Available online at www.sciencedirect.com
SCIENCE DIRECT®
BioSystems xxx (2006) xxx–xxx
BioSystems
www.elsevier.com/locate/biosystems

Qualitative analysis of the relation between DNA microarray data and behavioral models of regulation networks

A. Siegel[a,*], O. Radulescu[b], M. Le Borgne[a], P. Veber[a], J. Ouy[a], S. Lagarrigue[c]

[a] *IRISA, Symbiose, Campus de Beaulieu, 35042 Rennes Cedex, France*
[b] *IRMAR-Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France*
[c] *UMR Génétique Animale, Agrocampus Rennes-INRA, 65 Rue de Saint-Brieuc, CS 84215 Rennes, France*

Received 22 August 2005; received in revised form 20 September 2005; accepted 4 October 2005

Abstract

We introduce a mathematical framework that allows to test the compatibility between differential data and knowledge on genetic and metabolic interactions. Within this framework, a behavioral model is represented by a labeled oriented interaction graph; its predictions can be compared to experimental data. The comparison is qualitative and relies on a system of linear qualitative equations derived from the interaction graph. We show how to partially solve the qualitative system, how to identify incompatibilities between the model and the data, and how to detect competitions in the biological processes that are modeled. This approach can be used for the analysis of transcriptomic, metabolic or proteomic data.

Keywords: Systems biology; Qualitative analysis; Steady state shift; Sign algebra

1. Introduction

1.1. Systems biology: models and data

The field of *systems biology* appeared as a response to increasing need for analytical approaches in molecular biology. Its goals include modeling interactions, understanding the behaviour of a system from the interplay of its components, confronting the prediction of the model to data, and inferring models from data. Solutions to these challenges are often interdisciplinary.

The dynamical framework includes simulations and prediction of behaviours; models can be either qualitative or quantitative, as reviewed in (de Jong, 2002; Chaves et al., 2005; King et al., 2005). A first approach makes use of continuous models: the concentrations of products are modeled by continuous functions of time, governed by differential equations. This framework allows one to state biological properties of networks, eventually by using simulation software (Bakker et al., 1997; Eisenthal and Cornish-Bowden, 1998; Tyson et al., 2003; Mendes, 1997; Tomita et al., 1999). The properties of continuous models can be studied with convex analysis, linear and non-linear control techniques (Fell, 1997; Heinrich and Schuster, 1996; Papin et al., 2004; Angeli et al., 2004). Stochastic models transform reaction rates into probabilities and concentrations into numbers of molecules, allowing to understand how noise influences a system (Rao et al., 2002; Kaern et al., 2005). Finally, in the discrete models, each component is assumed to have a small number of qualitative states, and the regulatory interactions are described by discrete functions. Relevant discrete frameworks can be boolean (Kauffman, 1993; Sanchez and Thieffry, 2001), logical (Karp et al., 1996; Reiser et al., 2001), or Petri networks (Matsuno et al., 2000; Chaouiya et al., 2004). The bridge between con-

* Corresponding author. Tel.: +33 299847448; fax: +33 299847171.
E-mail address: anne.siegel@irisa.fr (A. Siegel)

0303-2647/$ – see front matter © 2005 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2005.10.006

tinuous and discrete models is made by piecewise linear differential models (de Jong et al., 2004; Ghoshn and Andomlin, 2004).

Each of these methods addresses in complementary ways dynamical properties such as the existence of attractors (limit cycles or steady states) and the behavior of these with respect to changes in the parameters (Thomas, 1973; Soulé, 2003; Chaves et al., 2005). They represent powerful tools to acquire a fine grained knowledge of the system at hand, but they need accurate data on chemical reactions kinetics or qualitative information. These data are scarcely available. Furthermore, these methods are also computationally demanding and their practical use is restricted to a limited number of variables.

Model identification addresses a different objective, that is, to form or modify a model consistently with a set of data. A first framework for identification consists in building models from scratch, using statistical techniques such as bayesian networks (Kaminski and Friedman, 2002; Nachman et al., 2004) or kernels (Yamanishi et al., 2004); these are particularly accurate when large amounts of data are available. Another efficient approach formalizes a priori knowledge as partially specified models. Fitting models to data is obtained by means of various techniques, depending on the class of models, that can be discrete (Bay et al., 2003; Zupan et al., 2001; Bryant et al., 2001; Reiser et al., 2001), continuous (Batt et al., 2005; Boyer and Viari, 2003; King et al., 2005) or hybrid (Calzone et al., 2005; Langley et al., 2005). Qualitative reasoning, hybrid system, constraint programming or model-checking allow either to identify a subset of active processes explaining experimental time-series data (Bay et al., 2003; Zupan et al., 2001; Bryant et al., 2001; Reiser et al., 2001) or to correct the models and infer some parameters from data (Batt et al., 2005; Chabrier-Rivier et al., 2004). The identification methods are limited to a few dozen components. Model correction or parameter regression can cope with up to hundreds of products (Chabrier-Rivier et al., 2004) provided that the biomolecular mechanisms and supplied kinetic data are accurate enough.

1.2. Steady state shift experiments and microarray data

Qualitative data such as DNA microarrays data cannot be easily used in most of the frameworks described above for two main reasons. First, the model-based identification approach has difficulties to take into account the errors and the variability that commonly affect measured expression levels in DNA microarrays. Secondly, time series data is not easily available and in many situations (for instance disease studies on clinical tissues) microarrays provide static data, meaning that they inform more on steady state shifts under perturbations than on the dynamics of the system.

In this paper, we develop a mathematical framework that allows us to check the compatibility between gene expression and metabolite concentrations differential data and a graphical model for the interactions among the measured products (metabolites and genes). If incompatibility is found we propose corrections to data or to the model. Our mathematical results connect network topology and the response to steady state shift experiments. Steady state shift experiments are useful tools in chemistry allowing in principle to recover the reaction mechanisms (Chevalier et al., 1993). We argue that similar approaches are well adapted to differential microarray experiments which compare gene expressions between two different states.

In our approach DNA microarrays are interpreted as qualitative data: we only consider the sign of the difference in the expression levels among the two situations. Additionally, we can consider qualitative data on metabolites, provided by biochemical measurements.

Biological literature is rich in knowledge about interactions between molecules. This information is scattered in many publications and it is not globally compared to the results of large scale experiments. A reason is the difficulty to translate the biological knowledge into models: this requires further information on chemical kinetic parameters and mechanisms.

In our models, we do not need such details. We simply use the biological information to build a labeled oriented graph (interaction graph). The nodes of this graph represent molecules (such as mRNA, proteins or metabolites). The edges represent interactions, labeled by their sign. Edge signs can be interpreted dynamically: a "+" on an edge between A and B means that an increase of the concentration of A increases the rate of production of B. A "−" on an edge stands for a decreasing effect. The necessary information to build such a graph can be collected either from interaction databases or can be manually extracted from the literature (Calvano et al., 2005).

Restricting ourselves to a qualitative framework allows us to compose the interaction rules, but it also tolerates redundancies. Particularly, we can include in the model information on indirect interactions, that is, the modeler can add an edge for A to B as soon as he knows that variations of A affect the variations of B, even if the underlying mechanism is not known.

1.3. Qualitative analysis

We present a formal analysis of interaction graphs based on a mathematical model, that allows to confront the interaction graph and qualitative data.

We suppose that the dynamics of the system can be described by a system of differential equations. An experiment is modeled as a steady state shift of the differential dynamical system. We perform transformations on the differential system that allow to connect the variations of products in a linear system. The quantitative system generalizes the discrete Laplace equation on graphs. Like for the Laplace equation, we solve the linear Dirichlet problem (Chung and Yau, 2000; Soardi, 1994; Campanino and Petritis, 2003). We also solve the nonlinear Dirichlet problem: this means relating the interior values to the values on the entrance boundary of the oriented graph even when the products variations are large.

We further suppose that experimental information on products is qualitative, that is, we can find out if the concentration of a product has increased, it has decreased, or it has not changed significantly. In order to exploit this information, qualitative models are derived from the interaction graph and from the differential dynamics: we transpose the quantitative equations into a linear qualitative system in the sign algebra (Kuipers, 1994; Travé-Massuyès and Dague, 2003). Provided that the signs of the interactions are constant for states within the range of experimental variations, the linear qualitative system applies to both small and large variations of the products.

We have developed several complementary methods to analyze the system of qualitative equations and to test whether experimental data are solutions of the system.

- The *graph valuation algorithm* is an automated method to partially or totally solve the qualitative system of equations. When there is no solution, incompatibilities between the model and the data are detected, as well as the place where these incompatibilities occur. When a solution exists, the algorithm predicts the signs of variations of products that have not been measured.
- The *essential balance computation* allows us to detect the variables in the system that are influenced by competitive pathways and to find which pathway wins the competition. Eventually, this computation detects the presence of errors in the data.

The modeling and the algorithms will be illustrated on a simplified model of regulated lipogenesis in liver. Then, we apply the algorithm to an extended model of lipogenesis. With these algorithms, we are able to validate or correct data and models. Also, they serve to emphasize products and pathways that have biological dynamic interest. Further improvements of the algorithms such as coding of qualitative equations over Galois fields and applying them to models containing hundreds of products will be presented elsewhere (Veber et al., 2005).

This paper is organized as follows. In Section 2 we describe our working example which is the regulation of lipogenesis in liver. Section 3 is devoted to modeling assumptions. We introduce the mathematical concepts and we show how the qualitative equations can be obtained. We detail in Section 4 two methods to solve the system of qualitative equations derived in Section 3. Section 5 is devoted to an independent method for analyzing incompatibilities and competitions. Section 6 presents the application of the algorithm to an extended model of lipogenesis.

2. Working example: regulation of the synthesis of fatty acids

Gene regulation associated with fatty acid synthesis in liver is our working example. The corresponding interactions are intricate and involve hundreds of molecules. By way of illustration we have kept as nodes of our illustrative model only the most important biological molecules and their intermediates. The choice of the observed molecules was not optimized; we shall see how our analysis suggests improvements on it.

Two ways of production of fatty acids coexist in liver. Saturated and mono-unsaturated fatty acids are produced from citrates due to the presence of a metabolic pathway composed of four enzymes, namely ATP citrate liase (ACL), acetyl-Coenzyme A carboxylase (ACC), fatty acid synthase (FAS) and Stearoyl-CoA desaturase 1 (SCD1). Polyunsaturated fatty acids (PUFA) such as arachidonic acid and docosahexaenoic acid are synthesized from essential fatty acids provided by nutrition; Delta-5 desaturase (D5D) and Delta-6 desaturase (D6D) catalyze the key steps of the synthesis of PUFA.

PUFAs play pivotal roles in many biological functions. They regulate the expression of genes that impact on lipid, carbohydrate, and protein metabolism. The effects of PUFA are mediated directly or indirectly as shown in the following examples. Direct effects are due to bindings leading to changes in the trans-activating activity of nuclear receptors (PPARα, peroxisome proliferator activated receptors; LXRα, liver-X-receptor α; HNF-4α). Indirect effects result in changes in the abundance of regulatory transcription factors (SREBP-1c, sterol regulatory element binding-protein; ChREBP; etc.) (Jump, 2004).

2.1. Variables in the model

We have considered in our model the transcription factors PPARα, LXRα and SREBP-1c (denoted by PPAR, LXR, SREBP), as they are synthesized from the corresponding genes. We have included the transactivating active forms of these nuclear receptors: LXR-a (denoting a complex LXRα:RXRα), PPAR-a (complex PPARα:RXRα) and SREBP-a (cleaved form of SREBP-1c). We have also considered SCAP (SREBP cleavage activating protein), a key enzyme involved in the cleavage of SREBP-1c. SCAP interacts with another family of proteins called INSIG, showing the complexity of molecular mechanism. We have included PUFA to symbolize metabolites. Finally, we have considered the enzymes ACL, ACC, FAS, SCD1 (implied in the fatty acid synthesis from citrate) and D5D, D6D (implied in PUFA synthesis).

2.2. Interactions in the model

Relations between the variables are as following. SREBP-a is an activator of the transcription of ACL, ACC, FAS, SCD1, D5D and D6D (Nara et al., 2002; Jump, 2004). LXR-a is an activator of the transcription of SREBP and FAS. It also indirectly activates ACL, ACC and SCD1 (Steffensen and Gustafsson, 2004). These indirect actions are kept in the model because we do not know whether they are only SREBP-mediated.

PUFA activates the formation of PPAR-a from PPAR. It inhibits the formation of LXR-a from LXR. It also inhibits the formation of SREBP-a (by inducing the degradation of mRNA and inhibiting the cleavage) (Jump, 2004). SCAP represents the activators of the formation of SREBP-a from SREBP. SCAP is inhibited by PUFA.

PPAR directly activates the transcription of SCD1, D5D, D6D (Miller and Ntambi, 1996; Tang et al., 2003; Matsuzaka et al., 2002).

