

## Systems biology

## Using chemical organization theory for model checking

Christoph Kaleta<sup>1,2,3,†</sup>, Stephan Richter<sup>1,2,†</sup> and Peter Dittrich<sup>1,2,\*</sup><sup>1</sup>Bio Systems Analysis Group, Jena Centre for Bioinformatics (JCB), <sup>2</sup>Department of Mathematics and Computer Science and <sup>3</sup>Department of Bioinformatics, Friedrich Schiller University Jena, Ernst-Abbe-Platz 2, D-07743 Jena, Germany

Received on July 22, 2008; revised on May 4, 2009; accepted on May 17, 2009

Advance Access publication May 25, 2009

Associate Editor: Olga Troyanskaya

## ABSTRACT

**Motivation:** The increasing number and complexity of biomodels makes automatic procedures for checking the models' properties and quality necessary. Approaches like elementary mode analysis, flux balance analysis, deficiency analysis and chemical organization theory (OT) require only the stoichiometric structure of the reaction network for derivation of valuable information. In formalisms like Systems Biology Markup Language (SBML), however, information about the stoichiometric coefficients required for an analysis of chemical organizations can be hidden in kinetic laws.

**Results:** First, we introduce an algorithm that uncovers stoichiometric information that might be hidden in the kinetic laws of a reaction network. This allows us to apply OT to SBML models using modifiers. Second, using the new algorithm, we performed a large-scale analysis of the 185 models contained in the manually curated BioModels Database. We found that for 41 models (22%) the set of organizations changes when modifiers are considered correctly. We discuss one of these models in detail (BIOMD149, a combined model of the ERK- and Wnt-signaling pathways), whose set of organizations drastically changes when modifiers are considered. Third, we found inconsistencies in 5 models (3%) and identified their characteristics. Compared with flux-based methods, OT is able to identify those species and reactions more accurately [in 26 cases (14%)] that can be present in a long-term simulation of the model. We conclude that our approach is a valuable tool that helps to improve the consistency of biomodels and their repositories.

**Availability:** All data and a JAVA applet to check SBML-models is available from

<http://www.minet.uni-jena.de/csb/prj/ot/tools>

**Contact:** dittrich@minet.uni-jena.de

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

## 1 INTRODUCTION

Reaction networks are widely used to model biological systems at various levels, including the molecular level (Le Novère *et al.*, 2006), the cellular level (Wodarz and Nowak, 1999), the ecological scale (Hofbauer and Sigmund, 1998) and the level of social

interactions (Dittrich *et al.*, 2003). Because of the growing number of models, methods for their verification are needed. Approaches that can be used for this purpose are flux balance analysis (FBA; Varma and Palsson, 1994), elementary mode analysis (Schuster *et al.*, 1999), extreme pathway analysis (Schilling *et al.*, 2000) and chemical organization theory (OT; Dittrich and Speroni di Fenizio, 2007). These methods have in common that they allow deriving constraints to dynamic behavior from the stoichiometric structure of a network. This stoichiometric structure is defined by the number of molecules of educts as well as the products each reaction consumes and produces. Thus, stoichiometry-based methods do not require a precise knowledge of the underlying reaction kinetics, which are often partly or totally unknown.

Here, we will use OT. An important property of chemical organizations is that every steady state and growth state<sup>1</sup> of a network corresponds to a chemical organization (Dittrich and Speroni di Fenizio, 2007, and Supplementary Material). These states we call the limit behavior of a model. However, this property is fulfilled only if a reaction network meets a condition formulated by Feinberg and Horn (1974): each reaction has a non-zero flux if and only if all of its educts have a positive concentration. Using this property, OT has already been applied to the prediction of growth phenotypes (Centler *et al.*, 2007) and the outcome of knockout experiments (Kaleta *et al.*, 2008), as well as in the design of chemical programs to solve NP-complete problems (Matsumaru *et al.*, 2007).

In a recent work, we used OT to assess the quality of a genome-scale reaction network of *Escherichia coli* by identifying species and reactions that could not be present in the limit behavior of the model during simulation (Centler *et al.*, 2008). We concluded that these species and reactions hint at missing knowledge as they were mostly part of pathways starting from or ending in dead-end species. Here, we want to extend this approach in two directions. First, we present a method for more accurately predicting the limit behavior of a reaction network if information on reactions kinetics is available. If modeled in Systems Biology Markup Language (SBML, Hucka *et al.*, 2003), the velocity of a reaction depends on the concentration of its educts, products and modifiers. A modifier is a species whose concentration affects the reaction velocity, but whose concentration itself is not changed by this reaction. Some modifiers, as for example catalysts or activators, are required to be present

\*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

<sup>1</sup>As growth state we define a situation where some species accumulate. An example is exponential growth in which, for instance, the overall amount of DNA increases given that there is a continuous supply (inflow) of nutrients.

for a non-zero reaction velocity. However, if we want a reaction to fulfill the Feinberg condition, such modifiers need to be added on its educt and product sides. Hence, since information necessary for the analysis using OT can be hidden in the kinetic laws, we present an algorithm for extracting this information. Second, using this approach, we demonstrate how knowledge of the organizational structure of a reaction network and thus of its limit behavior can help to uncover modeling inconsistencies. These inconsistencies are represented by species as well as by reactions that belong to no organization, indicating either incomplete knowledge, compounds missing from the specified growth media or modeling errors.

