPF and MPF since $\partial \dot{MPF}/\partial MPF = -k_2 Wee1 < 0$ and $\partial p\dot{MPF}/\partial pMPF = -k_2 Cdc25 < 0$, and not on Cdc25 since $\dot{Cdc25} = 0$. Note that a useful circuit analysis in this example would necessitate considering the reactions of formation of MPF and is beyond the scope of this paper.

On the other hand, in a reaction system, one can define an influence graph directly from the stoichiometry of the reactions, ignoring the kinetics, as follows:

Definition 2.14. *The* stoichiometric influence graph (SIG) *associated to a finite set R of reactions is the graph that has for vertices the molecular species, and for labeled edges the following set of signed edges:*
$$\{x \longrightarrow^+ y \quad | \quad either \ p_i(y) - r_i(y) > 0 \ and \ r_i(x) > 0,$$
$$or \ p_i(y) - r_i(y) < 0 \ and \ m_i(x) > 0, \ for \ some \ reaction \ i\}$$
$$\cup \ \{x \longrightarrow^- y \quad | \quad either \ p_i(y) - r_i(y) < 0 \ and \ r_i(x) > 0,$$
$$or \ p_i(y) - r_i(y) > 0 \ and \ m_i(x) > 0, \ for \ some \ reaction \ i\}$$

Intuitively, there is a positive (resp. negative) arc from x to y if x is a reactant in a reaction that produces more (resp. less) y than it consumes, or an inhibitor in a reaction that consumes more (resp. less) y than it produces.

Unlike the DIG, which needs to compute the sign of partial derivatives, the SIG can be easily computed in linear time in the number of reactions, assuming that the number of species per reaction is bounded, since it is sufficient to parse the stoichiometric coefficients of the reactions. As already shown in [36], the SIG is an over-approximation of the DIG:

Theorem 2.15 ([36])**.** *For any finite set R of well-formed reactions, the DIG of R is a subgraph of the SIG of R.*

We show here that, even in the presence of inhibitors, the SIG is in fact identical to the DIG with an extra assumption. Let us say that a tuple of molecular species (x, y) is in *conflict in an influence graph* if we have both $x \longrightarrow^+ y$ and $x \longrightarrow^- y$.

Theorem 2.16. *For any finite set R of well-formed reactions such that the SIG of R contains no conflict, the DIG and the SIG are identical.*

Proof. We just have to prove that the SIG is a subgraph of the DIG. Let us consider an arc $x \longrightarrow^+ y$ in the SIG. By Definition 2.14 there exists a reaction i with either $p_i(y) - r_i(y) > 0$ and $r_i(x) > 0$, or $p_i(y) - r_i(y) < 0$ and $m_i(x) > 0$. Since the reaction is well-formed, we have either $p_i(y) - r_i(y) > 0$ and $\partial f_i/\partial x(\vec{z}) > 0$, or $p_i(y) - r_i(y) < 0$ and $\partial f_i/\partial x(\vec{z}) < 0$, for some $\vec{z} \in \mathbb{R}_+^s$. Now, if $p_i(y) - r_i(y) > 0$ then f_i occurs in $\dot{y}$ with a positive sign. Since $\partial f_i/\partial x(\vec{z}) > 0$ and there is no conflict in the SIG, we thus get $\partial \dot{y}/\partial x(\vec{z}) > 0$, i.e. $x \longrightarrow^+ y$ is in the DIG. Similarly, if $p_i(y) - r_i(y) < 0$, f_i occurs in $\dot{y}$ with a negative sign and $\partial f_i/\partial x(\vec{z}) < 0$, hence $\partial \dot{y}/\partial x(\vec{z}) > 0$, i.e. $x \longrightarrow^+ y$ is in the DIG. The proof for an arc $x \longrightarrow^- y$ in the SIG is symmetrical. $\square$

Corollary 2.17. *The DIG of a finite set of well-formed reactions without conflict in its SIG, is independent of the kinetic expressions.*

Corollary 2.18. *The DIG of a finite set of well-formed reactions without conflict in its SIG, is computable in linear time in the number of reactions, when the number of species appearing in a reaction is bounded.*

The SIG of Example 2.10 is trivial to compute and since it contains no conflict, we can predict by Theorem 2.16 that it is identical to its DIG depicted in Example 2.13.

Example 2.19. *As for an example of conflict, in the simplified model of the yeast cell cycle of [48], the double activation reactions of MPF through Cdc25 and Wee1 [44], are simplified in a single autocatalytic reaction in parallel with a deactivation reaction:*

$$pMPF + MPF \longrightarrow 2 \times MPF$$
$$MPF \longrightarrow pMPF$$
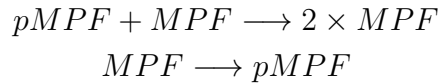
Such reactions create a conflict in the SIG, namely $MPF \longrightarrow^- pMPF$ and $MPF \longrightarrow^+ pMPF$. In general, there is a possibility that such conflicting direct influences in the SIG may be balanced in the ODEs and do not appear in the DIG. This situation is however quite pathological and rare in practice, and occurs when over-simplifications are made. For instance, Kohn's map of the cell cycle control [44] contains 800 reactions [8] and does not contain any conflict in its SIG [36]. The conflict of influences between MPF and pMPF in Tyson's model comes from the compression in one loop of the two positive circuits through Wee1 and Cdc25 respectively. The decompression of this loop makes disappear the influence conflict.