The activation of SCD1, D5D and D6D by both SREBP and PPAR is paradoxical because fatty acid synthesis (partially governed by SREBP) and oxidation (partially induced by PPAR) are antagonistic in liver. Nevertheless, PUFA have a regulatory role in oxidation. Hence, the induction of D5D and D6D gene by PPAR could be a compensatory response to the increased PUFA demand caused by induction of fatty acid oxidation.

2.3. Working data set: virtual fasting protocol

Our goal is to test whether data can fit properly with the full knowledge on the system. Such data can be an experimental set that have to be analyzed. It could also simply be a set of hypothetical data whose coherence have to be tested. Here, our working data set is inspired by the results of fasting extracted from literature as described hereafter, modified so that a potential error is introduced.

A compilation of recent literature on lipogenesis regulation indicates that SREBP, ACL, ACC, FAS and SCD1 decline in liver during the fasted state in rodents (Liang et al., 2002); this state is characterized by an inhibition of fatty acid synthesis and an activation of the fatty acid oxidation. However, Tobin et al. (Tobin et al., 2000) showed that fasting rats for 24 h increased the hepatic LXR mRNA and Matsuzaka et al. (Matsuzaka et al., 2002) observe no difference in either the hepatic D5D or D6D mRNA level between fasted and reefed mouse livers. Moreover, PUFA levels can be considered to be increased in liver following starvation because of the important lipolysis from adipose tissue as shown by Lee et al. in mice after 72 h fasting (Lee et al., 2004).

So we define our working data as results from a virtual experimentation on fasted animals compared to fed state and on which the hepatic mRNA of different genes could be quantified for example by DNA microarray analysis and the variation of hepatic PUFA could be measured by biochemical analysis. Hence, we assume that ACL, LXR, PPAR and PUFA increase while SREBP, ACC, FAS and SCD1 are supposed to decrease, in response to the fasting. We also assume that D5D and D6D are unchanged. By doing this, we voluntarily introduce a data that does not fit with results appearing in the literature: ACL should decline instead of inclining. This error is introduced to test whether the analyze is able to notice it and possibly correct it or not.

3. Steady state shift: qualitative description

We justify here the approach and impose some applicability limits. The basic mathematical object is a qualitative graph called the interaction graph. This graph represents the present knowledge on interactions for the biological process under study. Experimental data give variations of the concentrations of some molecules between two external conditions. The corresponding nodes of the interaction graph will be called *observed nodes*.

Differential model: The model describes the interactions between a set of molecules indexed by a set $I = \{1, \ldots, n\}$. The concentration of the molecule indexed by i is X_i. The quantity $\frac{dX_i}{dt}$ denotes the speed of variation of X_i.

We assume that the various concentrations involved in the model evolve according to a differential dynamics. The set of control parameters of the dynamics is denoted by **P**. The state vector $\mathbf{X} = (X_1, \ldots, X_n)$ satisfies a differential equation

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}, \mathbf{P}).$$

3.1. Interaction graph

3.1.1. Qualitative graph

The set of nodes $I \cup \{\mathcal{E}\}$ gathers all the molecules I in the model together with an extra generic node $\mathcal{E}$ to represent the exterior world.

The set of edges $A \subset (I \cup \{\mathcal{E}\}) \times (I \cup \{\mathcal{E}\})$ contains oriented edges joining j to i whenever $\frac{\partial F_i}{\partial X_j} \neq 0$. This description implies that the interaction graph contains no multi-arcs. Additionally, the set A contains edges from the exterior world $\mathcal{E}$ to I specifying which are the entrance nodes of the system. If there is an edge from $\mathcal{E}$ to $i \in I$, the node i is potentially affected by nodes that are not included in the model. If there is no edge between $\mathcal{E}$ and i, then the node i is supposed to be influenced only by I. Specifying these edges is an important modeling hypothesis.

The interaction graph $M = (I \cup \{\mathcal{E}\}, A, s)$ is signed by the sign function $s(j, i) = \text{sign}(\frac{\partial F_i}{\partial X_j}) \in \{+, -\}$ and weighted by the interaction coefficient $a(j, i) = \frac{\partial F_i}{\partial X_j}$. Notice that the sign function $s(j, i)$ is state dependent (it depends on the variables **X** and **P**). We suppose that $s(j, i)$ does not change within domains of experimental interest.

3.1.2. Exterior and parameters

The extra generic node $\mathcal{E}$ (exterior world) represents molecules that are not included in the model and that interact with the molecules in the model. **P** include external conditions such as food, temperature, stress, or internal conditions such as particularities of individual organisms, mutations, etc.

The only difference between $\mathcal{E}$ and **P** is dynamical. Indeed, $\mathcal{E}$ is rigorously a dynamical variable, while **P** is fixed during dynamics. From a static perspective $\mathcal{E}$ and the parameters **P** represent variables that may shift the steady state of the model. Throughout this paper we shall consider that there is no direct effect of **P** on the nodes of G: all influences arrive via the exterior $\mathcal{E}$. This modeling hypothesis applies to the analyzed example. It can be lifted with minor changes. If the modeler judges this assumption to be not reasonable, he can just add a (non-dynamical) variables **P** and connect them to the nodes that are directly influenced.

3.1.3. Extracting an interaction graph from biological facts described in the literature

Intuitively, an arc $j \to i$ describes the alteration of the speed of variation of the product i when the concentration of j is modified. We build the interaction graph associated to a given biological question from the biological literature. More precisely, we consider an edge from j to i in the graph when experimentation in a paper shows that the production of i is modified after a change in the concentration of j.

The interaction graph is supplied with negative self-interaction on all nodes of the graph, implying $s(i, i) = -$ for any i (in the absence of additional positive self-regulation). These result for instance from degradation processes or from growth induced dilution. It will be shown (Theorem 3.6 of Section 3) that negative self-interaction ensures the existence of a steady state.

Such an interaction graph might be incomplete: it is not necessary to gather all the biological facts in a model. Actually, we intend to understand whether experimental data fits with the theoretical model or not.

3.1.4. Example: fasting protocol (Fig. 1)

For the biological problem used as an illustration, we consider a model $M = (I \cup \{\mathcal{E}\}, A, s)$ with vertices $I = \{\text{PUFA, LXR, LXR - a, PPAR, PPAR-a, SREBP, SREBP-a, SCAP, ACL, ACC, FAS, SCD1, D5D, D6D}\}$. The interaction graph M is shown in Fig. 1. It is obtained by translating the biological information given in Section 2. As an example, since the active form of LXR (that is, LXR-a) is an activator of the transcription of SREBP, we consider an edge from LXR-a to SREBP, labeled by "+". We also consider a positive edge from SREBP to its active form SREBP-a since increasing the concentration of SREBP should increase the concentration of its active form. There is a positive edge from LXR-a to ACL because of the activation of ACL by the active form of LXR. Notice that this action is taken into account even if it is indirect: no detail on this action exists in the literature; a possible mechanism for this action involves SREBP as an intermediate, as described by the path LXR-a → SREBP → SREBP-a → ACL. However, since we do not know whether the action of LXR on ACL is only SREBP-mediated, the modeler has decided to keep an additional edge from LXR to ACL, redundant with the path containing SREBP.

3.1.5. Paths and loops on graph

A path $j \rightsquigarrow i$ is a sequence of nodes $\{i_1, i_2, \ldots, i_p\}$ such that $(i_k, i_{k+1}) \in A$, $i_1 = j$ and $i_p = i$. The nodes of a path are visited just once. Let the *path influence* $a_{j \rightsquigarrow i}$ denote the product of interaction coefficients along a path:

$$a_{j \rightsquigarrow i} = \prod_{k=1}^{k=p-1} a(i_k, i_{k+1}).$$

The sign of the path is $s(j \rightsquigarrow i) = \mathrm{sign}(a_{j \rightsquigarrow i})$.

A loop is a path with identical terminal nodes $i_1 = i_p$. In a loop all nodes are visited just once, excepting the terminal nodes.

3.1.6. Analyzed subgraph, observed nodes

The subset of the nodes in the model on which we perform the analysis is denoted by $G \subset I$. The nodes in G inherit the connections from M.

The set of observed nodes will be denoted as $\mathcal{O}$. One can define the nodes variations δX_i, $i \in \mathcal{O}$, given by experimental data on $\mathcal{O}$. The sign of these variations is denoted by $s(\delta X_i)$.

In our example, we consider $G = I$. Only some nodes are observed. The variations of observed nodes are shown in Fig. 1.

$$\mathcal{O} = \{\mathrm{PPAR}, \mathrm{LXR}, \mathrm{ACL}, \mathrm{ACC}, \mathrm{SREBP}, \mathrm{FAS}, \mathrm{SCD1}, \mathrm{PUFA}, \mathrm{D5D}, \mathrm{D6D}\}.$$

3.1.7. Entrance boundary

The analyzed subgraph G is naturally partitioned into two subsets: the entrance boundary of G and the interior nodes.

The entrance boundary of G, denoted $\top^{in} G$ is formed by all the nodes of G that are supposed to have entering connections from the outside:

$$\top^{in} G = \{i \in G \mid \exists j \in (I \cup \{\mathcal{E}\}) \setminus G,\ (j, i) \in A\}.$$

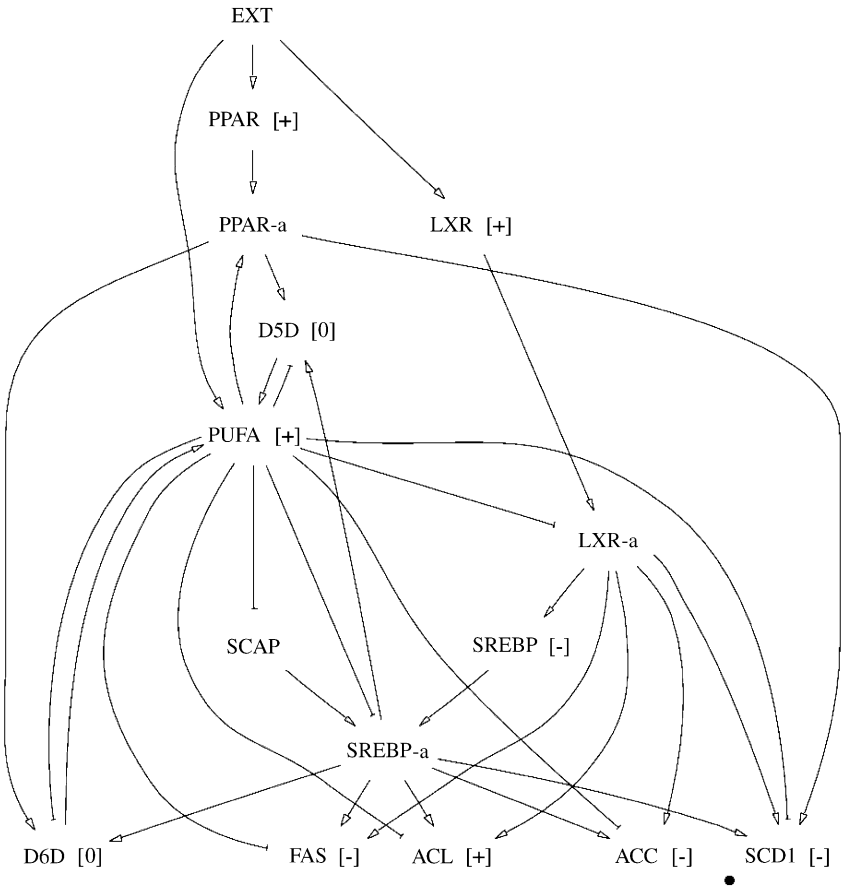


Fig. 1. Interaction graph A for a model $M = (I, A, s)$ of regulation of the synthesis of fatty acids. The node EXT stands for $\mathcal{E}$ and represents the exterior world. Self-regulation loops on nodes are omitted for sake of clarity. Observed variations are $s(\delta\mathrm{PPAR}) = +$, $s(\delta\mathrm{PUFA}) = +$, $s(\delta\mathrm{LXR}) = +$, $s(\delta\mathrm{SREBP}) = -$, $s(\delta\mathrm{SCD1}) = -$, $s(\delta\mathrm{FAS}) = -$, $s(\delta\mathrm{ACL}) = +$, $s(\delta\mathrm{ACC}) = -$, $s(\delta\mathrm{D5D}) = 0$, $s(\delta\mathrm{D6D}) = 0$.

The set of interior nodes is denoted $\mathring{G} = G \setminus \urcorner^{in} G$.

Notice that when $G = I$, the entrance boundary is entirely made of those nodes that have connections from $\mathcal{E}$.

In the model M of our working example, the analyzed subgraph is I. Its entrance boundary is $\urcorner^{in} I =$ {PPAR, LXR, PUFA}.

3.1.8. Predecessors

We denote by pred(i) the set of predecessors of i : $\text{pred}(i) = \{j \in I \cup \{\mathcal{E}\},\ j \neq i,\ (j,i) \in A\}$.