This work is structured as follows. In Section 2, we give a short outline of OT and present an algorithm that modifies the stoichiometric structure in a reaction network such that the Feinberg condition is fulfilled. We use this algorithm in Section 3 to demonstrate how these modifications affect the organizational structure of a model of the extracellular signal related-kinase (ERK)/Wnt-signaling pathway. In Section 4, we use our approach to find inconsistencies in a large-scale analysis of the models of the BioModels Database (Le Novère *et al.*, 2006) and compare our results with those obtained by other stoichiometric analysis techniques. Finally, we conclude in Section 5.

## 2 METHODS

### 2.1 Chemical OT

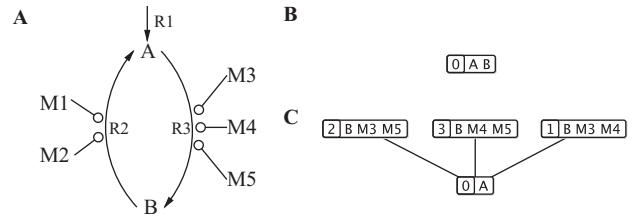
We define a reaction network  $(\mathcal{M}, \mathcal{R})$  by a set of molecular species  $\mathcal{M}$  and a set of reaction rules  $\mathcal{R}$ . A reaction rule  $\rho \in \mathcal{R}$  is defined by the stoichiometric coefficients  $l_{i,\rho}$  and  $r_{i,\rho}$  denoting the left- and right-hand sides of a reaction rule, respectively. Given a reaction rule  $\rho \in \mathcal{R}$ , we denote the set of reactant species and set of product species by  $\text{LHS}(\rho) := \{i \in \mathcal{M} | l_{i,\rho} > 0\}$  and  $\text{RHS}(\rho) := \{i \in \mathcal{M} | r_{i,\rho} > 0\}$ , respectively. With  $\mathbf{N} = (n_{i,\rho}) = (r_{i,\rho} - l_{i,\rho})$ , we denote the stoichiometric matrix of  $(\mathcal{M}, \mathcal{R})$ . W.l.o.g. we assume  $\mathbf{v}_\rho \geq 0$ ; hence a reversible reaction has two entries in  $\mathbf{v}$ .

Given a set  $A \subseteq \mathcal{M}$ , its set of reaction rules  $\mathcal{R}_A = \{\rho \in \mathcal{R} | \text{LHS}(\rho) \subseteq A\}$ , and the corresponding stoichiometric matrix  $\mathbf{N}_A$ , we say that  $A$  is closed if for all reaction rules  $\rho \in \mathcal{R}_A$ ,  $\text{RHS}(\rho) \subseteq A$ . Thus, we call  $A$  closed if there is no reaction with educts from  $A$  producing a species not in  $A$ .  $A$  is self-maintaining if there exists a strictly positive flux vector  $\mathbf{v}' \in \mathbb{R}_{>0}^{|\mathcal{R}_A|}$  such that all species in  $A$  are produced at a non-negative rate, that is,  $\mathbf{N}_A \mathbf{v}' \geq 0$  (Ditrich and Speroni di Fenizio, 2007). A set  $A$  that is closed and self-maintaining is called an organization (Fontana and Buss, 1994). An organization is called reactive if each of its species participates in at least one reaction of that organization. Elementary organizations are reactive organization that cannot be generated as union of other reactive organizations (Centler *et al.*, 2008).

Because organizations may share the same species, the set of organizations together with the set inclusion  $\subseteq$  form a partially ordered set that can be visualized in a Hasse diagram, providing a hierarchical view of the network under consideration: organizations are vertically arranged by size, with small organizations at the bottom. Two organizations are connected by a line, if the upper contains the lower organization and no other organization exists between them. For simplicity, only species appearing for the first time, i.e. which are not element of a lower organization, are displayed.

### 2.2 Analyzing reaction networks with modifiers

In this section, we introduce an algorithm that allows application of OT to reaction network models containing modifiers. As an example, we use a phosphorylation cycle, a typical motive found in signaling networks (Fig. 1A). The network consists of seven molecular species  $\mathcal{M} = \{A, B, M1, \dots, M5\}$  and three reactions  $\mathcal{R} = \{R1, R2, R3\}$ .



**Fig. 1.** (A) Example network (phosphorylation cycle) with seven species and three reaction rules. (B and C) Hasse diagrams of elementary organizations of the unprocessed and processed networks, respectively. Only species appearing for the first time in each organization are displayed.

For the reactions  $\mathcal{R} = \{R1: \emptyset \rightarrow A, R2: B \rightarrow A, R3: A \rightarrow B\}$ , we assume the following kinetic laws (omitting rate constants and units):

$$\begin{aligned} v_{R1} &= 1 \\ v_{R2} &= [B](1 + [M1] + [M2]) \\ v_{R3} &= [A]([M3][M4] + [M3][M5] + [M4][M5]) \end{aligned}$$

This model can be formulated in SBML, with  $M1, M2$  being modifiers of reaction  $R2$  and  $M3, M4, M5$  being modifiers of reaction  $R3$ , while not appearing as reactants.

Our algorithm consists of two steps: first, we examine the kinetic law of each reaction to detect minimal sets of modifiers that are necessary for that reaction to have a positive flux. Then, we use this information to adapt a reaction's set of reactants in order to more faithfully reflect the algebraic structure of the network used for computation of chemical organizations.

**2.2.1 Step 1: identifying sets of essential modifiers** In this first step, we identify all minimal supporting modifier sets of each reaction. Given a reaction  $\rho \in \mathcal{R}$ , a *minimal supporting modifier set* (*supporting set*, for short) is defined as a minimal set of modifiers that need positive concentrations (while all others are absent) to allow reaction  $\rho$  to have a positive flux. If at least one of these modifiers is additionally set to a zero concentration, the flux of the reaction is constrained to zero. There might be several possibly overlapping supporting sets. With respect to a certain reaction, a modifier is called *essential* if it is contained in all supporting sets of the reaction.