Thomas's necessary condition for a system to exhibit multi-stationarity is the existence of a positive circuit, i.e., a simple oriented cycle such that the product of the signs of its edges is positive, in the DIG [40]. That condition has proven useful to reason about gene interaction networks and predict the possibilities of multi-stationarity, i.e. cell differentiation. However, Thomas's original condition provides no information in presence of reactions with two reactants, since a reaction like for instance $A + B \longrightarrow C$ immediately creates a positive circuit of negative influences between A and B in the associated SIG and DIG for any reasonable kinetics. This counter-example has been recently rule out in [38], where it is shown that Thomas's conditions can be made stronger for reactions models, by labeling the influence edges by the reactions they come from, and by restricting the analysis of circuits to circuits labeled by different reactions. With this stronger condition for multi-stationarity, the analysis of labeled circuits in the DIG of a reaction system does provide information on its capabilities of exhibiting multi-stationarity. Theorem 2.16 shows, perhaps surprisingly, that for well-formed reaction systems without conflicts, the DIG is essentially independent from the kinetics, and in fact identical to the SIG, which is easy to compute and can be used to perform multi-stationarity analysis by circuit analysis.
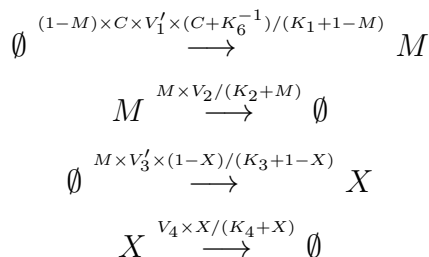
3. Reaction System Inference Algorithm

In this section we present an algorithm to infer a reaction system from an arbitrary ODE system, and study its properties. The algorithm proceeds in two steps: one first step for infering hidden molecules corresponding to linear invariants of the ODE system, and one second step for infering the reactions.

3.1. Motivating Example

As remarked in Section 2.1, any ODE model can be transcribed in a reaction system using artificial synthesis and degradation reactions for each molecular species, with the positive, respectively negative, terms of the differential equation for the variables as kinetic expressions. While preserving the ODE semantics and thus ODE simulations, such a transcription prevents the use of structural methods and stochastic simulations to analyze the system.

Such a transcription has nevertheless been used in biomodels.net to write the ODE model of [46] in SBML and create `BIOMD0000000008.xml`. This model adds a control mechanism to the cell-cycle model of Goldbeter et al. in [49] but with this transcription in SBML, the reaction graph is not even connected.

Here are some of the reactions of this model (after expansion of the macros used in the original writing) which illustrate the problem:
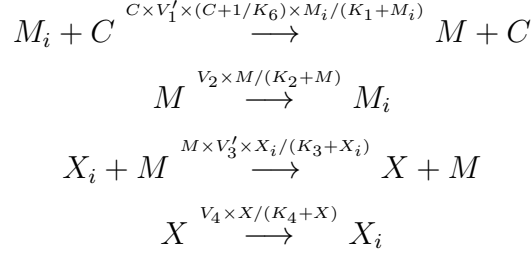
$$\emptyset \xrightarrow{(1-M)\times C\times V_1'\times(C+K_6^{-1})/(K_1+1-M)} M$$

$$M \xrightarrow{M\times V_2/(K_2+M)} \emptyset$$

$$\emptyset \xrightarrow{M\times V_3'\times(1-X)/(K_3+1-X)} X$$

$$X \xrightarrow{V_4\times X/(K_4+X)} \emptyset$$

One can notice that $\partial f_1/\partial C \neq 0$, where f_1 is the kinetic expression of the first reaction, but C is not a reactant nor an inhibitor. The model is therefore not well-formed.

One can also note that, though encoded in complicated MathML expressions, $1 - M$ (resp. $1 - X$) appears in the synthesis of M (resp. X) as a way to represent the inactive form of M (resp. X). Indeed, [49] states that *"(1 - M) thus represents the fraction of inactive (i.e., phosphorylated) cdc2 kinase, while (1 - X) represents the fraction of inactive (i.e., dephosphorylated) cyclin protease"*.

When applied to the ODE system associated to this model, the reaction system inference algorithm presented in the next two sections, infers two *hidden molecules* and the following

well-formed and strict reactions:

$$M_i + C \xrightarrow{C \times V_1' \times (C + 1/K_6) \times M_i/(K_1 + M_i)} M + C$$

$$M \xrightarrow{V_2 \times M/(K_2 + M)} M_i$$

$$X_i + M \xrightarrow{M \times V_3' \times X_i/(K_3 + X_i)} X + M$$

$$X \xrightarrow{V_4 \times X/(K_4 + X)} X_i$$

The two inactive forms are now explicitly represented by two inferred molecules, written M_i and X_i, and the actions of C on M and of M on X are properly transcribed. The reaction system inferred automatically from the ODE semantics is well-formed and strict, and in fact consistent with the graphical representation of the paper [49] where dashed arrows represent catalytic effects:
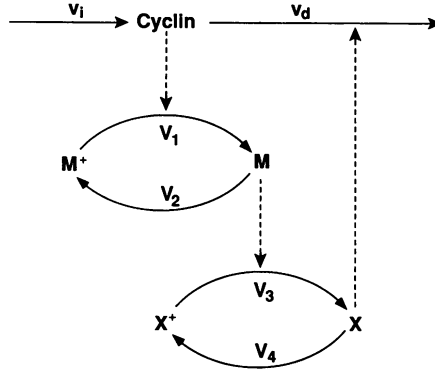
In that form, the inferred model is thus suitable for further structural analysis.

The following sections present the reaction system inference algorithm in two steps: first the algorithm for inferring hidden molecules corresponding, as above, to invariants, second the algorithm for inferring well-formed reactions whenever possible.

3.2. Inference Algorithm for Hidden Molecules

ODE models often contain algebraic invariants, i.e., algebraic equations relating variables of the model and that hold true in any solution of the ODE system. Among those, linear invariants $\sum \lambda_i x_i = \Lambda$, e.g. mass conservation invariants, or Petri-net place invariants, are an important particular case. A linear invariant can be used to simplify a model by eliminating one variable and replacing it with a linear expression. This may have several advantages, but when writing the model with reactions, such simplifications performed on the ODE system need be reversed in order to restore the correct structure of the reactions on eliminated molecular species, as shown for instance in the previous section with the inactive forms M_i of M, and X_i of X.