3.2. Qualitative linear equations

3.2.1. Steady state shift

In our approach we suppose that the waiting time after a change of the parameters is sufficient for the biological system to reach steady state.

Steady states are characterized by:

$$F(\mathbf{X}, \mathbf{P}) = 0. \tag{1}$$

The steady state shift is the result of a change in the control parameters $\mathbf{P}$. In our example fasting is the only control parameter. Experimental data are interpreted as variations, between two steady states, of the variables associated with observed nodes.

This point of view, which is assumed in the rest of the paper, is not correct in all situations. It cannot be assumed for time series where observations are made at relatively close instants. In such a situation, it is not true that the system has reached steady state at each instant when it is observed. For our example, this hypothesis is justified since data come from metabolism of animals that where kept under normal diet or fasted during a long time before the measurements. Furthermore, we should check carefully if the system is not capable of autonomous oscillations (limit cycles) in which case steady state again can not be reached.

3.2.2. Small variation and self-susceptivity

We differentiate Eq. (1) in order to understand how the steady state changes for small changes in the parameters. For a node $i \in I$ one gets the equation:

$$\frac{\partial F_i}{\partial X_i}\delta X_i + \sum_{j \neq i} \frac{\partial F_i}{\partial X_j}\delta X_j + \sum_k \frac{\partial F_i}{\partial P_k}\delta P_k = 0.$$

We can notice that if $\frac{\partial F_i}{\partial X_j} \neq 0$ and $j \neq i$ if and only if there is an edge from j to i in the interaction graph, that is, $j \in \text{pred}(i)$. Let us consider a node i such that $\frac{\partial F_i}{\partial P_k} = 0$: the production or the consumption of X_i does not depend directly on the parameters.

Property 3.1. *Let us consider a system whose steady state depends on the parameters. Let i be a node that satisfies the following conditions*:

- *the node i is not directly influenced by the parameters (all influences come via its predecessors), i.e.* $\frac{\partial F_i}{\partial P_k} = 0$,
- $\left(\frac{\partial F_i}{\partial X_i}\right) \neq 0$,

then the variation in i can be entirely calculated from the variations in the predecessors nodes:

$$\delta X_i = -\left(\frac{\partial F_i}{\partial X_i}\right)^{-1} \sum_{k \in \text{pred}(i)} a(k,i)\delta X_k. \tag{2}$$

Using a mechanical analogy, $\chi_i = -\left(\frac{\partial F_i}{\partial X_i}\right)$ can be called self-susceptivity: it is the ratio between the force produced by the predecessors of i and the variation in i. As discussed, there are always effects that produce self-susceptivity (degradation, growth induced dilution, self-regulation). The variation δX_i can be calculated from the force exerted by the predecessors only if the self-susceptivity χ_i is non-zero. A zero self-susceptivity would represent a non-generic case when the sum of all these effects cancel exactly.

3.2.3. Sign algebra

If we only know the signs of the variations and of the interaction coefficients, we intend to find the relations between the signs of the variations and the signs of the interaction coefficients. To do so, we transcribe Eq. (2) into the sign algebra. Remind that $s(\delta X_i)$ stands for the sign of the variation of X_i.

The sign algebra is defined as the set $\mathcal{S}$ of subsets of $\{0, +, -\}$, that is,

$$\mathcal{S} = \{\{0\}, \{+\}, \{-\}, \{0,+\}, \{0,-\}, \{+,-\}, \{0,+,-\}\}.$$

We denote by $? = \{0, +, -\}$ the subset standing for an undetermined sign.

The set $\mathcal{S}$ is then provided with the sum and product rules defined on singletons as follows:

$$\{+\} + \{-\} = ?,$$

$$\{+\} + \{+\} = \{+\}, \{+\} + \{0\} = \{+\},$$

$$\{-\} + \{-\} = \{-\}, \{-\} + \{0\} = \{-\},$$

$$\{+\} \times \{-\} = \{-\}, \{+\} \times \{+\} = \{+\},$$

$$\{-\} \times \{-\} = \{+\}, \{+\} \times \{0\} = \{0\},$$

$$\{-\} \times \{0\} = \{0\}.$$

These rules are extended to $\mathcal{S}$ by the union rule: $(A \cup B) + C = (A + C) \cup (B + C)$.

An equality in an equation means that the sign corresponding to the left hand side (l.h.s.) and the sign of the right hand side (r.h.s.) have a non-empty intersection (Kuipers, 1994; Travé-Massuyès and Dague, 2003).

3.2.4. Generalization to large variations

Eq. (2) represents a differential constraint that should be satisfied by (small) variations of the concentrations X_i. Naturally, we may ask whether there is also a constraint among large variations. This is equivalent to asking whether there is any globally defined function relating the value X_i to the concentrations of its predecessors that we denote by the vector $\hat{X}^{(i)}$. If there is no direct influence of the parameters $\mathbf{P}$ on X_i, stationarity of X_i reads:

$$F_i(X_i, \hat{X}^{(i)}) = 0 \tag{3}$$

We have the following.

Property 3.2. *If* $\frac{\partial F_i}{\partial X_i} < -C, \quad C > 0$, *if there is no direct influence of* $\mathbf{P}$ *on* X_i *and if* $F_i(0, \hat{X}^{(i)}) > 0$ *then there is a function* $\phi_i : \mathbb{R}_+^{n_i} \to \mathbb{R}$ (n_i *being the number of predecessors of* i) *such that* $X_i = \Phi_i(\hat{X}^{(i)})$. *The differential of* Φ_i *is*:

$$\mathrm{d}\Phi_i = -\left(\frac{\partial F_i}{\partial X_i}\right)^{-1} \sum_{j \in \mathrm{pred}(i)} \frac{\partial F_i}{\partial X_j} \mathrm{d}X_j.$$

Now, let us suppose that the steady state shifts from the vector X^1 to the vector X^2 and that the projection of the steady state onto the predecessors of i shifts from $\hat{X}^{(i,1)}$ to $\hat{X}^{(i,2)}$. In order to obtain the variation of X_i it is enough to integrate the differential of Φ_i on a segment $C^{1,2}$ connecting $\hat{X}^{(i,1)}$, $\hat{X}^{(i,2)}$:

$$\Delta X_i = \int_{C^{1,2}} -\left(\frac{\partial F_i}{\partial X_i}\right)^{-1} \sum_{j \in \mathrm{pred}(i)} \frac{\partial F_i}{\partial X_j} dX_j. \tag{4}$$

3.2.5. Qualitative system of equations

Transcribing Eq. (4) in the sign algebra provides the following result.

Theorem 3.3. *Supposing that the conditions of* Property 3.2 *are satisfied and supposing that the signs of* $\frac{\partial F_i}{\partial X_j}(\Phi_i(\hat{X}^{(i)}), \hat{X}^{(i)})$ *are constant on the segment* $C^{1,2}$, *then the signs of finite variations of steady state variables are connected by*:

$$s(\Delta X_i) = \sum_{k \in \mathrm{pred}(i)} s(k, i) s(\Delta X_k). \tag{5}$$

Remark. Although the ends of the segment $C^{1,2}$ are steady states there is no reason for the interior points to fulfill stationarity equations (they fulfill only one of the stationarity equations, namely Eq. (3)). Nevertheless, supposing that the absolute values of the derivatives $\frac{\partial F_i}{\partial X_j}$ are bounded (which is a natural modeling assumption) it follows from Eq. (4) that the excursion of Φ_i and therefore of X_i is bounded when $\hat{X}^{(i)}$ describes the segment $C^{1,2}$. This means that the segment $C^{1,2}$ corresponds to a range of states reasonably close to the experimental range, thus justifying the modeling assumption concerning the signs. This reasoning warns against the naive interpretation of interaction signs that assumes constancy of these signs over arbitrary ranges of the concentrations. The problem is not trivial and should be considered carefully.

Example. Theorem 3.3 provides the following equations for the working example M (for ease of notations, $s(\Delta A)$ is denoted as A). One can apply Theorem 3.3 because the self-susceptivities are positive in the absence of positive self-regulation. Nodes i that satisfy the hypothesis of Theorem 3.3 are those which are not connected to the exterior world $\mathcal{E}$ (implying that the first condition is satisfied).

System 1

```
(1)  PPAR-a = PPAR + PUFA
(2)  LXR-a = -PUFA + LXR
(3)  SREBP = LXR-a
(4)  SREBP-a= SREBP + SCAP -PUFA
(5)  ACL= LXR-a  + SREBP-a - PUFA
(6)  ACC= LXR-a + SREBP-a - PUFA
(7)  FAS=  LXR-a + SREBP-a - PUFA
(8)  SCD1 = LXR-a  + SREBP-a - PUFA + PPAR-a
(9)  SCAP=  -PUFA
(10) D5D= PPAR-a + SREBP-a - PUFA
(11) D6D = PPAR-a + SREBP-a - PUFA
```

Known values are the following:

Values 1

```
PPAR=+, PUFA=+, LXR=+, SREBP=-, SCD1=-,
FAS=-, ACL=+, ACC=-, D5D=0, D6D=0
```

3.3. Influences and their transmission across a boundary

In order to analyze data, we need to transform the qualitative system derived in Theorem 3.3 to obtain a system where the variables are functions of a fixed set of variables. In general, this kind of system transformation in the sign algebra is a NP-complete problem (Dormoy, 1988).

In the following paragraphs we show that quantitatively this operation is equivalent to solving the Dirichlet problem for a discrete operator on the interaction graph. A solution to the Dirichlet problem can be found under mild conditions. We use this result for deriving a large class of qualitative systems.

3.3.1. Transmission of influences on graphs

One can read Eq. (2) as the propagation of influence on the interaction graph: the concentration variation of the molecule i is an average of the influences from its predecessors. Let us define as follows the discrete operator $L : \mathbb{R}^I \to \mathbb{R}^I$:

$$(L\delta X)_i = \delta X_i + \left(\frac{\partial F_i}{\partial X_i}\right)^{-1} \sum_{k\in \mathrm{pred}(i)} a(k,i)\delta X_k. \qquad (6)$$

Then Eq. (2) reads $L\delta X = 0$ which is analogous to the Laplace equation on graphs. Let us remind that the Laplace equation fulfills an analogous property: the value of an harmonic function (satisfying the Laplace equation) in a node is the average of the values of the function on the predecessors of this node.

The analogy can be pushed further as one may try to define the influence of the boundary on an interior point (question known as the Dirichlet problem) (Chung and Yau, 2000).

Laplace equations on graphs were intensively studied in connection to electrical networks and random walks on undirected graphs (Soardi, 1994). Let us notice that our model describes propagation of influence on oriented graphs. This could be related to random walks on oriented graphs. Nevertheless, the propagation operators resulting from random walks have negative non-diagonal elements (Campanino and Petritis, 2003), while in our case there are no such constraints on the signs of the interactions. Little is known on the properties of random walks on oriented graphs. For instance it is not known whether the solution to the Dirichlet problem has a representation in terms of stopped random walks or not. The connection between propagation of influences and random walks on oriented graphs could thus be profitable both for biology and for applied mathematics.

3.3.2. Influence of the boundary on the interior

From now, we fix a subnetwork $G \subset I$ of the model. We have already defined the entrance boundary $\urcorner^{in} G$, formed by all the nodes of G that are supposed to have entering connections from the outside. The set of interior nodes was denoted $\mathring{G} = G \setminus \urcorner^{in} G$.