**Determination of supporting sets:** to decide whether a set of modifiers is a supporting set for a particular reaction, we follow a straightforward approach. If a set of modifiers is a supporting set, a positive concentration of only these modifiers allows a non-zero flux, while a positive concentration of only a proper subset of these modifiers constrains the flux to zero. Following this idea, we implemented FormulaChecker, which tries to compute the velocity of each reaction in terms of modifier concentrations. All variables in the kinetic law that represent undefined parameters or educt or product species are not further resolved; i.e. they are treated as symbols. The modifiers we want to test to determine whether they belong to a supporting set are also treated as symbols. The remaining modifiers are set to zero concentration. Function calls are resolved by application of their respective parameters, if necessary. Applying FormulaChecker can lead to two different results for the reaction velocity:

- (1) The result is zero. In this case the tested modifier set is not a supporting set. Let  $\{M3\}$ , for example, be the set to be checked in  $R3$ . Setting the concentrations of the remaining modifiers to zero results in  $v_{R3} = 0$ . Thus,  $\{M3\}$  is not a supporting set of  $R3$ . This also applies to the sets  $\{M4\}$  and  $\{M5\}$ .
- (2) The result is non-zero. Thus, it might be a constant only depending on parameters, or a formula, dependent on variables. Checking  $\{M3, M4\}$  in  $R3$  yields the kinetic law  $v_{R3} = [A]([M3][M4])$ . Since we know that  $\{M3\}$ ,  $\{M4\}$  and  $\{M5\}$  do not represent supporting sets,  $\{M3, M4\}$

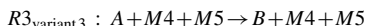
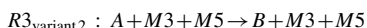
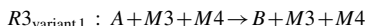
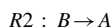
has to be a supporting set. In contrast, if we check the empty set in  $R_2$  by setting  $M_1$  and  $M_2$  to zero values in the kinetic law, we obtain  $v_{R_2} = [B]$ . In consequence, neither  $\{M_1\}$  nor  $\{M_2\}$  represent supporting sets of  $R_2$ ; the supporting set is the empty set, and no further tests are required.

*Finding all supporting sets:* in order to find all supporting sets of a reaction, the algorithm analyzes the power set of the reaction's set of modifiers to ensure that all supporting sets are found. The sets are checked in increasing size order, trying to avoid testing the whole power set of modifiers. If we find that a set of modifiers is a minimal supporting set, we do not need to test any of its supersets.

Looking at  $R_3$  in the example, after the empty set, all single-modifier sets are checked. We find that neither  $M_3$  nor  $M_4$  nor  $M_5$  allow a positive flux if standing alone. In the next step all two-element sets are tested. Since all these sets allow a positive flux of  $R_3$ , but none of the smaller ones, we conclude that  $\{M_3, M_4\}$ ,  $\{M_3, M_5\}$  and  $\{M_4, M_5\}$  are the supporting sets. In consequence, we do not have to test the superset  $\{M_3, M_4, M_5\}$ .

**2.2.2 Step 2: adapting the reactions** In the second step, each reaction possessing at least one supporting set is processed. For each supporting set the reaction is duplicated and the modifiers of the supporting set are added as catalysts to the duplicate reaction. Finally, the original reaction is removed from the model. In order to preserve the dynamics of the original model in the processed model, the kinetic law of each of the duplicate reactions is divided by the number of derived reactions, i.e. the number of supporting sets. The duplicate reactions get new names of the form  $[old\_reaction\_name]_{variant [number]}$ .

For our example, we obtain the following set of reaction rules  $\mathcal{R} = \{R_1, R_2, R_{3_{variant1}}, R_{3_{variant2}}, R_{3_{variant3}}\}$  with



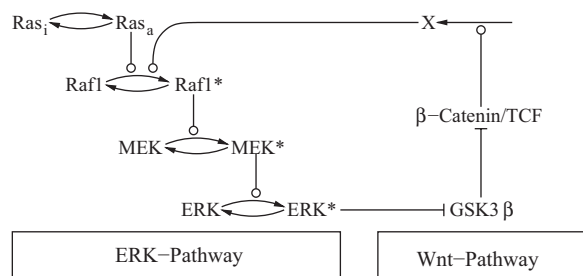
For a more detailed outline of the processing of the kinetic laws, see the Supplementary Material.

**2.2.3 Example application** Applying the algorithm to our example, we can see several effects of the processing of the kinetic laws (see Figs 1B and C for the Hasse diagrams of elementary organizations). Two trends are superimposed. First, some organizations vanish, including the organization solely containing  $A$  and  $B$  in the unprocessed network. In the processed network, a reaction still converts  $B$  to  $A$ . In order to replenish  $B$ , one pair of the modifiers  $M_3$ ,  $M_4$  and  $M_5$  is necessary. Thus,  $\{A, B\}$  does not fulfill the self-maintenance condition in the processed network. Second, some organizations appear for the first time, as in the case of the organization containing  $A$  in the processed network. In the original network, the set  $\{A\}$  was not closed since  $R_3$  unconditionally produced  $B$  from  $A$ .

### 3 ORGANIZATIONAL STRUCTURE OF THE ERK/WNT-SIGNALING PATHWAY

In order to demonstrate the utility of the incorporation of kinetic laws into the analysis with OT, we analyze the model *BIOMD149*<sup>2</sup> from the BioModels database (Le Novère *et al.*, 2006) containing an integrated ERK and Wnt/ $\beta$ -catenin signaling pathway (Figure 2).