A preprocessor is first applied before the reaction inference algorithm, in order to reverse the elimination of linear invariants and infer hidden molecules. The expressions f for which new molecules are introduced need be chosen with care in order to avoid the introduction of useless variables. Restricting the search to expressions of the form $k - x$ or $k - x - y$ where k is a constant or parameter, and x and y are molecule concentrations, has proven useful in practice. This leads to

Algorithm 3.1. *(Hidden molecule inference)*

input: *ODE system O over variables $\{x_1, \ldots, x_s\}$,*
    1. *iteratively replace in O any expression of the form $-x + y$ by $y - x$,*
    2. *for each expression of the form $k - x - y$ in O where k is a numerical constant or a parameter, and x and y are variables,*
        (a) *introduce a new variable z with time derivative $\dot{z} = -\dot{x} - \dot{y}$, and functional dependency equation $z = k - x - y$,*
        (b) *substitute any occurrence of $k - x - y$ in O by z,*
        (c) *substitute any occurrence of $k + v - x - y$ in O for any expression v, by $v + z$,*
        (d) *substitute any occurrence of $k - x + w - y$ in O for any w, by $v + z$,*
    3. *for each expression $k - x$ appearing in O where k is a constant or a parameter and x a variable,*
        (a) *introduce a new variable z with time derivative $\dot{z} = -\dot{x}$ and functional dependency equation $z = k - x$,*
        (b) *substitute any occurrence of $k - x$ in O by z,*
        (c) *substitute any occurrence of $k + v - x$ in O for any expression v, by $z + v$,*
output: *ODE system O over variables $\{x_1, \ldots, x_s\}$ and hidden molecule variables $\{z_1, \ldots, z_k\}$, together with functional dependency equations $z_j = f_j(x_1, \ldots, x_s)$ for $1 \leq j \leq k$.*

Proposition 3.2 (Soundness). *Let O be an ODE system over variables $\{x_1, \ldots, x_s\}$. The ODEs computed by Algorithm 3.1 for the time derivatives of $x_1, \ldots, x_s$, are mathematically equivalent to the equations in O given the functional dependency equations $z_j = f_j(x_1, \ldots, x_s)$ for the hidden molecules.*

Proof. We prove that each step of the algorithm replaces equal by equal, and thus that the whole execution preserves the mathematical equivalence of the equations. First, step (1) is a purely syntactical transformation that does not change the ODE system O. Now note that all the other changes are of two forms. Either the introduction of a new variable z such that $\dot{z} = \sum \lambda_i \dot{x}_i$, together with the functional dependency equation $z = k + \sum \lambda_i x_i$, steps (2a) and (3a). Since the differential equation on z is indeed the time derivative of the definition of z, this does not change the equations on $\dot{x}_i$. Or the replacement of $k + \sum \lambda_i x_i$ by z, steps (2b-d) and (3b-c), which are equal from the previous definition of z. $\qquad\square$

3.3. Inference Algorithm for Reactions

The inference algorithm for reactions is based on a syntactical normal form for ODE systems which facilitates the recognition of common subterms in the equations.

We consider ODEs and kinetic laws written in MathML as terms with mathematical operations and functions (e.g.$+$, $-$, $/$, $\times$, etc.), constants of $\mathbb{R}$ and variables representing species concentrations and parameters. It is beyond the scope of this paper to precisely describe the mathematical expressions allowed and the symbolic computation performed. However, let us call *non-decomposable* a term that:

- its functor (top function symbol) is neither $+$ nor $-$;

- cannot be reduced at top-level by the algebraic laws of distributivity of the product and division on addition and subtraction, e.g. if its functor is $\times$ (resp. $/$) then the arguments (resp. the numerator) are not sums.

Definition 3.3. *A reaction $r \mathbin{/} m \overset{f}{\longrightarrow} p$ over molecular species $\{x_1, \ldots, x_s\}$ is* non-decomposable *if f is syntactically a non-decomposable term.*

Definition 3.4. *A mathematical expression is in* additive normal form *if it is of the form $\sum_{i=1}^{k} c_i \times t_i$ where c_i are integers and t_i are distinct non-decomposable terms without integer coefficients.*

An ODE system is in additive normal form if each equation is in additive normal form, i.e. if it is of the form

$$\dot{x}_i = \sum_{j=1}^{l} c_{i,j} \times t_j, \ 1 \le i \le s$$

where l is the number of non-decomposable terms t_j in the system.

Additive normal forms are not unique, but any ODE system can be written in additive normal form through standard algebraic transformations (such as the distributivity of $\times$ over $+$). The non-decomposability condition excludes the composition of several reactions in a single one with a sum as kinetic expression. In particular, we have:

Proposition 3.5. *Any non-decomposable well-founded reaction system, such that its ODE semantics is a polynomial ODE system, is strict.*

Proof. First notice that a polynomial kinetics once in additive normal form results in a sum of monomials as non-decomposable terms. Now, from the second condition of well-formedness in Definition 2.5, for each reaction (r, m, p, f) we have $r(x_j) > 0$ implies $\exists \vec{x}, \partial f / \partial x_j(\vec{x}) > 0$, but since f is a monomial, this implies that f has degree at least 1 in x_j, and therefore that $f(x_1, \ldots, x_s) = 0$ when $x_j = 0$, i.e., (r, m, p, f) is strict. □

Now, given an ODE system in additive normal form, the following algorithm can infer an equivalent reaction system by sorting the terms of the ODEs, and creating one reaction for each term (formalized in Prop. 3.9 below). This algorithms requires checking

the sign of a partial derivative, and as described in Section 4.1, such checks can be arbitrarily difficult for arbitrary mathematical expressions, but can be over-approximated. We thus assume given a test program (exact or not) for testing the sign of partial derivatives: `partial_has_pos_val`(f, x) that answers if yes or no the partial derivative of the function f with respect to variable x takes a non-negative value for some input in $\mathbb{R}^s$, and such that $\exists \vec{y} \ \partial f / \partial x(\vec{y}) > 0 \Rightarrow$ `partial_has_pos_val`(f, x). For computability reasons, the reverse implication is not required. These tests are used in Steps 4(c) and 4(d).