By differentiating the components of Eq. (1) corresponding only to the interior nodes of G one gets for every $i \in \mathring{G}$:

$$\sum_{j\in\mathring{G}} \frac{\partial F_i}{\partial X_j}\delta X_j + \sum_{k\in\urcorner^{in}G} \frac{\partial F_i}{\partial X_k}\delta X_k + \sum_{l\in I\setminus G} \frac{\partial F_i}{\partial X_l}\delta X_l + \frac{\partial F_i}{\partial P}\delta P = 0.$$

Noticing that $\frac{\partial F_i}{\partial X_l} = 0$ for all $i \in \mathring{G}, l \in I \setminus G$ and assuming that $\frac{\partial F_i}{\partial P} = 0$ for all $i \in \mathring{G}$ it follows:

$$\forall i \in \mathring{G}, \sum_{j\in\mathring{G}} \frac{\partial F_i}{\partial X_j}\delta X_j = -\sum_{k\in\urcorner^{in}G} \frac{\partial F_i}{\partial X_k}\delta X_k. \qquad (7)$$

Eq. (7) is the starting point for the proof of the following theorem (details can be found in the Appendix A):

Theorem 3.4. (Linear Dirichlet solution) *Let G be a subgraph of I. Let $\mathring{A}$ be the restriction of the jacobian of F to the interior nodes: $\mathring{A}_{ij} = \frac{\partial F_i}{\partial X_j}, i, j \in \mathring{G}$. Let us suppose that*:

- $\det(\mathring{A}) \neq 0$,
- *there is no direct influence of the parameters on the interior nodes of G, i.e. $\frac{\partial X_i}{\partial P_k} = 0, \forall i \in \mathring{G}$.*

Then, the response of $i \in \mathring{G}$ to small changes on the boundary of G is given by:

$$\delta X_i = \sum_{j\in\urcorner^{in}G} \sum_{j\rightsquigarrow i\in\mathcal{P}^{\circ}_{G}} \frac{a_{j\rightsquigarrow i}}{\mathring{C}_{j\rightsquigarrow i}}\delta X_j, \qquad (8)$$

with the following notations:

- $\mathcal{P}^{\circ}_{G}$ *denotes the set of paths included in G, starting on the boundary and that do not return to the boundary*: $\mathcal{P}^{\circ}_{G} = \{j\rightsquigarrow i = (i_1 = j, \ldots, i_k, \ldots, i_p = i), i_k \notin \urcorner^{in}G, \forall k > 1\}$;
- $a_{j\rightsquigarrow i} = \prod_{k=1}^{p-1} a(i_k, i_{k+1})$;
- $\mathring{C}_{j\rightsquigarrow i} = (-1)^{l_{k(j)\rightsquigarrow i}+1} \frac{\det\mathring{A}}{\det\mathring{A}_{\{k(j)\rightsquigarrow i\}^c}}$ *denotes the "path modulus" of the internal path $k(j)\rightsquigarrow i$ where*
  - $k(j) \in \mathring{G}$ *is the second node after j of the path $j\rightsquigarrow i$;*
  - $l_{k(j)\rightsquigarrow i}$ *the length of the path $k(j)\rightsquigarrow i$, i.e. the number of arcs*;
  - $\det\mathring{A}_{\{k(j)\rightsquigarrow i\}^c}$ *is the principal minor obtained by eliminating all the lines and all the columns whose indices are in the path $k(j)\rightsquigarrow i$ from the jacobian $\mathring{A}$. Conventionally, if the resulting minor is empty we choose it equal to one*: $\det\mathring{A}_{\mathring{G}^c} = 1$.

If $k(j) = i$ the path modulus becomes the node modulus $\mathring{C}_i = -\frac{\det \mathring{A}}{\det \mathring{A}_{\{i\}^c}}$.

If we consider $G = \{i\} \cup \mathrm{pred}(i)$ to be the set of a predecessors of a node i we obtain the Property 3.1 as a corollary.

3.4. Moduli and sign algebra

3.4.1. Path modulus

Eq. (8) describes the influence on an interior node of variations on the boundary. j acts with a force $a_{j\rightsquigarrow i}\delta X_j$ *on the node i, along the path* $j\rightsquigarrow i$. This force is bigger when the product of interaction coefficients $\frac{\partial F_{i_{k+1}}}{\partial X_{i_k}}$ along the path is bigger.

$\mathring{C}_{k(j)\rightsquigarrow i}$ is the ratio force/response and therefore can be called *path modulus*. According to Theorem 3.4, the contribution to the modulus comes only from the internal part of the path: $k(j)\rightsquigarrow i$. If the internal part reduces to a point ($k(j) = i$, meaning that the path from j to i is a direct arc), then the contribution to the modulus comes from the node i only: the path modulus becomes node modulus. A large path (or node) modulus implies a small response at the end of the path, even if the force is big. Therefore, the modulus can be related to sensitivity.

3.4.2. Signs of moduli

One way to test the compatibility between microarray data and interaction models is to compare predicted and observed signs of the variations. In order to do so, we need to know the signs of the path moduli. The following property is very useful:

Property 3.5. *If the interior of the subgraph G contains no positive loops, then all internal path moduli* $\mathring{C}_{k(j)\rightsquigarrow i}$ *and node moduli* $\mathring{C}_i$ *in Theorem 3.4 are strictly positive.*

If G has no positive self-loops, than all the self-susceptivities χ_i *are positive.*

Following a reasoning similar to the one leading to Property 3.2 and to Theorem 3.3, and using Property 3.5 and Theorem 3.4, we can write down (under some conditions) linear qualitative equations valid for large variations.

Let $\mathring{X}_G$,$\hat{X}_G$ be the set of variables internal to G, and on the entrance boundary, respectively. If the parameters P have direct influence on the interior of G, then the stationarity equations for G (*non-linear Dirichlet problem*) read:

$$F_G(\mathring{X}_G, \hat{X}_G) = 0 \tag{9}$$

where F_G is the restriction of F to G.

Let us suppose that steady state shifts from X^1 to X^2 and that the projection of the steady state onto the boundary variables shifts from $\hat{X}^1$ to $\hat{X}^2$.

Then, we obtain (see Appendix A) the following.

Theorem 3.6. *Suppose that*

- *the parameter variations have no direct influence on the interior of G, i.e.* $\frac{\partial F_i}{\partial P_k} = 0, \forall i \in \mathring{G}$;
- *the restriction of the Jacobian of F to* $\mathring{G}$ *is non-singular*;
- *the interaction graph has no positive loops in the interior of G*;
- *the restriction of F to G has the form*

$$F_G = \psi^G(\mathring{X}_G, \hat{X}_G) - \Lambda_G \mathring{X}_G \tag{10}$$

where ψ^G *is bounded and satisfies*

$$\psi_i^G(\ldots, X_i = 0, \ldots, \hat{X}_G) > 0, \forall i \in \mathring{G} \tag{11}$$

and Λ_G *is a diagonal matrix with positive entries*;

Then there is a function Φ_G *such that* $\mathring{X}_G = \Phi_G(\hat{X}_G)$ *gives the unique solution of the non-linear Dirichlet problem* (Eq. (9)).

If furthermore the signs $s(j\rightsquigarrow i)$ *of the path influences from the boundary to the interior point i are constant for* $\mathring{X}_G$ *on a segment connecting* $\hat{X}_G^1$, $\hat{X}_G^2$ *and* $\mathring{X}_G = \Phi_G(\hat{X}_G)$, *then the signs of the finite variations of the internal variables* X_i *satisfy the following relation for every* $i \in \mathring{G}$:

$$s(\Delta X_i) = \sum_{j \in \overset{\neg}{}^{in} G} \sum_{j\rightsquigarrow i \in \mathcal{P}^{\circ}_G} s(j\rightsquigarrow i) s(\Delta X_j). \tag{12}$$

Remark. The decomposition in Eq. (10) corresponds to a regulation term ψ^G and to a linear dissipation term $\Lambda_G \mathring{X}_G$. The condition Eq. (11) naturally confines the dynamics to positive concentrations. A remark similar to the one for Theorem 3.3 holds as to the boundedness of the variations of $\mathring{X}_G$ when the boundary variables describe the segment from $\hat{X}^1$ to $\hat{X}^2$.

4. Model and data compatibility

In this section we introduce two methods of qualitative analysis of biological systems and experimental data. These methods use the systems of qualitative equations generated by Theorem 3.3.

Nevertheless, Theorem 3.6 can be used to generate other systems of qualitative equations. Indeed, it is enough to consider any covering of the network I by subgraphs $\{G_j\}_{j=1,p}$. Then, the Eq. (12) written for all

nodes $i \in G_j$ and for all subgraphs G_j provide another system of qualitative equations.

Hence, even if algorithms for incompatibility are presented for system 1 in this section, they can be easily adapted for another system of qualitative equations as well.

4.1. Solving qualitative systems

Let us recall the qualitative equations provided by Theorem 3.3.

System 1

```
(1)  PPAR-a = PPAR + PUFA
(2)  LXR-a = -PUFA + LXR
(3)  SREBP = LXR-a
(4)  SREBP-a= SREBP + SCAP -PUFA
(5)  ACL= LXR-a  + SREBP-a - PUFA
(6)  ACC= LXR-a + SREBP-a - PUFA
(7)  FAS=  LXR-a + SREBP-a - PUFA
(8)  SCD1 = LXR-a  + SREBP-a - PUFA + PPAR-a
(9)  SCAP=  -PUFA
(10) D5D= PPAR-a + SREBP-a - PUFA
(11) D6D = PPAR-a + SREBP-a - PUFA
PPAR=+, PUFA=+, LXR=+, SREBP=-, SCD1=-,
FAS=-, ACL=+, ACC=-, D5D=0, D6D=0
```

4.1.1. Compatible system

A valuation of the qualitative system is called compatible if it satisfies:

1. all nodes are determined (have values different from ?);
3. all the observed nodes are given the observed values;
3. the qualitative Eq. (5) are fulfilled.

Solving the qualitative system means finding all compatible valuations. If there is no compatible valuation, the system is said to be incompatible.

The interest of finding compatible valuations lies in the following property: any solution of the quantitative system provides a compatible valuation. Consequently, the incompatibility of the qualitative system implies that the quantitative system is incompatible as well. This immediately indicates a contradiction between the model and the observations. Furthermore, the signs corresponding to a solution of the quantitative system should be given by one of the solutions of the qualitative system.

Let us notice that when the system is compatible, the number of compatible valuations can be big and this could bring very limited information on the biological system. In general we would like to know which kind of information one can extract from compatible valuations. In order to do this an useful concept is the following notion of competition.

4.1.2. Competitions

We say that there is a competition or a balance on a node i if the right side of the Eq. (5) is equal to "?". The biological interpretation of this indetermination is that the node i is submitted to competitive actions of different signs. An example is provided by the node LXR-a in the working example, that satisfies:

$$\text{PUFA} = +,$$
$$\text{LXR} = +,$$
$$\text{LXR-a} = -\text{PUFA} + \text{LXR} = ?$$

Determining the sign of the balanced node says which paths win the competition. If the variation of the balanced node is zero we say that the competition is neutral. If this variation is negative (positive) we say that the negative (positive) paths win the competition, or that the competition leans toward the negative (positive) direction. The information on competitions has biological interest, because it replaces undetermined actions by determined ones and facilitates the understanding of the interaction graph.

In the next subsections we use simple rules of qualitative algebra to solve System 1 and extract information on competitions. Then, we propose an algorithmic approach to solve the system obtained from Theorem 3.3 by using these rules systematically. Algorithmic methods related to competitions are detailed in Section 5.

4.2. Hand solving of qualitative equations

Let us solve System 1 by using the rules of the sign algebra and the following computation rule:

Property 4.1. *If $A = X + B$ and if the sign of A is opposite to the sign of B, then the sign of X is determined and it is equal to the sign of A.*

The result of the application of these rules to System 1 is summarized as following.

Property 4.2. *System 1 is not compatible with the observed variations.*

Simple change of the sign of ACL (ACL$= -$) *renders it compatible and determined. The values of non-observed variables are* PPAR-a $= +$, LXR-a $= -$, SCAP $= -$, SREBP-a $= -$.

In this case there are competitions on LXR-a, SCD1 *that are observed to lean towards the negative directions. There are also competitions on* D5D, D6D

that are observed to be neutral. There is no competition on FAS, ACC *and* ACL.

Proof. From Eq. (1) PPAR-a = +. From Eq. (3) LXR-a = SREBP = −. From Eq. (9) SCAP = −. From Eq. (4) SREBP-a = −.

We have already some biologically interesting results: by Eq. (2), LXR-a is supposed to be undetermined, that is, the result of a competition between the positive action of LXR and the negative action of PUFA. Using Eq. (3) we have shown that the competition on LXR-a should lean toward the negative direction. From Eq. (8) it follows that there is a competition on SCD1. Data on SCD1 implies that this competition leans toward the negative direction.

From Eq. (5) and Property 4.1, SREBP-a = +. This is inconsistent with the previously found value, which proves that the system is incompatible. There is necessarily an error, either in the interactions or in the data. Let us assume that the interactions are correct. Then, considering that data on ACL is wrong is the simplest way to make the system compatible (because Eqs. (6) and (7) do not force signs of FAS and ACC).

With the correction ACL = −, the unknown values are determined in the same way as before. From Eq. (1) PPAR-a = +. From Eq. (3) LXR-a = SREBP = −. From Eq. (9), SCAP = −. From Eq. (4) SREBP-a = −. Now Eq. (5) is satisfied with no competition. The same is true for Eqs. (6) and (7): there are no competitions on ACC and FAS. From Eqs. (2) and (8) we find that there are competitions on LXR-a, SCD1 which leans towards the negative direction. From Eqs. (10) and (11) we find that there are competitions on D5D, D6D. The observed variations of D5D, D6D are vanishing, therefore these competitions are neutral. □

Manual resolution of a qualitative system is only possible for small examples. In the next section we describe an algorithm that allows the systematic application of these rules to large systems.

4.3. Graph valuation algorithm

4.3.1. Basic ideas underlying the algorithm

A natural way to reduce the number of unknowns in a system of equations is to proceed by elimination. This means to compute the value of one unknown from an equation and then substitute this value in all the other equations. Unfortunately, in the sign algebra, the usual elimination rules do not apply (qualitative equality of undetermined quantities is not transitive).