<sup>2</sup>We abbreviate the official name of the BioModels by reducing the number to three digits. The original name of the model is *BIOMD0000000149*.

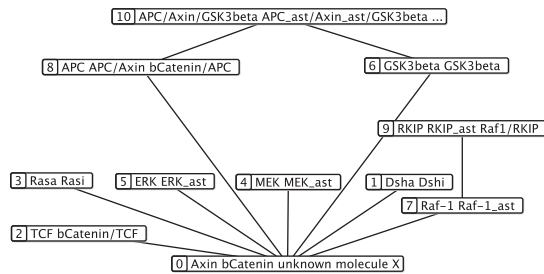


**Fig. 2.** Simplified representation of the reaction network from *BIOMD149* (Kim *et al.*, 2007) combining the ERK- and Wnt-signaling pathways. The Wnt signal, serving as input to both pathways, is not shown. Lines with circles represent essential modifiers identified with the presented approach. Lines ending in orthogonal bars indicate inhibition.

This model is based on the work of Kim *et al.* (2007), who described a positive feedback loop between these two pathways important in the development of some cancer. The positive feedback loop works through a yet unknown mechanism modeled by a species called 'molecule X'. The transcription of this molecule is modeled to be upregulated by a complex of  $\beta$ -catenin and T-cell factor (TCF). The availability of  $\beta$ -catenin is regulated by active glycogen synthase kinase  $3\beta$  (GSK- $3\beta$ ), which in turn is inactivated by phosphorylated ERK. According to the model, X upregulates the signaling through the ERK-pathway. The rates of phosphorylation of the different levels of the ERK-pathway are modeled with kinetic laws. Thus, a high concentration of phosphorylated Raf increases the rate of phosphorylation of MEK, which in turn increases the rate of phosphorylation of ERK.

Without the processing of the kinetic laws the network contains 384 reactive organizations generated from the union of 11 elementary organizations. After processing, the network contains 150 reactive organizations generated from the union of 18 elementary organizations. Thus, the number of reactive organizations declines, while the number of elementary organizations increases. Figures 3 and 4 depict the Hasse diagram of elementary organizations of both networks. The Hasse diagram of the unprocessed network (Fig. 3) displays a very simple structure. The smallest organization already contains X. From the kinetic law of the production reaction of X, it can be determined that a positive concentration of the complex  $\beta$ -catenin/TCF is required for a non-zero flux of this reaction. But this is not taken into account since this constraint is modeled through the modifiers of the reaction and not on the level of substrates and educts as required by the Feinberg condition. Consequently, the different levels of the ERK-signaling pathway are also present independent of each other. This can be observed by the presence of the corresponding phosphorylated and dephosphorylated proteins directly above the smallest organization in the Hasse diagram.

From a simulation perspective, the reactive organizations of the original network would indicate a state of the network where, for example, MEK and MEK\* as well as the input species could be constantly present (Fig. 3, organization 4). However, by examining the kinetic laws of the phosphorylation from MEK to MEK\*, we find that this reaction has a flux of zero if the species Raf1\* is not present. Thus, only the dephosphorylation of MEK would have a positive flux, finally using up all MEK\*. After



**Fig. 3.** Hasse diagram of elementary organizations of *BIOMD149* without processing of the kinetic laws. Only species appearing for the first time in each organization are shown. For example organization 9 contains the species displayed in the nodes corresponding to organization 0, 7 and 9. Not all species in organization 10 are displayed. A list of abbreviations can be found in the Supplementary Material. Phosphorylated forms of a protein are denoted by the suffix ‘\_ast’. Active/Inactive forms by the suffix ‘a’/‘i’.



**Fig. 4.** Hasse diagram of elementary organizations of *BIOMD149* after the processing of the kinetic laws. Only species appearing for the first time in each organization are shown. Not all species in organization 12 are displayed. Naming follows the same conventions as in Figure 3. The different pathways for upregulation of the ERK-signaling pathway are indicated. In comparison to Figure 3, we find, for example, the node corresponding to organization 6 above the node corresponding to organization 2 (corresponding to the nodes labeled 7, respectively, 3 in Fig. 3). This corresponds to the conclusion that a positive concentration of Rasa and Rasi is required for the presence of Raf1 and Raf1\* in the limit behavior. Comparison with Figure 3 shows that this conclusion can be drawn only if the kinetic laws are processed.

processing of the corresponding kinetic law, Raf1\* is identified as an essential modifier and added as a catalyst to the reaction, as seen in the Hasse diagram of the processed network (Fig. 4). The organization containing the species MEK\* and MEK (Fig. 4, organization 8) is situated above the organization containing Raf1\* (Fig. 4, organization 6).

From this perspective, the processing of the kinetic laws can be seen as adding mechanistic detail to the reactions. Thus, when we

find Raf1\* necessary for the phosphorylation of MEK to MEK\*, the addition of the modifier Raf1\* as catalyst corresponds to the complex formation between Raf1\* with MEK prior to phosphorylation. The approach to consider kinetic laws in OT can be seen as refinement of the reactions of a model making use of the additional information present in kinetic laws. Even though OT does not explicitly require the kinetic laws of a reaction network, knowledge about them can be used to better predict the limit behavior of a reaction network. Conversely, in the sense of the Feinberg condition, the underlying mechanisms are modeled more accurately on the stoichiometric level of the network if this approach is used.