Algorithm 3.6. *(Reaction inference)*

input: *ODE system O over variables for molecular concentrations,*
   `partial_has_pos_val` *test*
   1. *rewrite O into additive normal form*
   2. *compute the set $\mathcal{T}$ of all terms appearing in O*
   3. *let $R := \emptyset$*
   4. *for each non-decomposable term t in $\mathcal{T}$,*
      (a) *let $r := \emptyset$, $p := \emptyset$, $m := \emptyset$*
      (b) *for each variable x where t occurs with integer coefficient c in $\dot{x}$ in O,*
          i. *if $c < 0$ then $r(x) := -c$,*
          ii. *if $c > 0$ then $p(x) := c$,*
      (c) *for each variable x such that $r(x) = 0$ and `partial_has_pos_val`(t, x),*
          i. *$r(x) := 1$,*
          ii. *$p(x) := p(x) + 1$,*
      (d) *for each variable x such that `partial_has_pos_val`$(-t, x)$,*
          i. *$m(x) := 1$,*
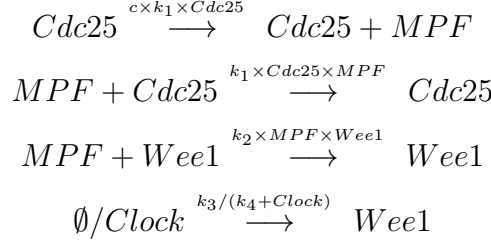   (e) *$R := R \cup \{r \ / \ m \xrightarrow{t} p\}$,*
output: *reaction system R.*

Example 3.7. *The model of three reactions of Example 2.10 has one invariant: $pMPF + MPF$ is indeed a constant c (the sum of initial values of $pMPF$ and MPF) since $p\dot{M}PF + \dot{M}PF = 0$. One variable, e.g. $pMPF$, can thus be eliminated and replaced by $c - MPF$. This yields the following ODE system, where all k_i are positive:*

$$
\begin{aligned}
\dot{MPF} &= k_1 \times (c - MPF) \times Cdc25 - k_2 \times MPF \times Wee1 \\
\dot{Wee1} &= k_3 / (k_4 + Clock) \\
\dot{Cdc25} &= 0 \\
\dot{Clock} &= 0
\end{aligned}
$$

When applied to this system, using the test for partial derivatives described in Section 4.1,
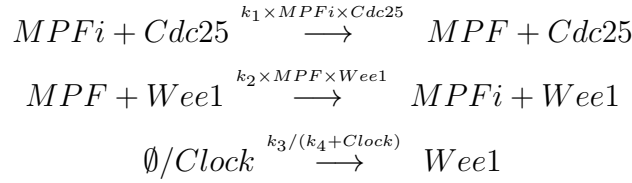
Algorithm 3.6 infers the following reactions:

$$Cdc25 \xrightarrow{c \times k_1 \times Cdc25} Cdc25 + MPF$$

$$MPF + Cdc25 \xrightarrow{k_1 \times Cdc25 \times MPF} Cdc25$$

$$MPF + Wee1 \xrightarrow{k_2 \times MPF \times Wee1} Wee1$$

$$\emptyset/Clock \xrightarrow{k_3/(k_4+Clock)} Wee1$$

However, by applying first the hidden molecule inference Algorithm 3.1, a hidden molecular species $MPFi$ is introduced for the expression $c - MPF$. This hidden molecule corresponds to the linear invariant $MPFi + MPF = c$. We have

$$\dot{MPF_i} \quad = \quad -k_1 \times MPF_i \times Cdc25 + k_2 \times MPF \times Wee1$$

and when applied to this ODE system after the preprocessing step, Algorithm 3.6 now computes the correct reactions:

$$MPFi + Cdc25 \xrightarrow{k_1 \times MPFi \times Cdc25} MPF + Cdc25$$

$$MPF + Wee1 \xrightarrow{k_2 \times MPF \times Wee1} MPFi + Wee1$$

$$\emptyset/Clock \xrightarrow{k_3/(k_4+Clock)} Wee1$$

By counting the loops, one can easily check

Proposition 3.8 (Time complexity)**.** *On an ODE system O in additive normal form, Algorithm 3.6 computes a reaction system in time $O(n \times t)$, where n is the number of variables and t is the number of non-decomposable terms in O.*

By executing symbolically the algorithm, one can similarly check that the result is characterized in mathematical terms by

Proposition 3.9 (Inferred reactions)**.** *Given an ODE system in additive normal form with appearing terms $\mathcal{T} = \{f_1, \ldots, f_t\}$: $\dot{x}_i = \sum_{u=1}^{t} c_{i,u} \times f_u$ for $1 \leq i \leq s$. The reaction system inferred by Algorithm 3.6 is the set of non-decomposable reactions*

$$\{ r_u/m_u \xrightarrow{f_u} p_u \}_{1 \leq u \leq t}$$

$$where \quad r_u = \sum_{\{i \ | \ c_{i,u} < 0\}} (-c_{i,u}) \times x_i + \sum_{\{i \ | \ c_{i,u} \geq 0, \ \texttt{partial\_has\_pos\_val}(f_u, x_i)\}} x_i,$$

$$p_u = \sum_{\{i \ | \ c_{i,u} > 0\}} c_{i,u} \times x_i + \sum_{\{i \ | \ c_{i,u} \geq 0, \ \texttt{partial\_has\_pos\_val}(f_u, x_i)\}} x_i$$

and m_u is the set of variables x such that $\texttt{partial\_has\_pos\_val}(-f_u, x)$.