In the proposed algorithm, we will rely on the rules used in Section 4.2. Let us formulate them more precisely.

- Eq. (5) implies that the sign of ΔX_i is determined if $s(k,i)s(\Delta X_k)$ are determined and have the same value for all the predecessors k of the node i. This first elementary rule is used in the forward propagation step of the algorithm.
- The second rule follows from Property 4.3. Suppose that

$$\Delta X_i = s(k,i)s(\Delta X_k) + \sum_{l\neq k} s(l,i)s(\Delta X_l).$$

If both the signs of $\sum_{l\neq k} s(l,i)s(\Delta X_l)$ and $s(\Delta X_i)$ are determined and if they are opposite then $s(k,i)s(\Delta X_k) = s(\Delta X_i)$. The second rule is used in the backward propagation step of the algorithm.

In the case of a sign algebra with 0 value, the first rule is unchanged while the second one needs a straightforward adaptation.

4.3.2. The graph valuation algorithm

1. *Initialization*: All the nodes are given a value in {+, −, ?}. Observed nodes are initialized with the value corresponding to the observed variations of the concentration of the molecules. The other nodes are initialized with the value ?. The value associated with a node i is denoted val(i).

   Let us consider two sets of nodes U and T. The nodes with values in {+, −} are put in T. The remaining nodes (with value "?") are put in U.
2. *Forward propagation*: Create a set F that contains every node of U that is also a successor of a node in T.

   Repeat: Find a node i in F which has at least a predecessor and such that for all predecessors j of i, $s(j,i)\text{val}(j) = +$, or for all predecessors j of i, $s(j,i)\text{val}(j) = -$. Then
   - set $\text{val}(i) = s(j,i)\text{val}(j)$,
   - remove i from F,
   - remove the successors of i which are in U and put them in F,
   - put i in T.

   until no node found
3. *Backward propagation*: Create a set B that contains every node of U that is also a predecessor of a node in T.

   Repeat: Find a node j in B which has a successor i in T, such that $\text{val}(i) \times \sum_{k\in\text{pred}(i),k\neq j} s(k,i)\text{val}(k) = -$ or $\sum_{k\in\text{pred}(i),k\neq j} s(k,i)\text{val}(k) = 0$. Then

- val(j) = $s(j, i)$val(i),
- remove j from B,
- remove the predecessors of j which are in U and put them in B,
- put j in T.

until no node found

4. *Alternate propagations.* Alternate forward and backward propagations until no new value is added to T.

Notice that each time that a node is valuated, it is moved to T and it never returns to F or B. T is an increasing subset of a finite set, so that the algorithm stops.

4.3.3. Application to the detection of incompatibilities

All the valuations introduced by the algorithm are unique sign choices compatible with the equations provided by Theorem 3.3. By doing so, no compatible valuation is lost. More precisely:

Property 4.3. *The graph valuation algorithm provides a new graph such that any compatible valuation of the nodes in the initial graph is a compatible valuation of the new graph.*

Nevertheless, nothing guarantees that a compatible valuation exists, i.e. that all Eq. (5) are satisfied. The resulting graph may contain incompatibilities. These incompatibilities are detected simply by checking Eq. (5) on each node.

4.3.4. Criterion to compare model and data 1

If the graph output by the algorithm described below contains a node where Eq. (5) is not s atisfied, then experimental data and the model are incompatible.

4.3.5. Working example

In the example M the initialization is

$$\text{PUFA} = +, \text{LXR} = +, \text{LXR-a} = ?, \text{PPAR} = +,$$
$$\text{PPAR-a} = ?, \text{SREBP} = -, \text{SREBP-a} = ?,$$
$$\text{ACL} = +, \text{ACC} = +, \text{FAS} = -, \text{SCD1} = -,$$
$$\text{SCAP} = ?, \text{D5D} = ?, \text{D6D} = ?$$

We assume that the variations of D5D and D6D are not known.

The forward propagation begins with

$$F = \{\text{PPAR-a}, \text{LXR-a}, \text{SREBP-a}, \text{SCAP}, \text{D6D}, \text{D5D}\}$$

and ends with $F = \emptyset$, PPAR-a $= +$ (from PUFA $= +$ and PPAR $= +$), SREBP-a $= -$ (from SREBP $= -$) and SCAP $= -$ (from PUFA $= +$). D5D and D6D are removed from F and keep their value ?

The backward propagation begins with $B = \{\text{LXR-a}\}$ and ends with $B = \emptyset$, LXR-a $= -$ (from SREBP $= -$), or LXR-a $= +$ (from ACL $= +$, SREBP-a $= -$ and PUFA $= +$). The second forward propagation begins with $F = \{\text{D6D}, \text{D5D}\}$ and stops with $F = \emptyset$ but no new value for D5D or D6D.

After running the algorithm, if LXR-a$= -$ Eq. (5) is not satisfied at node ACL and if LXR-a$= +$ Eq. (3) is not satisfied at node SREBP. This shows a contradiction between the model and the values assigned to some nodes by experimental data or hypothesis.

The graph valuation algorithm is not complete for the resolution of linear qualitative systems. As a consequence, the algorithm concludes only when it detects an incompatibility or when all the values are determined.

5. Assessment of competitions

The graph valuation algorithm aims at detecting incompatibilities or producing compatible valuations. Nevertheless, even when these tasks are successful the result may be difficult to use. If for instance data is wrong for a node that happens to be very connected to the rest of the graph, then many incompatibilities appear and it is difficult to localize and to correct the error. Also, if there are not enough observations then compatible valuations are produced in huge number. It is difficult to extract useful biological information from this situation.

In this section, we develop another approach, that exploits the global qualitative information on the propagation of influences on the graph. This approach is complementary to the graph valuation algorithm since it identifies competition processes and the places where these occur.

5.1. Competitions

As already explained, we say that there is a competition or a balance on a node i if the right side of the Eq. (5) is equal to "?". The biological interpretation of this indetermination is that the node i is submitted to competitive actions of different signs.

This definition can be extended to any equation connecting i to other nodes in the graph instead of Eq. (5) (related to the predecessors of a node i). Such equations are provided by Theorem 3.6 that can be applied to every subgraph G containing i. Then, we say that there is a competition on a node i when there exists an equation whose left side is $s(\delta X_i)$ and whose right side is equal to "?".

5.2. Set of equations describing the sign of variation of a variable X_{i_0}

Let $M = (I \cup \{\mathcal{E}\}, A, s)$ be a network with no positive loop in the interior of I and such that the sign function s is constant within the entire range of parameter variations. Let us focus on a variable X_{i_0} with index i_0 in the model M. From Theorem 3.6, as soon as X_{i_0} is an internal variable of a subgraph G of M, the sign of variation of X_{i_0} can be expressed as a combination of the signs of paths starting on the boundary of G and never coming back to it:

$$\forall G \subset I, X_{i_0} \in \mathring{G}, s(\Delta X_{i_0}) \sum_{\substack{j \in \gamma^{in} G \\ j \rightsquigarrow i_0 \in \mathcal{P}_G^{\circ}}} s(j \rightsquigarrow i_0) s(\Delta X_j). \tag{13}$$

The *set of equations describing the variable* X_{i_0} is the set of Eq. (13) for all subgraphs G such that X_{i_0} is an internal variable of G.

5.3. Example of SREBP

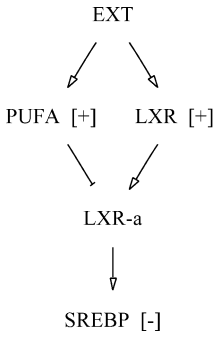
In the working example of lipogenesis, let us consider the variable $X =$ SREBP. We want to determine all the subgraphs G that contain X as an internal variable. A necessary and sufficient condition is that G contains all the predecessors of X, that is, LXR-a. In each subgraph G, SREBP is determined by all the paths that start from the boundary of G and never go back to it. Notice that any path arriving in SREBP contains LXR-a. If it is longer than the path LXR-a → SREBP, it must contain PUFA or LXR, that both belong to the boundary of $\mathring{G}$ (since they are connected to the exterior). Hence, there are three paths to SREBP that start from the boundary of such subgraphs G and never come back to it. They are given in Table 1, together with the set of equations associated to SREBP.

5.4. Paths appearing in an Eq. (13)

Let $\mathcal{P}(M)$ contain all paths of M that appear in an equation (13). Hence, a path $j \rightsquigarrow i$ belongs to $\mathcal{P}(M)$ if there exists a subgraph $G \subset I$ such that j belongs to the boundary of G and $j \rightsquigarrow i$ is an internal path of G. The reasoning detailed in the example allows to characterize this set. A full proof is given in the Appendix A.

Property 5.1. *A path $j \rightsquigarrow i \subset I$ belongs to $\mathcal{P}(M)$ iff* $\mathrm{pred}(j) \neq \emptyset$ *and no node of $j \rightsquigarrow i$ except possibly j is connected to the exterior $\mathcal{E}$.*

Table 1
System of qualitative equations describing the variations of SREBP; sign of influences associated to each path



(P_1) LXR-a → SREBP,
(P_2) PUFA → LXR-a → SREBP,
(P_3) LXR → LXR-a → SREBP.

$$s(\Delta \mathrm{SREBP}) = \underset{(-)}{s(P_1)s(\Delta \mathrm{LXR\text{-}a})} \tag{14}$$

$$s(\Delta \mathrm{SREBP}) = \underset{(-)}{s(P_2)}\underset{(-)}{s(\Delta \mathrm{PUFA})} + \underset{(+)}{s(P_3)s(\Delta \mathrm{LXR})} \tag{15}.$$

The graph is restricted to nodes having an influence on SREBP such that this influence possibly starts from the boundary (PPAR, PUFA, LXR) but does not come back to it.

If so, there exists a subgraph G such that the variations of X_i satisfy:

$$s(\Delta X_i) = s(\Delta X_j)s(j \rightsquigarrow i) + \sum_{\substack{k \in \gamma^{in} G \\ k \rightsquigarrow i \in \mathcal{P}_{\mathring{G}} \\ k \rightsquigarrow i \neq j \rightsquigarrow i}} s(k \rightsquigarrow i) s(\Delta X_k).$$

5.5. Example of SREBP-a

Let us fix $X =$ SREBP-a. As an application of Property 5.1 seven paths start from the boundary of a graph G such that $X \in \mathring{G}$ and the path never come back to the boundary. They are given in Table 2 together with the equations describing SREBP-a. Notice that Eqs. (19)–(21) correspond to Eqs. (16)–(18) in which $s(P_7)s(\Delta \mathrm{PUFA})$ has been replaced by $s(P_{10})s(\Delta \mathrm{SCAP})$. This corresponds to the particular case of SCAP, whose only predecessor is PUFA, so that SCAP can be considered either on the boundary or in the interior of a subgraph G, providing two different equations.

5.6. Influences

Property 5.1 suggests the following definition. For a path $j \rightsquigarrow i \in \mathcal{P}(M)$ we say that $s(\Delta X_j)s(j \rightsquigarrow i)$ is the *influence of $j \rightsquigarrow i$ on the variable X_i*.