In agreement with the results of Kim *et al.* (2007), we find an alternative route for the activation of the ERK-pathway, indicated by the organizations 3, 7, 9 and 11 in Figure 4. Through the action of the complex  $\beta$ -catenin/TCF, the transcription of X is upregulated and, thus, bypasses the activation of Raf by Ras. A constant activation of  $\beta$ -catenin/TCF, for example through a mutation, can result in a decoupling from any signal and consequently lead to a constant upregulation of the ERK-signaling pathway, as is often found in cancer (Kim *et al.*, 2007). In the unprocessed network, we do not obtain these results.

## 4 LARGE-SCALE ANALYSIS OF BIOMODELS

In order to demonstrate the utility of our approach, we analyze the models of the 11th release<sup>3</sup> of the BioModels database (Le Novère *et al.*, 2006). This database contains 185 manually curated models of biological networks in SBML format.

SBML allows species to be defined as external. Thus, their concentration is assumed constant. For the computation of chemical organizations, we add an inflow and outflow reaction of the form  $\emptyset \rightarrow s$  and  $s \rightarrow \emptyset$  for each external species  $s$ . For all except 3 models, we were able to compute the reactive organizations using the deterministic algorithms for organization computation [see Centler *et al.* (2008) for algorithmic details]. For the remaining three models (*BIOMD014*, *BIOMD019* and *BIOMD049*), a heuristic based on a random walk strategy to determine organizations (Centler *et al.*, 2008) needed to be applied. Since we wanted to identify species appearing in no organization and each of these models did contain an organization encompassing the entire species set, computation of the complete set of organization was not necessary for these models.

A total of 172 models contained a non-empty organization. In the remaining 13 models only the empty organization was found, since they contained neither reactions nor species. An overview of the number of reactive organizations is given in Table 2 in the Supplementary Material. While 77 models contained only a single reactive organization, the highest number of organizations was found in *BIOMD175*, with 319 248 reactive organizations. An overview of the distribution of the number of organizations can be found in Figure 5.

Species participating in none of the reactions can drastically increase the number of organizations in a network. Thus, we computed only the reactive organizations in each network and omitted species participating in no reaction (in 24 models) from the analysis. In 31 models some species did not appear in

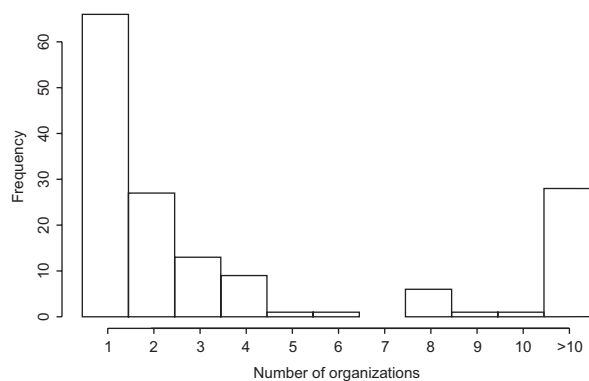
<sup>3</sup>The BioModels Database is updated in releases whereby models are corrected or added. We downloaded the models used in this work on October 20, 2008.



any reactive organization. A first analysis showed that this set contains many models where such behavior was intended. Thus, in several models the concentration of some species was set to a non-zero value at a given time point (e.g.  $t=0$ ). To take into account this short-time behavior, we added an inflow reaction for each such species. Doing this, we found that only five models with species absent from any reactive organization remained: *BIOMD044*, *BIOMD093*, *BIOMD094*, *BIOMD143* and *BIOMD151* (Table 1). By analyzing the reactions in which the missing species participated and comparing the SBML models to their description in the corresponding publications, we found the potential inconsistencies. We identified all these inconsistencies as actual modeling errors.

#### 4.1 Resolving network inconsistencies

In three of the five models, *BIOMD093*, *BIOMD094* and *BIOMD143*, we identified reactions that were set to irreversible despite their kinetic laws producing negative fluxes in the course of the simulation, as described in the corresponding publications. Thus,



**Fig. 5.** Histogram of the number of reactive organizations in the models of the BioModels Database. Please note that this number includes six models with more than 1000 organizations (listed below >10 organizations).

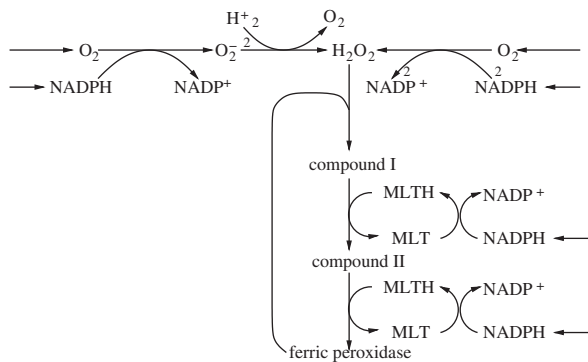
they were indeed reversible and we modified them accordingly. Repeating the analysis, we found all species present in the reactive organizations of *BIOMD093*. In *BIOMD094*, missing species remained. However, this was an intended behavior since a gene knockout was modeled (Yamada *et al.*, 2003).