Theorem 3.10 (Soundness)**.** *The ODE semantics of the reaction system inferred by Algorithm 3.6 from an ODE system O is equal to O.*

Proof. Let us suppose without loss of generality that $O = \{\dot{x}_i = \sum_{u=1}^{t} c_{i,u} \times f_u \mid 1 \leq i \leq s\}$ is in additive normal form. The inferred reaction system is the set $\{r_u/m_u \xrightarrow{f_u} p_u\}_{1 \leq u \leq t}$ where $r_u = \sum_{\{i \mid c_{i,u} < 0\}} (-c_{i,u}) \times x_i + \sum_{\{i \mid c_{i,u} \geq 0, \text{ partial\_has\_pos\_val}(f_u, x_i)\}} x_i$, $p_u = \sum_{\{i \mid c_{i,u} > 0\}} c_{i,u} \times x_i + \sum_{\{i \mid c_{i,u} \geq 0, \text{ partial\_has\_pos\_val}(f_u, x_i)\}} x_i$, and m_u is the set of variables y such that $\texttt{partial\_has\_pos\_val}(-f_u, x_i)$.

The ODE system associated to these reactions is thus
$\{\dot{x}_i = \sum_{u=1}^{t} (p_u(x_i) - r_u(x_i)) \times f_u\}_{1 \leq i \leq s} = \{\dot{x}_i = \sum_{u=1}^{t} c_{i,u} \times f_u\}_{1 \leq i \leq s} = O$.

Note that it does not depend on the test $\texttt{partial\_has\_pos\_val}$. $\square$

Algorithm 3.6 always computes a non-decomposable reaction system with an equivalent associated ODE system but this reaction system may not be well-formed. In particular, step 3b adds a variable x to the reactants of the reactions even if x does not appear in the kinetic expression f of the reaction. Therefore the algorithm may infer reactions with reactants that do not occur in the kinetic expression, as required for instance by Example 2.9.

We can measure the completeness of the method by showing that, at least, if we start from a well-formed reaction model, generate the ODE semantics, and from the ODE system solely, infer back a reaction model, the algorithm does infer a well-formed reaction model.

First, it is clear that the algorithm infers non-decomposable kinetics (Prop. 3.9) in which any variable appearing in the kinetics appears in the reaction as either reactant (step 4b, or step 4c for catalysts), inhibitor (step 4d) or both:

Proposition 3.11. *The reactions inferred by Algorithm 3.6 contain no reaction with a molecular species x appearing in the kinetic expression f with $\partial f / \partial x \neq 0$, and not appearing as a reactant or inhibitor.*

This proposition remains true even if the sets of variables for which the partial derivatives are positive (4c) or negative (4d) are over-approximated. However, for completeness an exact test is necessary.

Theorem 3.12 (Weak completeness)**.** *When applied to the ODE semantics of a non-decomposable well-formed reaction system such that $\forall 1 \leq i \leq n, 1 \leq j \leq s \; \partial f_i / \partial x_j > 0 \Leftrightarrow \texttt{partial\_has\_pos\_val}(f_i, x_j)$, Algorithm 3.6 does infer a non-decomposable well-formed reaction system. Furthermore, if the ODE system is polynomial, the inferred model is strict.*

Proof. Let us consider the ODEs associated to a well-formed non-decomposable reaction system $R = \{ r_i \; / \; m_i \xrightarrow{f_i} p_i \}_{i=1,\dots,n}$. The ODE system is of the form $O = \{\dot{x}_j = \sum_{i=1}^{n} (p_i(x_j) - r_i(x_j)) \times f_i \mid 1 \leq i \leq m\}$ which is an additive normal form after evaluation of the integers $p_i(x_j) - r_i(x_j)$. By Prop. 3.9, the inferred reaction system is $\{ r_i'/m_i' \xrightarrow{f_i} p_i'\}_{1 \leq i \leq n}$ where f_i is non-decomposable by hypothesis,
$r_i' = \sum_{\{j \mid p_i(x_j) < r_i(x_j)\}} (r_i(x_j) - p_i(x_j)) \times x_j + \sum_{\{j \mid p_i(x_j) \geq r_i(x_j), \; \partial f_i / \partial x_j > 0\}} x_j$
$p_i' = \sum_{\{j \mid p_i(x_j) > r_i(x_j)\}} (p_i(x_j) - r_i(x_j)) \times x_j + \sum_{\{j \mid p_i(x_j) \geq r_i(x_j), \; \partial f_i / \partial x_j > 0\}} x_j,$

17

and $m_i' = m_i$.

Now for any variable x_j, we have $x_j \in r_i'$ if and only if $x_j \in r_i$ since either $p_i(x_j) < r_i(x_j)$ or $\partial f_i / \partial x_j > 0$. Similarly $x_j \in p_i'$ if and only if $x_j \in p_i$ since either $p_i(x_j) > r_i(x_j)$ or $p_i(x_j) = r_i(x_j)$ and $\partial f_i / \partial x_j > 0$, These equalities between the sets (not multisets) of reactants, products and inhibitors suffice to show the well-formedness of the inferred reactions.

Strictness in the polynomial case follows from Prop. 3.5. $\qquad\qquad\square$

Since we do not restrict ourselves to Mass Action kinetics, our algorithm may well infer reactions with other kinetic expressions in cases where purely Mass Action reactions were possible. This is an important difference between our algorithm and the previous algorithms, which are restricted to Mass Action kinetics [30, 31, 32]. Furthermore, even if we restrict to polynomial ODEs and Mass Action kinetics for reaction, further conditions are necessary to grant the unicity of the solution [32].
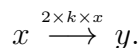
Example 3.13. *For instance, given the ODE system*
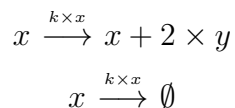
$$\dot{x} = -2kx = -\dot{y},$$

our algorithm infers the reaction

$$2 \times x \xrightarrow{k \times x} 2 \times y,$$
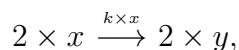
whereas a Mass Action kinetic reaction model for this system is

$$x \xrightarrow{2 \times k \times x} y.$$

Furthermore, another Mass Action reaction system exists for this ODE system:

$$x \xrightarrow{k \times x} x + 2 \times y$$

$$x \xrightarrow{k \times x} \emptyset$$

4. Evaluation Results on biomodels.net

The ability to infer a reaction system from ODEs can be turned into some automatic curation algorithm, as was done in Theorem 3.12, by inferring the reactions from the ODE semantics of a starting reaction system. In this section, we evaluate this form of curation on repository of structured models.