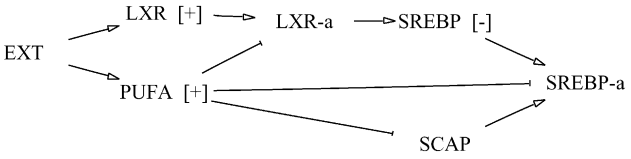
Let us also define the sets of positive and negative influences acting on X_i, denoted by $\mathcal{I}^+(i)$ and $\mathcal{I}^-(i)$.

$$\mathcal{I}^+(i) = \{j \rightsquigarrow i \in \mathcal{P}(M),\ j \in \mathcal{O},\ s(\Delta X_j)s(j \rightsquigarrow i) = +\}$$
$$\mathcal{I}^-(i) = \{j \rightsquigarrow i \in \mathcal{P}(M),\ j \in \mathcal{O},\ s(\Delta X_j)s(j \rightsquigarrow i) = -\}$$

The set of all influences is denoted by $\mathcal{I}(i)$.

$$\mathcal{I}(i) = \mathcal{I}^+(i) \cup \mathcal{I}^-(i).$$

Table 2
System of qualitative equations describing the variations of SREBP-a; sign of influences associated to each path



(P_4) PUFA $\rightarrow$ SREBP-a,
(P_5) $SREBP \rightarrow$ SREBP-a,
(P_6) PUFA $\rightarrow$ LXR-a $\rightarrow$ SREBP $\rightarrow$ SREBP-a,
(P_7) PUFA $\rightarrow$ SCAP $\rightarrow$ SREBP-a,
(P_8) LXR-a $\rightarrow$ SREBP $\rightarrow$ SREBP-a,
(P_9) LXR $\rightarrow$ LXR-a $\rightarrow$ SREBP $\rightarrow$ SREBP-a,
(P_{10}) SCAP $\rightarrow$ SREBP-a.

$$s(\Delta\text{SREBP-a}) = \underset{(-)}{s(P_4)s(\Delta\text{PUFA})} + \underset{(-)}{s(P_5)s(\Delta\text{SREBP})} + \underset{(-)}{s(P_7)s(\Delta\text{PUFA})} \quad (16)$$

$$s(\Delta\text{SREBP-a}) = s(P_8)s(\Delta\text{LXR-a}) + \underset{(-)}{s(P_4)s(\Delta\text{PUFA})} + \underset{(-)}{s(P_7)s(\Delta\text{PUFA})} \quad (17)$$

$$s(\Delta\text{SREBP-a}) = \underset{(+)}{s(P_9)s(\Delta\text{LXR})} + \underset{(-)}{s(P_4)s(\Delta\text{PUFA})} + \underset{(-)}{s(P_6)s(\Delta\text{PUFA})} + \underset{(-)}{s(P_7)s(\Delta\text{PUFA})} \quad (18)$$

$$s(\Delta\text{SREBP-a}) = \underset{(-)}{s(P_4)s(\Delta\text{PUFA})} + \underset{(-)}{s(P_5)s(\Delta\text{SREBP})} + s(P_{10})s(\Delta\text{SCAP}) \quad (19)$$

$$s(\Delta\text{SREBP-a}) = s(P_8)s(\Delta\text{LXR-a}) + \underset{(-)}{s(P_4)s(\Delta\text{PUFA})} + s(P_{10})s(\Delta\text{SCAP}) \quad (20)$$

$$s(\Delta\text{SREBP-a}) = \underset{(+)}{s(P_9)s(\Delta\text{LXR})} + \underset{(-)}{s(P_4)s(\Delta\text{PUFA})} + \underset{(-)}{s(P_6)s(\Delta\text{PUFA})} + s(P_{10})s(\Delta\text{SCAP}) \quad (21)$$

5.7. Examples of SREBP and SREBP-a

The observed variations are $\Delta\text{PUFA} = +$, $\Delta\text{SREBP} = +$ and $\Delta\text{LXR} = +$, so that the positive and the negative influences are

$$\mathcal{I}^+(\text{SREBP}) = \{P_3\}$$
$$\mathcal{I}^-(\text{SREBP}) = \{P_2\}$$
$$\mathcal{I}^+(\text{SREBP-a}) = \{P_9\}$$
$$\mathcal{I}^-(\text{SREBP-a}) = \{P_4, P_5, P_6, P_7\}.$$

The influence of paths starting from LXR-a and SCAP is not taken into account since the variations of these molecules are not observed.

5.8. Counterbalanced influences

In Table 2, the right sides of Eqs. (15), (16) and (18) are undetermined since they are the sums of at least two terms of different signs. This means that there are competitions on SREBP and SREBP-a. The observation of SREBP shows that the competition between P_2 and P_3 leans toward the negative direction: we say that the influence of P_3 is counterbalanced by the influence of P_2. Since SREBP-a is not observed, we can say that the influence of P_9 is counterbalanced by the influence of P_4, P_5, P_6, P_7, but we do not know the inclination of the competition.

This example justifies the following.

Definition 5.2. The set of *counterbalanced influences* on i is denoted by $\mathcal{C}(i)$. It is defined as follows:

- if i is observed, the influence of $j \rightsquigarrow i$ is counterbalanced if it is different from the variation of X_i, i.e. if $s(\Delta X_j)s(j \rightsquigarrow i) \neq s(\Delta X_i)$. Then

  $\text{if} s(\Delta X_i) = +\text{then} \mathcal{C}(i) = \mathcal{I}^-(i)$
  $\text{if} s(\Delta X_i) = -\text{then} \mathcal{C}(i) = \mathcal{I}^+(i)$
  $\text{if} s(\Delta X_i) = 0\text{then} \mathcal{C}(i) = \mathcal{I}(i).$

- If i is not observed, the influence of $j \rightsquigarrow i$ is counterbalanced if there exists another path $k \rightsquigarrow i$ whose influence on X_i has a different sign, i.e. if $\mathcal{I}^+(i)$ and $\mathcal{I}^-(i)$ are both nonempty. Then
  - if $i \notin \mathcal{O}, \mathcal{I}^+(i) \neq \emptyset$ and $\mathcal{I}^+(i) \neq \emptyset$ then $\mathcal{C}(i) = \mathcal{I}(i)$,
  - if $i \notin \mathcal{O}, \mathcal{I}^+(i) \neq \emptyset$ or $\mathcal{I}^-(i) = \emptyset$ then $\mathcal{C}(i) = \emptyset$.

A direct consequence of the definition is that an influence which is not counterbalanced absorbs all the influences in the equations it appears in, so that it does not introduce indetermination's:

Property 5.3. *If a path $j \rightsquigarrow i$ is not counterbalanced then the sign of its influence on the node i is the same as the sign of ΔX_i.*

To summarize, influences are counterbalanced when they compete with other influences. When we can observe which is the result of this competition, the counterbalanced influence is the dominated one. Influences are not counterbalanced either if they are non-competing, or if they dominate their competitors.

Examples. In the example of SREBP-a, positive influences $\mathcal{I}^+(\text{SREBP-a}) = \{P_9\}$ compete with negative ones $\mathcal{I}^-(\text{SREBP-a}) = \{P_4, P_5, P_6, P_7\}$ for determining the unknown sign of the variation of SREBP-a. All these influences are considered to be counterbalanced:

$$\mathcal{C}(\text{SREBP-a}) = \mathcal{I}(\text{SREBP-a}) = \{P_9, P_4, P_5, P_6, P_7\}.$$

The observed negative variation of SREBP implies that positive influences are dominated, hence counterbalanced:

$$\mathcal{C}(\text{SREBP}) = \mathcal{I}^+(\text{SREBP}) = \{P_3\}.$$

5.9. Redundant and essential influences

Let us consider the influences of P_1 and P_2 on SREBP. The observed sign of SREBP implies that P_1 is dominated (thus counterbalanced) by P_2. With no loss, we can skip P_1 in our analysis of influences. We can not skip P_2. We say that P_2 is *essential*, while P_1 is *redundant*. Let us now consider the influences of P_9 and P_6 on SREBP-a. These two influences are competing, but we do not know which one wins. We decide to consider that both are redundant.

Generally, we call essential those influences that can not be skipped without loss. The discussion of the examples justifies the following.

Definition 5.4. A path is said to have an *essential influence* on a node i if it belongs to the set $\mathcal{E}(i)$ defined as follows:

$$\mathcal{E}(i) = \{j \rightsquigarrow i = (j_0 = j, j_1, \ldots, j_l = i) \in \mathcal{I}(i),$$
$$\forall 0 < k < l,\ j \rightsquigarrow j_k \notin \mathcal{C}(k)\}.$$

5.10. Essentially balanced nodes

Let us consider LXR-a. It collects two essential influences of different signs (from PUFA and LXR). We say that LXR-a is *essentially balanced*.

Let us also consider SREBP. This node collects two competing influences: P_1 and P_2. Nevertheless, we know that P_1 is dominated, hence redundant. We say that SREBP is *not essentially balanced*.

More generally, we have the following.

Definition 5.5. An observed variable X_i is said to be *essentially balanced* if it collects essential influences that are also counterbalanced, i.e. if $\mathcal{E}(i) \cap \mathcal{C}(i) \neq \emptyset$.

A non-observed variable X_i is said to be *essentially balanced* if it collects essential influences with opposite signs, that is $\mathcal{E}(i) \cap \mathcal{I}^+(i) \neq \emptyset$ and $\mathcal{E}(i) \cap \mathcal{I}^-(i) \neq \emptyset$.

An essential balance on a variable X_i gathers competing influences such that no information allows to say that one of the competing influences has been previously absorbed in a previous competition. This implies that the essential balance is localized precisely at the node i. The variation of X_i gives the inclination of the competition.

From the above discussion it follows:

Criterion to compare model and data 2. Let us consider a model of molecular influences. Let us consider that the interaction graph contains no positive loop made of nodes that are not connected to the exterior. Let X_i be a variable that collects essential influences, i.e. $\mathcal{E}(i)\emptyset$. Then several situations are possible:

1. X_i is essentially balanced. Then there is a non-redundant competition process localized at i. If X_i is observed, we can say which one of the competing influences is dominating.
2. X_i is not essentially balanced and all the collected essential influences have the same sign.
   If X_i is observed then:
   (a) if the sign of ΔX_i coincides with the one of the essential influences then we can say nothing.
   (b) if the sign of ΔX_i is different from the one of the essential influences then there is probably an error in data.

5.11. Complementarity with the graph valuation algorithm

As already stated, the assessment of competitions is not devoted to proving incompatibilities. For instance the case 2(b) of the Criterion 2 suggests, but does not prove the presence of an error. Its utility is not obvious in our simple working example, but should appear from its application to more complex networks. Case 2(b) of Criterion 2 is a hint for the location of the error.

Case 1 of Criterion 2 is obviously useful as it allows to localize non-redundant competitions. These non-redundant competitions are usually important regulation checkpoints.

5.12. Working example and biological discussion

Computation on the graph related to the working examples *M* of regulation of lipid synthesis showed that there are competitions on nine nodes among 11: LXR-a, SREBP, SREBP-a, ACL, ACC, FAS, SCD1, D5D, D6D. However, some of these competitions are redundant. Among the 32 counterbalanced paths, 20 are essential. Essentially balanced nodes are LXR-a, ACL and SCD1, D5D and D6D:

- LXR-a collects non-redundant competitive paths.
- ACL collects four essential negative influences and no essential positive influence. This suggests that the data on ACL is false.

  The system has to be studied more precisely, in order to check whether there is a real incompatibility between the model and the data. A way to do this is by using the methods of the previous section. An alternative way is a refined assessment of competitions (see Section 5.13).
- SCD1, D5D and D6D are observed nodes that collect essential negative and positive influences. This suggests that a non-redundant competition occurs on each of these nodes.

  The observed negative variation on SCD1 suggests that the negative path through SREBP dominates the positive path through PPAR.

Consequently, two points can be extracted from such an analysis: at the SCD1 level, the negative pathway containing SREBP dominates the positive pathway containing PPAR. At the ACL level, the data on ACL should be checked carefully or paths in the model leading to ACL should be studied in detail.

5.13. Improvements

Our discussion of competitions aims at identifing and localizing essential competitions. Let us describe modifications of the definitions that allow us to refine the analysis.

Definitions 5.2 and 5.4 declare redundant influences that compete with others, even if the result of the competition is unknown. We have did so in order to localize essential competitions: once they occur somewhere, competitions are transported along the graph as non-essential.

With altered definitions we can identify and localize incompatibilities. Let us change the Definition 5.2 of counterbalanced influences by skipping its second part. This means deciding that influences of different signs on an unobserved node are not counterbalanced. Then, according to Definition 5.4, the redundant influences are only those that are dominated somewhere. Hence the case 2(b) of the Criterion 2 identifies an incompatibility: it is impossible to act with non-dominated influences of even sign on a node and to obtain a different sign on the node. This is actually the case of the node ACL in our working example. This node collects non-dominated negative influences and one dominated positive influence.

Thus, by altering Definitions 5.2 and 5.4 we can detect incompatilities, but we can no longer localize competitions. In order to satisfy both purposes we can introduce two types of redundancy, one suitable for incompatibility detection, the other for localizing competitions.

6. Application to an extended model of the synthesis of fatty acids

We applied the graph valuation algorithm and the essential balance computation to an extended model of the synthesis of fatty acids shown in Fig. 2. The model contains 59 products and 110 interactions implied in the synthesis metabolism. The product list includes ACL, ACC, D5D, D6D, FAS, LXR, PPAR, PUFA, SCAP, SCD1 that were introduced in the working example given in Fig. 1. Compared to the working example, this new model includes neither the active forms of SREBP and LXR, nor SCAP. The model was obtained as a compilation of interactions extracted from the litterature. Notice that in this model, only the nodes without predecessors (that is, INS and retinoid) are connected to the exterior. Hence, we suppose implicitely that the model is complete enough to explain all the possible actions among products.