In *BIOMD143*, we still found some species absent after we had changed reactions with negative fluxes in the simulation to be reversible. This model describes the oscillatory metabolism of activated neutrophils (Olsen *et al.*, 2003). A simplified and decompartmentalized version of the relevant reactions is depicted in Figure 6. The species absent from the reactive organizations are hydrogen from cytoplasm and phagosome. The model contains only reactions consuming these two species. During simulation even negative concentrations of both species appear. The reason for the consumption of these species is inconsistent modeling of the stoichiometry of the reactions and an inconsistent kinetic law. Cytoplasmatic and phagosomal hydrogen are consumed together with superoxide ( $O_2^-$ ) to produce hydrogen peroxide ( $H_2O_2$ ). In the course of the disposal of  $H_2O_2$  by ferric peroxidase in the phagosome, additional four protons from melatonin (MLTH) are consumed to produce the initial form of ferric peroxidase. With the exception of ferric peroxidase and free radicals of melatonin (MLT), all species are consumed without producing equivalent products. Thus, the disposal of  $H_2O_2$  by ferric peroxidase consumes oxygen and protons. The model contains an inflow for NADPH and  $O_2$ . Oxidation of NADPH by oxygen or free radicals of melatonin can produce superoxide and melatonin, respectively. Thus, there is a constant inflow of NADPH and oxygen that can replenish the consumed species. However, the kinetic law of the production of superoxide from  $O_2^-$  and hydrogen does not depend on the concentration of hydrogen in the model. Together with a zero initial concentration of hydrogen, the simulation of the model leads to a negative concentration of this species. Making the rate law dependent on the concentration of hydrogen resolves the problem of negative concentration of hydrogen. Additionally, either removing the inconsistencies in the stoichiometry or adding an inflow for

**Table 1.** Selected results from the large-scale analysis

Model	Description	Species/ Reactions	Reactive Orgs.	First Step		Second Step	
				OT (spec./rea.)	FBM (rea.)	OT (spec./rea.)	FBM (rea.)
BIOMD037	Sporulation control network in <i>P. polycephalum</i> (Goldbeter, 1991)	12/14 (14)	1 (2)	3/6	10	12/14	14
BIOMD044	Model of intracellular calcium oscillations (Borghans <i>et al.</i> , 1997)	7/8 (8)	2 (2)	3/4	5	6/7	7
BIOMD093	JAK/STAT signal transduction pathway (Yamada <i>et al.</i> , 2003)	34/48 (48)	5 (3)	11/16	30	31/43	43
BIOMD094	JAK/STAT signal transduction pathway (Yamada <i>et al.</i> , 2003)	34/47 (47)	2 (3)	5/5	27	24/24	40
BIOMD143	Oscillatory metabolism of activated neutrophils (Olsen <i>et al.</i> , 2003)	20/20 (20)	1 (1)	4/4	4	7/5	5
BIOMD149	Crosstalk between Wnt and ERK Pathways (Kim <i>et al.</i> , 2007)	28/39 (39)	150 (384)	28/39	39	–	–
BIOMD151	IL-6 signal transduction in hepatocytes (Singh <i>et al.</i> , 2006)	68/114 (114)	80 (96)	49/71	111	19/14	112

See Supplementary Material for the entire table. The five models in which inconsistencies have been identified are shaded in light gray. The first 4 columns give general details about the models. Numbers in brackets indicate the number of reactions of the original network that can increase through processing of the kinetic laws. The number of species remains constant. The fourth column gives the number of reactive organizations in the modified and (in brackets) the original network. In the fifth and sixth columns species and reactions that can be present in the limit behavior of the processed network are given. OT denotes the predictions by OT, and FBM the predictions by flux-based methods. In some cases, FBM identifies more reactions to be present in the limit behavior than OT. These cases are shaded in dark gray. The seventh and eighth columns give the same numbers when inflow reactions for species with an event setting their concentration to a positive value at a certain time point are added. In cases where the original network already contained all species, those numbers are omitted.



**Fig. 6.** Simplified representation of the reactions of *BIOMD143*. As a result of inconsistent stoichiometries hydrogen and oxygen are consumed in the course of detoxification of hydrogen peroxide. There is only an inflow of oxygen, and the consumption of hydrogen does not depend upon its concentration. Consequently, a simulation leads to a negative concentration of hydrogen.

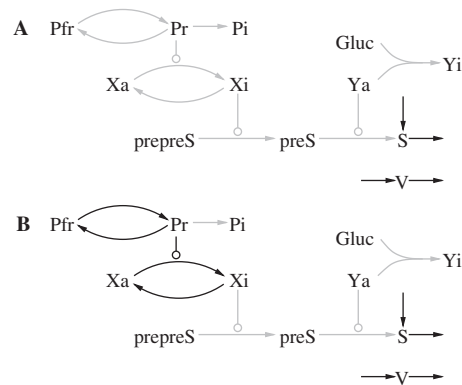
hydrogen allows positive concentrations of this species during simulation.

The reasons for the missing species in *BIOMD044* are very similar. Here, a species is modeled to serve as a pseudo-substrate to a reaction that could have been modeled without substrate. The kinetic law governing the reaction does not depend upon the concentration of this substrate. Since it is not produced by any other reaction, negative concentrations appear in the course of the simulation. Replacing the respective reaction by an inflow reaction resolves the problem.

In *BIOMD151* almost all species are absent from reactive organizations. This network represents an integrated model of the JAK/STAT and ERK-signaling pathways regulated by IL-6 in hepatocytes (Singh *et al.*, 2006). A detailed analysis of the model and the set of ordinary differential equations presented in Singh *et al.* (2006) showed that a complex formation step was missing, such that the signal from IL-6 could not be transmitted to the subsequent signaling pathways. Only the complex dissociation reaction was present. During simulation it had a negative flux, mimicking the complex formation reaction. Adding the missing step produced a model in which all species appeared in a reactive organization.