4.1. Computability Issues

Since, like in SBML, we allow arbitrary mathematical expression for kinetic expressions, checking the well-formed conditions may raise arbitrary difficult symbolic computation problems. These conditions can be checked however by doing some approximations.

In our implementation in Biocham [1, 2][2], the `partial_has_pos_val` proceeds as follows: the kinetic expressions are first normalized as if they were polynomials, stopping when a non-polynomial operator (anything else than $+$, $-$ and $\times$) is found. For the polynomials, the exact computation of the sign of any partial derivative is easy. For the other terms, either they are recognized as a standard kinetics (like Hill functions) and once again the exact sign is extracted, or they are considered unknown and for any variable appearing we will assume that it is possible that $\partial f / \partial x$ becomes positive for some values, and negative for some values. This is a conservative over-approximation.

With these provisions, different syntactical conditions may indicate that a reaction is not well-formed. The conditions for a reaction to be ill-formed can be classified into three categories:

1. "K not R" indicates that the concentration of a compound appears in the kinetic law of a reaction, but this compound is neither a reactant nor an inhibitor of the reaction;
2. "R not K" indicates that some compound is marked as reactant or inhibitor in a reaction, but does not appear in the kinetic expression;
3. "Negative" indicates that a kinetic expression may be negative with non-negative concentration values.

Indeed, in a well-formed reaction with kinetic expression f, if a species x is neither a reactant nor an inhibitor, then $\partial f / \partial x = 0$, hence x should not appear in the kinetic expression f. Similarly, if a species is a reactant or an inhibitor, then $\partial f / \partial x \neq 0$, so x should appear in f. Moreover, f should be non-negative.

These ill-formedness conditions are checked in Biocham using the previous approximations. They correspond to the warning messages that Biocham can raise when loading a reaction system.

4.2. Global analysis

The 424 models from the curated branch of the latest version (release 24) of the `biomodels.net` repository [35] were used as benchmark to test our reaction system inference algorithm, and compare the results with the original writing of the models in SBML. Out of those 424 models only 361 define *reactions* with proper *kineticLaws*. The other ones only describe systems through events and rules, or with no kinetic information, and thus have no ODE semantics.

Our curation algorithm reads the SBML model, extracts the corresponding ODE system and infers from it a new reaction system.

Table 4.2 summarizes the result of the procedure, as detected by Biocham warnings. Over the 361 reaction systems of the original curated part of biomodels.net with ODE semantics, our algorithm reveals *hidden molecules* in 58 models, 173 models with "K not R" warning, 123 models with "R not K" warning and 157 models with "negative kinetics" warning. Our algorithm is able to automatically curate the writing of these models with reactions by reducing the number of non well-formed models with a warning by more than the half, from 65% to 29%.

---

[2]`http://lifeware.inria.fr/biocham/`

	"K not R"	"R not K"	"Negative"	Any warning
Original	173	123	157	234 (64.81 %)
Inferred	0	67	70	103 (28.53 %)

Table 1: Number of models having a "K not R", "R not K", or "negative kinetics" warning among the original 361 models of the curated part of biomodels.net, and among the reaction systems automatically inferred from their ODE semantics. "Any warning" reflects model for which there was at least one of the three warning.
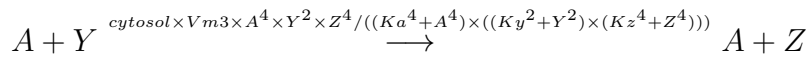
As predicted by Proposition 3.11, the Algorithm 3.6 completely removes the "K not R" warnings. For the two other warnings, since the algorithm focuses on *non-decomposable* kinetics, it results in curated models quite close to the original ones, but does not tackle thoroughly the case of reactions with rates independent of some reactant, for the reasons illustrated in Example 2.9 of for any other reason. Therefore, 103 over 361 models remain with a non well-formedness warning.

4.3. Model inconsistencies studied in [19]

In [19], the authors also scan the biomodels.net repository and report finding 5 inconsistencies: models 44, 93, 94, 143 and 151. Their diagnostics is as follows, some reaction fluxes become negative during the simulations of those models because of missing reversibility indications in models 93, 94 and 143. In the two first cases they report that adding the reverse reactions makes the models consistent, whereas for 143 it is also necessary to change some kinetic law. For model 151 they report a "missing step", but since the opposite reaction is part of the model, once again this amounts to adding a reverse reaction to an existing one. Finally, for model 44 they describe that the issue is that some kinetic expression does not depend on one of the reactants of the reaction, making it possible for that reactant's concentration to become negative.

For models 93, 94 and 151, which indeed are flagged by the "Negative" warning, our algorithm correctly adds the missing reverse reactions, directly from the kinetic expressions. The models automatically curated this way do not raise any warning at the end.

For model 44, the automatic curation allows us to get rid of a "K not R" warning by transforming the reaction $v3$

$$A + Y \xrightarrow{cytosol \times Vm3 \times A^4 \times Y^2 \times Z^4 / ((Ka^4 + A^4) \times ((Ky^2 + Y^2) \times (Kz^4 + Z^4)))} A + Z$$

into

$$Z + A + Y \longrightarrow 2 \times Z + A$$

with the same kinetics.

However, as expected, the "R not K" warning identified by Kaleta *et al.* remains, the obtained model is still not well-formed. The same happens with model 143 where indeed a "R not K" warning remains after automatic curation, in accordance with the earlier results.