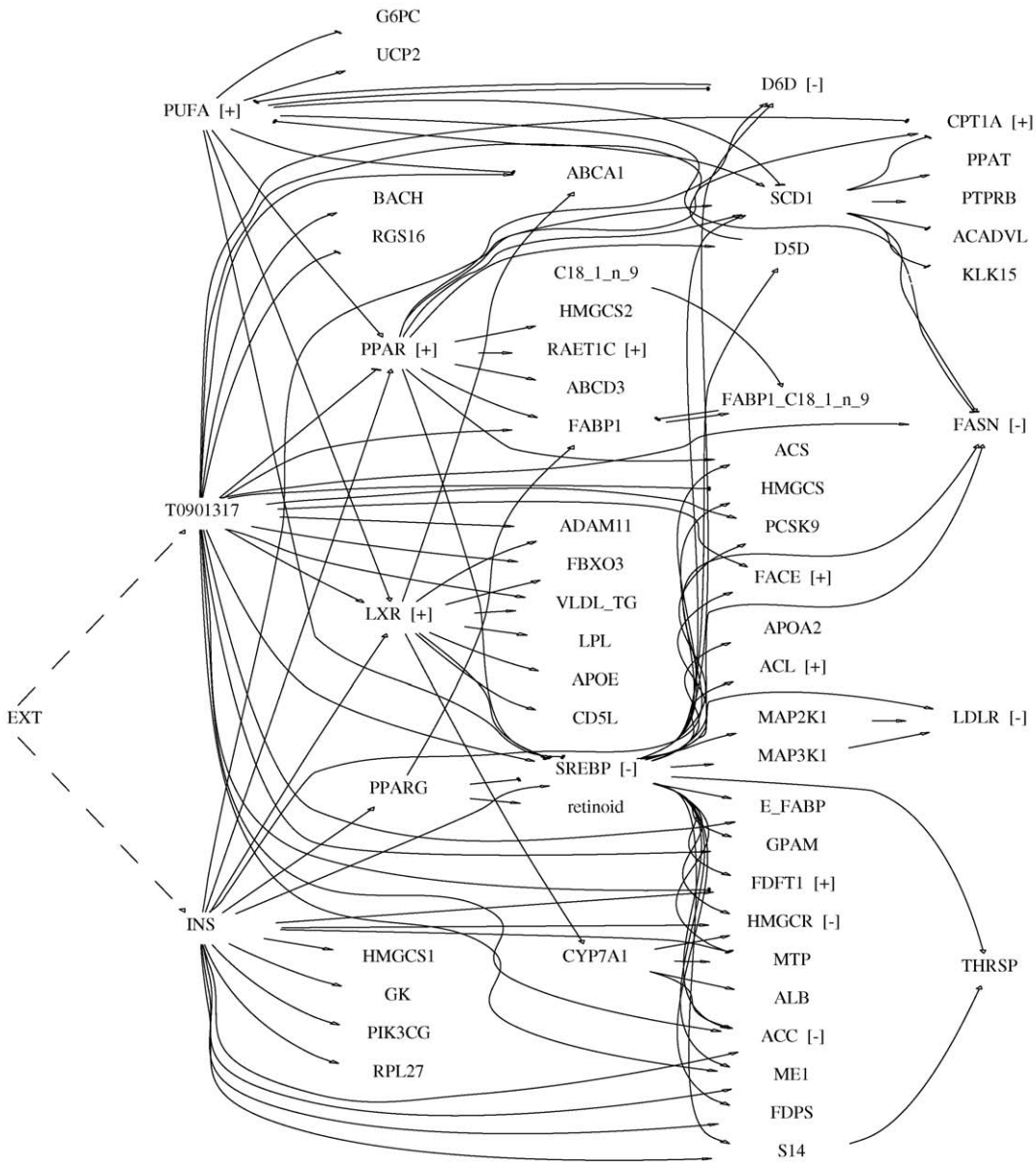
We have used a theoretical experimental set of measurements for 16 products, built as a result of a virtual experimentation on fasted animals. Hence, we suppose that PUFA, PPAR, LXR, ACL increase while ACC, D6D, FAS and SREBP decrease.

6.1. Backward–forward algorithm

The backward–forward algorithm states that the system of qualitative equations is incompatible with the set of data. The qualitative equation that cannot be satisfied by the data is the equation corresponding to ACL.

6.2. Essential balances

The result of the computation of essential balances is shown in Table 3. Six products are essentially bal-



[+]	PUFA, CPTIA, LXR, RAETIC, PPAR ACL, FACE, FDFT1
[−]	D6D, SREBF1, ACC, FAS, HMGCR, LDLR

Fig. 2. Interaction graph for an extended model of regulation of the synthesis of fatty acids. The model contains 59 nodes and 110 interactions. Sixteen products are supposed to be observed. The list of products is given in Table 3.

anced, among which D5D, D6D, SCD1 and SREBP. In the working example, an essential balance occured on the active form of LXR, denoted LXR-a; it cannot occur in this model since LXR-a is not considered any more. Hence, the essential balance on LXR-a on the working example is shifted in the new model as a balance on the successor of LXR-a, that is, SREBP.

16 products collect only positive pure influences. Four of them were supposed to be positive during the experimentation, which is confirmed by the analysis. The other

Table 3
Computation of essential influences on the extended model of the synthesis of fatty acids shown in Fig. 2

	Essentially balanced	Positive essentially influenced	Negatively essentially influenced	Not essentially influenced
Data=+		PUFA, CPT1A, LXR, RAET1C, PPAR	ACL, FACE, FDFT1	
Data=-	D6D, SREBF1		ACC, FAS, HMGCR, LDLR	
No data	ACS, ABCA1, D5D, SCD1	ABCD3, APOE, CD5L, ADAM11, CYP7A1, FBXO3, FABP1, HMGCS2, LPL, MTP, UCP2, VLDL_TG	APOA2, E_FABP,FDPS, G6PC, GPAM, HMGCS, MAP2K1, MAP3K1, ME1, PCSK9, S14, THRSP	ACADVL, ALB, BACH, C18_1_n_9, FABP1_C18_1_n_9, GK, HMGCS1, INS, KLK15, PIK3CG, PPARG, PPAT, PTPRB,RGS16, RPL27, T0901317, retinoid

The computation suggests that competitions occur in D6D, SREBP, ACS, ABCA1, D5D and SCD1. Moreover, the competition should incline in the negative direction in D6D and SREBP. The computation also suggests that the data about ACL, FACE and FDFT1 have to be corrected. Finally, the computation suggests that ACS, ABCA1, D5D and SCD1 either incline in the positive direction, or collect a competition process. Similarly, ABCD3, APOE, CD5L, ADAM11, CYP7A1, FBXO3, FABP1, HMGCS2, LPL, MTP, UCP2 and VLDL_TG either incline in the negative direction, or collect a competition process.

12 products collect only positive influences, meaning that either they are balanced, or they incline in the positive direction.

Nineteen products collect only negative pure influences. Among them, ACL, FACE and FDFT1 were supposed to incline in the positive direction during the experimentation. We suspect that there is an error in the data.

Nothing can be said about the 17 products that do not collect eseential influence.

6.3. Backward–forward algorithm on a corrected set of data

The backward–forward algorithm on the set of data stated that the set of data was not compatible with the model. The computation of essential balances suggested that the data on ACL, FACE and FDFT1 were false. Hence, we propose as a new set of data the following set:

[+]	PUFA, CPT1A, LXR, RAET1C, PPAR
[−]	D6D, SREBF1, ACC, FAS, HMGCR, LDLR, ACL, FACE, FDFT1

Then the backward–forward algorithm terminates so that this set of data is compatible.

Finally, the backward–forward algorithm states that the inclinaison of nonobserved values suggested by the essential balanced algorithm is compatible with the interaction graph.

[+]	PUFA, CPT1A, LXR, RAET1C, PPAR, ACS, ABCA1, D5D, SCD1, ABCD3, APOE, CD5L, ADAM11, CYP7A1, FBXO3, FABP1, HMGCS2, LPL, MTP, UCP2, VLDL_TG
[−]	D6D, SREBF1, ACC, FAS, HMGCR, LDLR, ACL, FACE, FDFT1

7. Remarks and conclusion

In this paper we have discussed how a qualitative theoretical model of a mixed gene and metabolic network can be confronted to DNA-microarray and metabolic data.

Using a differential description of dynamics we have obtained quantitative equations connecting the theoretical variations of genes and metabolic products to the changes of the external conditions. In their linear version, that connects small variations of genes and products, these equations are analogous to the discrete linear Laplace equations on graphs that have been studied in the context of Markov chains (Chung and Yau, 2000; Soardi, 1994; Campanino and Petritis, 2003). In order to calculate the variations we have to compute the matrix inverse of the Jacobian of the differential dynamical system. This matrix inverse is in fact the discrete Green function of the problem. Contrary to (Chung and Yau, 2000) we have not searched for explicit formulas for the Green function (difficult to obtain for irregular graphs), but we have emphasized how one node collects influences from the other nodes. A similar path representation of the Green function is used in statistical mechanics (Campanino and Petritis, 2003).

We have extended the quantitative equations to the case of large variations of the products and we have established general linear qualitative equations that are valid both for small and large variations. The validity of the approach relies on uniform signs of the interaction coefficients (elements of the Jacobian) within the range of the experimental conditions.

We have introduced the graph valuation algorithm that can reduce or solve the qualitative system. The computation rules are not those of usual linear spaces, but belong to the sign algebra.

When the qualitative system is incompatible, it is generally difficult to propose corrections. When it is compatible, we may obtain many solutions. For this reason we have also introduced a complementary approach which is the study of competitions. This approach treats all nodes on equal footing. It does not prove incompatibilities, but it identifies biologically important competition processes and the places where they occur.

The analysis of the examples taken from the regulation of lipogenesis suggests that this type of methods could be used to optimize experimental protocols. The optimization criteria are not formalized here. Nevertheless, as a general rule, if a node is observed we would like to have observations as complete as possible on its predecessors. In the working example, SCAP is not observed. Instead, we have predicted the sign of its variation. The experimental knowledge of SCAP variation would further facilitate the localization of the discovered incompatibility between interaction and data.

Acknowledgements

This research was supported by the ACI IMPBio, French Ministry for Research program on interdisciplinarity. We would like to thank M. Crouzeix, I. Gruais, Y. Guivarch, D. Petritis, E. Pécou, R. Kant, and C. Soulé for many inspiring discussions.

Appendix A. Proofs

Proof of Property 3.2. From $\frac{\partial F_i}{\partial X_i} < -C$ it follows that for any fixed $\hat{X}^{(i)}$, $F_i(X_i, \hat{X}^{(i)})$ depends monotonically on X_i. Furthermore, F_i becomes negative for large enough X_i. As it is positive for $X_i = 0$, by the intermediate value theorem it vanishes somewhere. Eq. (3) has an unique solution $X_i = \Phi_i(\hat{X}^{(i)})$ for any $\hat{X}^{(i)}$. The differential of Φ_i follows from the implicit function theorem. □

Proof of Theorem 3.4. From Eq. (7) it follows:

$$\delta X_i = -\sum_{j\in\mathring{G}} \mathring{B}_{ij} \sum_{k\in\neg^{in}G} \frac{\partial F_j}{\partial X_k}\delta X_k, \tag{22}$$

where $\mathring{B} = \mathring{A}^{-1}$.

Let us consider the following well-known property of the inverse matrix (Bloom, 1979):

$$\mathring{A}^{-1}_{ij} = (-1)^{i+j}\frac{\det(M_{ji})}{\det\mathring{A}} \tag{23}$$

where M_{ij} is the minor obtained by deleting line i and column j in matrix $\mathring{A}$.

Furthermore, it is a simple exercise (Bloom, 1979) to show that:

$$M_{ji} = \begin{cases} (-1)^{i+j}\sum_{j\rightsquigarrow i}\left[(-1)^{l_{j\rightsquigarrow i}}a_{j\rightsquigarrow i}\det\mathring{A}_{(j\rightsquigarrow i)^c}\right] & \text{if } i \neq j \\ \det\mathring{A}_{\{j\}^c} & \text{if } i = j. \end{cases} \tag{24}$$

where $j\rightsquigarrow i$ is any path leading from j to i with no loop, $l_{j\rightsquigarrow i}$ the length of the path, $a_{j\rightsquigarrow i}$ the product of elements of $\mathring{A}$ along this path, $\mathring{A}_{(j\rightsquigarrow i)^c}$ is the principal minor defined by the set of indices complementary to those in the path. Conventionally, $\det\mathring{A}_{\overset{\circ}{G}^c} = 1$ (determinant of the empty matrix).