## 4.2 Comparison with flux-based methods

Next, we will compare our results with those obtained with flux-based methods, including FBA (Varma and Palsson, 1994), elementary mode analysis (Schuster *et al.*, 1999) and extreme pathway analysis (Schilling *et al.*, 2000). These methods can be used to check whether a certain reaction can be present in a steady-state flux obeying the irreversibility constraint. Thus, they can predict whether a reaction can be present in the limit behavior of a reaction network. In FBA this can be done directly, while elementary mode analysis and extreme pathway analysis return a set of vectors spanning the solution space of the steady-state condition. However, since OT also takes into account growth states, in which some species accumulate, the steady-state condition is adapted accordingly (details can be found in the Supplementary Material). Furthermore, since we only want to know whether a reaction can appear in any steady state or growth state, we do not need to apply



**Fig. 7.** Reaction network from *BIOMD037* modeling the sporulation control in *P.polycephalum* from Marwan (2003). Lines ending in circles indicate essential modifiers for a reaction. Light gray reactions cannot have a positive flux in the limit behavior, according to **A** OT and **B** flux-based methods. Abbreviations: Pr, active photoreceptor; Pi, inactive photoreceptor; (pre/prepre)S, sporulation signal (and precursors); Ya/i, active/inactive glucose receptor; Gluc, glucose; Xa/i, active/inactive signal transducer.

these methods directly, but can use a linear programming approach similar to FBA, outlined in the Supplementary Material.

We compared the predictions of flux-based methods to those of OT for the models of the BioModels Database. With OT, we identified 31 models where some reactions did not appear in any reactive organization. The same 31 models are identified using flux-based methods. However, when analyzing the predicted set of available reactions in detail, we found differences in 25 of the 31 models. Due to the definition of self-maintenance, the set of available reactions is a subset of those predicted by flux-based methods. Thus, in all 25 cases, flux-based methods found reactions present in the limit behavior that indeed could not maintain a positive flux in a long-term simulation.

The reason for this difference closely follows a concept presented in Kaleta *et al.* (2006): a steady-state flux in a network uses some species that cannot be produced at a positive rate. In this flux, these species might be interconverted into each other or act as catalysts. Further assume that there is a reaction steadily draining some of the unproducible species. Thus, they will finally vanish. In consequence, this steady-state flux cannot be part of any steady state of the complete network. If a particular reaction is present only in such steady-state fluxes, it is predicted to be present in the limit behavior of a reaction network by flux-based methods, while OT correctly identifies it as absent since it correctly takes into account the drain of the unproducible species. We will outline this concept in more detail using *BIOMD037*, a model of the sporulation control network in *Physarum polycephalum* by Marwan (2003) (Fig. 7). While OT predicts 8 of the 12 reactions to be absent from the limit behavior (Fig. 7A), flux-based methods identify only four such reactions (Fig. 7B). The differentially predicted reactions account for the interconversion of Pfr to Pr and Xi to Xa. Flux-based methods find a flux where the conversion of Pfr to Pr and vice versa is in equilibrium. However, this does not take into account that there is also a reaction irreversibly converting Pr to Pi. Thus, a non-zero concentration of Pr will be depleted by the conversion into Pi. In consequence, there is no reactive organization containing Pfr and Pi.

Additionally, we find an interesting case in the interconversion of Xa to Xi and vice versa. The conversion of Xi to Xa requires the presence of Pr. Flux-based methods identify an equal flux of both reactions as a feasible flux, since Pr acts only as a catalyst. However, the analysis using OT shows that such a flux also requires the presence of Pr. Thus, both species cannot persist in the limit behavior since Pr, required for the reaction of Xi to Xa, will vanish over time. Since Xa is steadily converted to Xi, only this species would finally remain. This demonstrates how our approach takes the kinetic laws into account which is not possible using flux-based methods.

In two of the models in which we identified inconsistencies, *BIOMD094* and *BIOMD151*, predictions for the presence of reactions in the limit behavior between OT and flux-based methods differ. In *BIOMD151*, OT predicts nine reactions to be present, while flux-based methods identify 112 of the 114 overall reactions. As outlined above, flux-based methods can predict only the same or a larger set of reactions to be present in the limit behavior. Thus, the search for inconsistencies is simplified by reducing the size of the system to analyze if OT is used. This is also corroborated by three models in the uncurated branch of the BioModels Database containing inconsistencies. In all three models, flux-based methods predict more reactions to be present in the limit behavior than OT (see Supplementary Material for further details).

## 5 CONCLUSIONS

In this work, we demonstrated that information hidden in kinetic laws affects the results obtained from chemical organization theory (OT). We presented an approach that is able to uncover this information. This approach enabled us to refine the chemical organizations in 41 of the 185 models (22%) of the BioModels Database. The Hasse diagram of organizations of the processed model of a combined ERK/Wnt-signaling pathway took into account the different levels of phosphorylation in the signaling cascade, while the set of organizations of the unprocessed network did not. Furthermore, the Hasse diagram of organizations demonstrated several possible pathways for constant upregulation of this pathway, an important event in carcinogenesis consistent with the results of Kim *et al.* (2007).

Analyzing the 185 models of the BioModels Database, we checked the behavior of the models during long-term simulation (limit behavior). Thus, we found 31 models where several species could not persist in a long-term simulation. Furthermore, we identified five models in which some species could not be present at all during simulation. This was due to inconsistent reversibility constraints in two models, negative concentrations of some species during simulation in other two models and a missing reaction in the fifth model. In the non-curated branch of the BioModels Database, we identified the models with modeling errors. Comparing the set of species present and the reactions having a non-zero flux in the limit behavior, we found OT able to predict those sets more accurately in 25 models (14%) compared with flux-based methods like FBA, elementary mode analysis and extreme pathway analysis. These models account for 81% of the models in which the set of species and reactions present in the limit behavior of the model did not encompass the entire set of species and reactions. In five of the 8 models of both branches of the BioModels Database in which we

detected modeling errors, OT made more accurate predictions in comparison to flux-based methods.