5. Conclusion

We have described an algorithm for trying to infer a meaningful reaction system from a system of ordinary differential equations. This algorithm is based on a general consistency condition between the kinetic expression and the structure of a reaction in terms of its reactants, products and inhibitors.

We have shown some general properties enjoyed by the influence graph of the Jacobian sign matrix associated to such well-formed reaction systems. These theoretical results militate for distinguishing between catalysts and inhibitors in the modifiers of a reaction, and for using structural analysis methods before fixing parameter values and going to simulations.

We have also evaluated the capability of our reaction inference algorithm to automatically curate the writing of ODE models with reactions by applying it to the ODE models generated from the SBML models of the curated part of `biomodels.net`. In particular, we have shown that the inference of well-formed reactions from the ODEs, combined with the inference of hidden molecules corresponding to linear invariants, is sufficient to automatically curate the writing of some ODE models of the cell cycle with consistent reactions. On the whole curated part of the `biomodels.net` repository, we have shown that our automatic curation method significantly improves the writing of the models with reactions by reducing the number of non well-formed reaction systems from 65% to 29%.

Although the primary concern of SBML is to provide a common format for exchanging models and doing simulations, we believe that stronger consistency conditions should be enforced in SBML to perform structural analyses, and that the strict well-formedness conditions presented in this paper should be verified by non reversible reactions.

References

[1] F. Fages, S. Soliman, Formal cell biology in BIOCHAM, in: M. Bernardo, P. Degano, G. Zavattaro (Eds.), 8th Int. School on Formal Methods for the Design of Computer, Communication and Software Systems: Computational Systems Biology SFM'08, Vol. 5016 of Lecture Notes in Computer Science, Springer-Verlag, Bertinoro, Italy, 2008, pp. 54–80. `doi:10.1007/978-3-540-68894-5_3`.

[2] 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`.

[3] R. Thomas, A.-M. Gathoye, L. Lambert, A complex control circuit : regulation of immunity in temperate bacteriophages, European Journal of Biochemistry 71 (1) (1976) 211–227.

[4] R. Thomas, On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations, Springer Ser. Synergetics 9 (1981) 180–193.

[5] R. Thomas, Regulatory networks seen as asynchronous automata : a logical description, Journal of Theoretical Biology 153 (1991) 1–23.

[6] S. Eker, M. Knapp, K. Laderoute, P. Lincoln, J. Meseguer, M. K. Sönmez, Pathway logic: Symbolic analysis of biological signaling, in: Proceedings of the seventh Pacific Symposium on Biocomputing, 2002, pp. 400–412.

[7] G. Bernot, J.-P. Comet, A. Richard, J. Guespin, A fruitful application of formal methods to biological regulatory networks: Extending thomas' asynchronous logical approach with temporal logic, Journal of Theoretical Biology 229 (3) (2004) 339–347.

[8] N. Chabrier-Rivier, M. Chiaverini, V. Danos, F. Fages, V. Schächter, Modeling and querying biochemical interaction networks, Theoretical Computer Science 325 (1) (2004) 25–44.

[9] F. Fages, S. Soliman, N. Chabrier-Rivier, Modelling and querying interaction networks in the biochemical abstract machine BIOCHAM, Journal of Biological Physics and Chemistry 4 (2) (2004) 64–73.

[10] L. Calzone, N. Chabrier-Rivier, F. Fages, S. Soliman, Machine learning biochemical networks from temporal logic properties, in: G. Plotkin (Ed.), Transactions on Computational Systems Biology VI, Vol. 4220 of Lecture Notes in BioInformatics, Springer-Verlag, 2006, pp. 68–94, cMSB'05 Special Issue. `doi:10.1007/11880646_4`.

[11] V. N. Reddy, M. L. Mavrovouniotis, M. N. Liebman, Petri net representations in metabolic pathways, in: L. Hunter, D. B. Searls, J. W. Shavlik (Eds.), Proceedings of the 1st International Conference on Intelligent Systems for Molecular Biology (ISMB), AAAI Press, 1993, pp. 328–336.

[12] I. Zevedei-Oancea, S. Schuster, Topological analysis of metabolic networks based on petri net theory, In Silico Biology 3 (29).

[13] D. Angeli, P. D. Leenheer, E. D. Sontag, A petri net approach to persistence analysis in chemical reaction networks, in: Biology and Control Theory: Current Challenges, Vol. 357 of LNCIS, Springer-Verlag, 2007, pp. 181–216.

[14] C. Chaouiya, E. Remy, D. Thieffry, Petri net modelling of biological regulatory networks, Journal of Discrete Algorithms 6 (2) (2008) 165–177. `doi:10.1016/j.jda.2007.06.003`.

[15] C. Rohr, W. Marwan, M. Heiner, Snoopy - a unifying petri net framework to investigate biomolecular networks, Bioinformatics 26 (7) (2010) 974–975. `doi:10.1093/bioinformatics/btq050`.

[16] S. Soliman, Invariants and other structural properties of biochemical models as a constraint satisfaction problem, Algorithms for Molecular Biology 7 (15). `doi:10.1186/1748-7188-7-15`.

[17] S. Gay, S. Soliman, F. Fages, A graphical method for reducing and relating models in systems biology, Bioinformatics 26 (18) (2010) i575–i581, special issue ECCB'10. `doi:10.1093/bioinformatics/btq388`.

[18] S. Gay, F. Fages, T. Martinez, S. Soliman, C. Solnon, On the subgraph epimorphism problem, Discrete Applied Mathematics 162 (2014) 214–228. `doi:10.1016/j.dam.2013.08.008`.

[19] C. Kaleta, S. Richter, P. Dittrich, Using chemical organization theory for model checking, Bioinformatics 25 (15) (2009) 1915–1922. `doi:10.1093/bioinformatics/btp332`.

[20] S. Schuster, D. A. Fell, T. Dandekar, A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks, Nature Biotechnology 18 (2002) 326–332.