Using Eqs. (23) and (24) it follows:

$$\delta X_i = -\sum_{k\in\neg G}\frac{\det\mathring{A}_{\{i\}^c}}{\det\mathring{A}}\frac{\partial F_i}{\partial X_k}\delta X_k - \sum_{j\in\mathring{G}, j\neq i}\sum_{j\rightsquigarrow i} \times \left[\sum_{k\in\neg G}(-1)^{l_{j\rightsquigarrow i}}a_{j\rightsquigarrow i}\frac{\det\mathring{A}_{\{j\rightsquigarrow i\}^c}}{\det\mathring{A}}\frac{\partial F_j}{\partial X_k}\delta X_k\right].$$

We can identify in the sum the contributions of direct (one arc) and indirect (at least two arcs) paths from the boundary to the interior. Thus:

$$\delta X_i \sum_{k\in\neg^{in}G}\left[\sum_{k\rightsquigarrow i}^{\text{indirect}}\frac{a_{k\rightsquigarrow i}}{C_{k(j)\rightsquigarrow i}} + \sum_{k\rightsquigarrow i}^{\text{direct}}\frac{a(k,i)}{C_i}\right]\delta X_k$$

from which it follows Eq. (8). □

Proof of Theorem 3.6. The existence of a solution of the nonlinear Dirichlet problem Eq. (9) is a rather standard mathematical result based on Eqs. (10) and (11) (Radulescu et al., 2005). The uniqueness of such a solution follows from the absence of positive loops in the interior of G (Thomas's conjecture proven in (Soulé, 2003)). Thus, the unique solution of Eq. (9) can be represented as a function $\mathring{X}_G = \Phi_G(\hat{X}_G)$. The differential of Φ_G is the solution of the linear Dirichlet problem, given by Eq. (8). Integrating this differential on the segment connecting $\hat{X}1$, $\hat{X}2$ and passing to sign algebra we find Eq. (12). We also need the fact that all internal moduli are positive which by Property 3.5 follows from the absence of positive loops. □

Proof of Property 3.5. If there are no positive loops, then it can be shown (see Soulé (Soulé, 2003)) that the signs of the principal minors of the Jacobian $\det\mathring{A}$ are alternating. Thus:

$$\text{sign}(\det\mathring{A}_{\{j\}^c}) = (-1)^{\#\mathring{G}-1}$$

$$\mathrm{sign}(\det \mathring{A}_{(j\rightsquigarrow i)^c}) = (-1)^{\#\mathring{G}-l_{j\rightsquigarrow i}-1}$$

$$\mathrm{sign}(\mathring{A}_{jj}) = \mathrm{sign}(a_{jj}) = -1$$

from which it follows that all the moduli are positive. □

Proof of Property 5.1. If $P = j\rightsquigarrow i = (j, j_1, \ldots, i)$ contains an intermediate node j_k that is connected to the exterior, then for every subgraph G, j_k belongs to the boundary of G, hence $P \notin \mathcal{P}_{\mathring{G}}$.

Conversely, if P contains no intermediate node connected to the exterior $\mathcal{E}$ let $G = \{j_0 = j, j_1, \ldots, j_n = i\} \cup_{0<k\leq n} \mathrm{pred}(j_k)\}$. Then $P \in \mathcal{P}_{\mathring{G}}$, $j \in \urcorner^{in} G$, hence $P \in \mathcal{P}(M)$.

In order to prove the second part of the Property, Theorem 3.6 can be applied to G. □

References

Angeli, J., Ferrell, J.J., Sontag, E., 2004. Detection of multi-stability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. PNAS 1822–1827.

Bakker, B.M., Michels, P.A.M., Opperdoes, F.R., Westerhoff, H.V., 1997. Glycolysis in bloodstream from trypanasoma brucei can be understood in terms of the kinetics of the glycotic enzymes. J. Biol. Chem. 272, 3207–3215.

Batt, G., Ropers, D., de Jong, H., Geiselmann, J., Mateescu, R., Page, M., Schneider, D., 2005. Validation of qualitative models of genetic regulatory networks by model checking: analysis of the nutritional stress response in escherichia coli. Bioinformatics 21 (Suppl. 1), i19–i28, special issue ISMB-2005.

Bay, S., Shrager, J., Pohorille, A., Langley, P., 2003. Revising regulatory networks: from expression data to linear causal models. J. Biomed. Inform. 35, 289–297.

Bloom, D., 1979. Linear Algebra and Geometry. Cambridge University Press, Cambridge.

Boyer, F., Viari, A., 2003. Ab initio reconstruction of metabolic pathways. Bioinformatics 19 (Suppl. 2).

Bryant, C., Muggleton, S., Olivier, S., Kell, D., Reiser, P., King, R., 2001. Combining inductive logic programming, active learning and robotics to discover the function of genes. Electron. Trans. Artif. Intel. 5, 1–36.

Calvano, S., Xiao, W., Richards, D., Felciano, R., Baker, H., Cho, R., Chen, R., Brownstein, B., Cobb, J., Tschoeke, S., Miller-Graziano, C., Moldawer, L., Mindrinos, M., Davis, R., Tompkins, R., Lowry, S., IA, L. S. C. R. P., published online Augest 31, 2005. A network-based analysis of systemic inflammation in humans. Nature.

Calzone, L., Chabrier-Rivier, N., Fages, F., Soliman, S., 2005. A machine learning approach to biochemical reaction rules discovery. In: Francis, J., Doyle III, (Eds.), Proceedings of Foundations of Systems Biology and Engineering FOSBE'05, Santa Barbara, pp. 375–379

Campanino, M., Petritis, D., 2003. Random walks in randomly oriented lattices. Markov Processes and Related Fields 9, 391–412.

Chabrier-Rivier, N., Chiaverini, M., Danos, V., Fages, F., Schächter, V., 2004. Modeling and querying biomolecular interaction networks. Theor. Comp. Sci. 325 (1), 25–44.

Chaouiya, C., Remy, E., Ruet, P., Thieffry, D., 2004. Qualitative modelling of genetic networks: from logical regulatory graphs to standard petri nets. Lect. Note Comput. Sci. 3099, 137–156.

Chaves, M., Albert, R., Sontag, E., 2005. Robustness and fragility of boolean models for genetic regulatory networks. J. Theor. Biol. 235, 431–449.

Chevalier, T., Schreiber, T., Ross, J., 1993. Toward a systematic determination of complex reaction mechanisms. J. Phys. Chem. 97.

Chung, F., Yau, S., 2000. Discrete green's functions. J. Combin. Theor. Ser. A 91, 191–214.

de Jong, H., 2002. Modeling and simulation of genetic regulatory systems: a literature review. J. Comp. Biol. 9 (1), 69–105.

de Jong, H., Gouzé, J.-L., Hernandez, C., Page, M., Sari, T., Geiselmann, J., 2004. Qualitative simulation of genetic regulatory networks using piecewise-linear models. Bull. Math. Biol. 66, 301–340.

Dormoy, J., 1988. Controlling qualitative resolution. In: Proceedings of the seventh National Conference on Artificial Intelligence, AAAI88', Saint-Paul, Minn.

Eisenthal, R., Cornish-Bowden, A., 1998. Propsects for antiparasitic drugs: the case of trypanasoma brucei, the causative agent of african sleeping sickness. J. Biol. Chem 272, 5500–5505.

Fell, D., 1997. Understanding the Control of Metabolism. Portland Press, London.

Ghoshn, R., Andomlin, C., 2004. Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling: delta-notch protein signalling. Syst. Biol. 1 (1), 170–183.

Heinrich, R., Schuster, S., 1996. The Regulation of Cellular Systems. Chapman and Hall, New York.

Jump, D., 2004. Fatty acid regulation of gene transcription. Crit. Rev. Clin. Lab. Sci. 41 (1), 41–78.

Kaern, M., A. Elston, T., Blake, W. J., Collins, J.J., 2005. Stochasticity in gene expression: from theories to phenotypes. Nat. Rev. Genet. 6, 451–464.

Kaminski, N., Friedman, N., 2002. Practical approaches to analyzing results of microarray experiments. Am. J. Resp. Cell Mol. Biol. 27, 125–213.

Karp, P., Riley, M., Paley, S., Pellegri, i-Toole, A., Krummmenacker, M., 1996. Eco-cyc: encyclopedia of escerichia coli genes and metabolism. Nucl. Acid Res. 24, 32–39.

Kauffman, S., 1993. The Origin of Order, Self-Organisation and Selection in Evolution. Oxford University Press, Oxford, U.K.

King, R., Garrett, S., Coghill, G., 2005. On the use of qualitative reasoning to simulate and identify metabolic pathways. Bioinformatics 21 (9), 2017–2026.

Kuipers, B., 1994. Qualitative reasoning. Mdeling and Simulation With Incomplete Knowledge. MIT Press.

Langley, P., Shiran, O., Shrager, J., Todorovski, L., Pohorille, A., 2005. Constructing explanatory process models from biological data and knowledge. Artif. Intell. Med., in press.

Lee, S., Chan, W., Lo, C., Wan, D., Tsang, D., Cheung, W., 2004. Requirement of pparalpha in maintaining phospholipid and triacylglycerol homeostasis during energy deprivation. J. Lipid. Res. 45 (11), 2025–2037.

Liang, G., Yang, J., Horton, J., Hammer, R., Goldstein, J., Brown, M., 2002. Diminished hepatic response to fasting/refeeding and liver x receptor agonists in mice with selective deficiency of sterol regulatory element-binding protein-1c. J. Biol. Chem. 277 (15), 9520–9528.

Matsuno, H., Doi, A., Nagasaki, M., Miyano, S., 2000. Hybrid petri net representation of gene regulatory network. Pac. Symp. Biocomput. 5, 341–352.

Matsuzaka, T., Shimano, H., Yahagi, N., Amemiya-Kudo, M., Yoshikawa, T., Hasty, A., Tamura, Y., Osuga, J., Okazaki, H., Iizuka, Y., Takahashi, A., Sone, H., Gotoda, T., Ishibashi, S., Yamada, N., 2002. Dual regulation of mouse Delta-5- and Delta-6-desaturase gene expression by srebp-1 and pparalpha. J Lipid Res 43 (1), 107–114.

Mendes, P., 1997. Biochemistry by numbers: simulation of biochemical pathways with gepasi 3. Trends Biochem. Sci. 22, 36–363.

Miller, C., Ntambi, J., 1996. Peroxisome proliferators induce mouse liver stearoyl-coa desaturase 1 gene expression. Proc. Natl. Acad. Sci. U.S.A. 93 (18), 9443–9448.

Nachman, I., Regev, A., Friedman, N., 2004. Inferring quantitative models of regulatory networks from expression data. Bioinformatics 20, i248–i256.

Nara, T., He, W., Tang, C., Clarke, S., 2002. The e-box like sterol regulatory element mediates the suppression of human Delta-6 desaturase gene by highly unsaturated fatty acids. Biochem. Biophys. Res. Commun. 296 (1), 111–117.

Papin, J., Stelling, J., Price, N., Klamt, S., Schuster, S., Bo, P., 2004. Comparison of network-based pathway analysis methods. Trend Biotechnol. 22, 400–405.

Radulescu, O., Lagarrigue, S., Siegel, A., Veber, P., Le Borgne, M., 2005. Topology and static response of interaction networks in molecular biology. J. R. Soc. Interface, in press.

Rao, C., Wolf, D., Arkin, A., 2002. Control exploitation and tolerance of intracellular noise. Nature 420, 231–237.

Reiser, P., King, R., Kell, D., Muggleton, S., Bryant, C., Oliver, S., 2001. Developing a logical model of yeast metabolism. Electron. Trans. Artif. Intel. 5, 223–244.

Sanchez, L., Thieffry, D., 2001. A logical analysis of the drosophila gap-gene system. J. Theor. Biol. 211, 115–141.

Soardi, P., 1994. Potential Theory on Infinite Networks. Springer, Berlin, Heidelberg, New York.

Soulé, C., 2003. Graphic requirements for multistationarity. Complexus 1, 123–133.

Steffensen, K., Gustafsson, J., 2004. Putative metabolic effects of the liver x receptor (lxr). Diabetes 53 (Supp 1), 36–52.

Tang, C., Cho, H., Nakamura, M., Clarke, S., 2003. Regulation of human Delta-6 desaturase gene transcription: identification of a functional direct repeat-1 element. J Lipid Res 44 (4), 686–695.

Thomas, R., 1973. Boolean formalization of genetic control circuits. J. Theor. Biol. 42, 563–585.

Tobin, K., Steineger, H., Alberti, S., Spydevold, O., Auwerx, J., Gustafsson, J., Nebb, H., 2000. Cross-talk between fatty acid and cholesterol metabolism mediated by liver x receptor-α. Mol Endocrinol 14 (5), 741–752.

Tomita, M., Hashimoto, K., Takahashi, K., Shimuzu, T., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J., Hutchinson, J., 1999. E-cell:software environment of whole-cell simulation. Bioinformatics 15, 72–84.

Travé-Massuyès, L., Dague, P. (Eds.), 2003. Modèles et Raisonnements Qualitatifs. Hermes sciences.

Tyson, J.J., Chen, C., Novák, B., 2003. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. Curr. Opin. Cell Biol. 15, 221–231.

Veber, P., Le Borgne, M., Siegel, A., Lagarrigue, S., Radulescu, O., 2005. Complex qualitative models in biology: a new approach. In: European Conference on Complex Systems 2005. Paris.

Yamanishi, Y., Vert, J.-P., Kanehisa, M., 2004. Protein network inference from multiple genomic data: a supervised approach. Bioinformatics 20, i363–i370.

Zupan, B., Bratko, I., Demsar, J., Beck, J., Kuspa, A., Shaunlsky, G., 2001. Abductive inference of genetic networks. In: Proceedings of the Eighth Conference on AI in Medicine in Europe: Artificial Intelligence Medicine. Lecture Notes In Computer Science, vol. 2101. pp. 304 – 313.