These results demonstrate that OT is a valuable tool in three important aspects of network design and analysis. First, when this approach is used to extract additional information from the kinetic laws of the reactions, the set of organizations corresponds to the potential steady state and growth states of a reaction network. Thus, important information about the dynamic structure of a reaction network can be uncovered. Second, OT can be used in an iterative fashion to assist in model building by identifying inconsistencies that need to be resolved. Third, OT more faithfully identifies parts of a network whose maintenance is not yet explained than flux-based methods. Thus, it is of particular interest for identifying gaps due to missing knowledge in large-scale metabolic networks as documented in Centler *et al.* (2008). In consequence, it can be beneficial for methods aiming to remove such inconsistencies (Kumar *et al.*, 2007; Reed *et al.*, 2006). In the other direction, our approach could be extended by these methods to automatically propose changes in order to remove inconsistencies. However, computational constraints currently prohibit the application of our deterministic algorithms to very large networks (e.g. more than 500 reactions). An approximation can be used for networks of this size, but the results require manual checking. A more efficient algorithm that will enable the application of OT to genome-scale networks is in development.

## ACKNOWLEDGEMENTS

The authors would like to thank four anonymous referees for very helpful comments. Furthermore, we thank Sandro Lambeck for helpful information on the Wnt/ $\beta$ -catenin pathway.

*Funding:* German Research Foundation (Grant Di852/4-2); the European Union, NEST-project ESIGNET no. 12789.

*Conflict of Interest:* none declared.

## REFERENCES

- Borghans, J.M. *et al.* (1997) Complex intracellular calcium oscillations. A theoretical exploration of possible mechanisms. *Biophys. Chem.*, **66**, 25–41.
- Centler, F. *et al.* (2007) Chemical organizations in the central sugar metabolism of *Escherichia coli*. In Deutsch, A. (eds) *Mathematical Modeling of Biological Systems*, Vol. 1. Birkhäuser, Boston, pp. 109–123.
- Centler, F. *et al.* (2008) Computing chemical organizations in biological networks. *Bioinformatics*, **24**, 1611–1618.
- Dittrich, P. and Speroni di Fenizio, P. (2007) Chemical organization theory. *Bull. Math. Biol.*, **69**, 1199–1231.
- Dittrich, P. *et al.* (2003) On the formation of social order – modeling the problem of double and multi contingency following Luhmann. *JASSS*, **6**.
- Feinberg, M. and Horn, F.J.M. (1974) Dynamics of open chemical systems and the algebraic structure of the underlying reaction network. *Chem. Eng. Sci.*, **29**, 775–787.
- Fontana, W. and Buss, L.W. (1994) ‘The arrival of the fittest’: toward a theory of biological organization. *Bull. Math. Biol.*, **56**, 1–64.
- Goldbeter, A. (1991) A minimal cascade model for the mitotic oscillator involving cyclin and CDC2 kinase. *Proc. Natl Acad. Sci. USA*, **88**, 9107–9111.
- Hofbauer, J. and Sigmund, K. (1998) *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge.
- Hucka, M. *et al.* (2003) The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.
- Kaleta, C. *et al.* (2006) Analyzing molecular reaction networks: from pathways to chemical organizations. *Mol. Biotechnol.*, **34**, 117–124.

- Kaleta,C. et al. (2008) Phenotype prediction in regulated metabolic networks. *BMC Syst. Biol.*, **2**, 37.
- Kim,D. et al. (2007) A hidden oncogenic positive feedback loop caused by crosstalk between Wnt and ERK pathways. *Oncogene*, **26**, 4571–4579.
- Kumar,V.S. et al. (2007) Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics*, **8**, 212.
- Le Novère,N. et al. (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.*, **34**, D689–D691.
- Marwan,W. (2003) Theory of time-resolved somatic complementation and its use to explore the sporulation control network in *Physarum polycephalum*. *Genetics*, **164**, 105–115.
- Matsumaru,N. et al. (2007) Toward organization-oriented chemical programming: a case study with the maximal independent set problem. In Dressler,F. and Carreras,I. (eds) *Advances in Biologically Inspired Information Systems*, Vol. 69 of *Studies in Computational Intelligence*, Springer, Berlin, pp. 147–163.
- Olsen,L.F. et al. (2003) A model of the oscillatory metabolism of activated neutrophils. *Biophys. J.*, **84**, 69–81.
- Reed,J.L. et al. (2006) Systems approach to refining genome annotation. *Proc. Natl Acad. Sci. USA*, **103**, 17480–17484.
- Schilling,C.H. et al. (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.*, **203**, 229–248.
- Schuster,S. et al. (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.*, **17**, 53–60.
- Singh,A. et al. (2006) Modeling regulatory mechanisms in IL-6 signal transduction in hepatocytes. *Biotechnol. Bioeng.*, **95**, 850–862.
- Varma,A. and Palsson,B.O. (1994) Metabolic flux balancing: basic concepts, scientific and practical use. *BioTechnology*, **12**, 994–998.
- Wodarz,D. and Nowak,M.A. (1999) Specific therapy regimes could lead to long-term immunological control of HIV. *Proc. Natl Acad. Sci. USA*, **96**, 14464–14469.
- Yamada,S. et al. (2003) Control mechanism of JAK/STAT signal transduction pathway. *FEBS Lett.*, **534**, 190–196.