[21] A. Varma, B. Palsson, Metabolic flux balancing: basic concepts, scientific and practical use, Nature Biotechnology 12 (10) (1994) 994–998.

[22] P. Dittrich, P. di Fenizio, Chemical organisation theory, Bulletin of Mathematical Biology 69 (4) (2007) 1199–1231. `doi:10.1007/s11538-006-9130-8`.

[23] M. Feinberg, Mathematical aspects of mass action kinetics, in: L. Lapidus, N. R. Amundson (Eds.), Chemical Reactor Theory: A Review, Prentice-Hall, 1977, Ch. 1, pp. 1–78.

[24] G. Shinar, M. Feinberg, Structural sources of robustness in biochemical reaction networks, Science 327 (5971) (2010) 1389–1391. `doi:10.1126/science.1183372`.

[25] G. Koh, H. Teong, M.-V. Clement, D. Hsu, P. Thiagarajan, A decompositional approach to parameter estimation in pathway modeling: a case study of the akt and mapk pathways and their crosstalk, Bioinformatics 22 (14) (2006) e271–e280.

[26] S. Soliman, Finding minimal P/T-invariants as a CSP, in: Proceedings of the fourth Workshop on Constraint Based Methods for Bioinformatics WCB'08, associated to CPAIOR'08, 2008.

[27] E. Grafahrend-Belau, F. Schreiber, M. Heiner, A. Sackmann, B. H. Junker, S. Grunwald, A. Speer, K. Winder, I. Koch, Modularization of biochemical networks based on a classification of petri net by T-invariants, BMC Bioinformatics 9 (90).

[28] F. Nabli, S. Soliman, Steady-state solution of biochemical systems, beyond S-Systems via T-invariants,

in: P. Quaglia (Ed.), CMSB'10: Proceedings of the 8th International Conference on Computational Methods in Systems Biology, CoSBi, ACM, 2010, pp. 14–22. `doi:10.1145/1839764.1839768`.

[29] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions, Journal of Physical Chemistry 81 (25) (1977) 2340–2361.

[30] V. Hárs, J. Tóth, On the inverse problem of reaction kinetics, in: M. Farkas (Ed.), Colloquia Mathematica Societatis János Bolyai, Vol. 30 of Qualitative Theory of Differential Equations, 1979, pp. 363–379.

[31] G. Szederkényi, J. R. Banga, A. A. Alonso, Inference of complex biological networks: distinguishability issues and optimization-based solutions., BMC systems biology 5 (1) (2011) 177+. `doi:10.1186/1752-0509-5-177`.

[32] S. Soliman, M. Heiner, A unique transformation from ordinary differential equations to reaction networks, PLoS One 5 (12) (2010) e14284. `doi:10.1371/journal.pone.0014284`.

[33] M. Hucka, et al., The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models, Bioinformatics 19 (4) (2003) 524–531.

[34] M. Hucka, S. Hoops, S. M. Keating, L. N. Nicolas, S. Sahle, D. Wilkinson, Systems biology markup language (SBML) level 2: Structures and facilities for model definitions, Nature Precedings`doi:10.1038/npre.2008.2715.1`.

[35] N. le Novère, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, J. L. Snoep, M. Hucka, BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems, Nucleic Acid Research 1 (34) (2006) D689–D691.

[36] F. Fages, S. Soliman, From reaction models to influence graphs and back: a theorem, in: Proceedings of Formal Methods in Systems Biology FMSB'08, no. 5054 in Lecture Notes in Computer Science, Springer-Verlag, 2008. `doi:10.1007/978-3-540-68413-8_7`.

[37] F. Fages, S. Soliman, Abstract interpretation and types for systems biology, Theoretical Computer Science 403 (1) (2008) 52–70. `doi:10.1016/j.tcs.2008.04.024`.

[38] S. Soliman, A stronger necessary condition for the multistationarity of chemical reaction networks, Bulletin of Mathematical Biology 75 (11) (2013) 2289–2303. `doi:10.1007/s11538-013-9893-7`.

[39] C. Soulé, Mathematical approaches to differentiation and gene regulation, C. R. Biologies 329 (2006) 13–20.

[40] C. Soulé, Graphic requirements for multistationarity, ComplexUs 1 (2003) 123–133.

[41] E. H. Snoussi, Necessary conditions for multistationarity and stable periodicity, Journal of Biological Systems 6 (1998) 3–9. `doi:10.1142/S0218339098000042`.

[42] J.-L. Gouzé, Positive and negative circuits in dynamical systems, Journal of Biological Systems 6 (1998) 11–15.

[43] M. Kaufman, C. Soulé, R. Thomas, A new necessary condition on interaction graphs for multistationarity, Journal of Theoretical Biology 248 (2007) 675–685.

[44] K. W. Kohn, Molecular interaction map of the mammalian cell cycle control and DNA repair systems, Molecular Biology of the Cell 10 (8) (1999) 2703–2734.

[45] M. Katsumata, Graphic representation of botts-morales equation for enzyme-substrate-modifier system, Journal of Theoretical Biology 36 (2) (1972) 327–338.

[46] T. S. Gardner, M. Dolnik, J. J. Collins, A theory for controlling cell cycle dynamics using a reversibly binding inhibitor., Proceedings of the National Academy of Sciences of the United States of America 95 (24) (1998) 14190–14195. `doi:10.1073/pnas.95.24.14190`.

[47] T. Matsuo, S. Yamaguchi, S. Mitsui, A. Emi, F. Shimoda, H. Okamura, Control mechanism of the circadian clock for timing of cell division in vivo, Science 302 (5643) (2003) 255–259.

[48] J. J. Tyson, Modeling the cell division cycle: cdc2 and cyclin interactions, Proceedings of the National Academy of Sciences 88 (16) (1991) 7328–7332.

[49] A. Goldbeter, A minimal cascade model for the mitotic oscillator involving cyclin and cdc2 kinase, PNAS 88 (20) (1991) 9107–9111.