Learning Boolean controls in regulated metabolic networks: a case-study

Kerian Thuillier, Caroline Baroukh, Alexander Bockmayr, Ludovic Cottret, Loïc Paulevé, Anne Siegel

Learning Boolean controls in regulated metabolic networks: a case-study

Kerian Thuillier[1], Caroline Baroukh[2], Alexander Bockmayr[3], Ludovic Cottret[2], Loïc Paulevé[4], and Anne Siegel[1]

[1] Univ Rennes, Inria, CNRS, IRISA, F-35000 Rennes, France
[2] LIPME, INRAE, CNRS, Université de Toulouse, Castanet-Tolosan, France
[3] Freie Universität Berlin, Institute of Mathematics, D-14195 Berlin, Germany
[4] Univ. Bordeaux, Bordeaux INP, CNRS, LaBRI, UMR5800, F-33400 Talence, France

Abstract. Many techniques have been developed to infer Boolean regulations from a prior knowledge network and experimental data. Existing methods are able to reverse-engineer Boolean regulations for transcriptional and signaling networks, but they fail to infer regulations that control metabolic networks. This paper provides a formalisation of the inference of regulations for metabolic networks as a satisfiability problem with two levels of quantifiers, and introduces a method based on Answer Set Programming to solve this problem on a small-scale example.

Keywords: Inference · Regulated metabolism · Satisfiability problem.

1 Introduction

During the last twenty years, both the amount and the type of available data have allowed scientists to consider intracellular processes as a whole. Boolean networks have been refined to include non-deterministic dynamics in order to model the response of regulatory interactions [16,2,5]. Similarly, the study of metabolism at steady state has led to various constraint-based approaches [19,17], which usually assume that internal metabolites are in a quasi-steady-state (QSS). The classical approach to analyze metabolic networks at steady state is flux balance analysis (FBA) [19]. In this approach, a linear function, e.g. biomass production, is optimized with respect to stoichiometric and thermodynamic constraints, resulting in a linear programming problem (LP).

However, both the Boolean approach for regulation and the QSS approximation for metabolism are often developed *"in solo"*, without considering that cellular biology is multi-layered in the sense that the metabolic layer interacts through feed-forward and feedback loops with the regulatory layer [4,27,21,9]. Indeed, cellular metabolism transforms nutrients into biomass constituents. Metabolic reactions are catalysed by enzymes, which themselves are controlled by a cascade of regulations involving other proteins, metabolites and abiotic factors, such as temperature and pH. A biological system thus has several layers of control, which mutually depend on each other. It cannot be simply viewed as a

purely hierarchical system because there are regulatory feed-forward and feedback mechanisms to inform each layer on the state of the other ones. In concrete terms, some compounds produced by the metabolic layer have the capability to block or induce signaling regulation cascades, which themselves can block or induce transcription of genes leading to changes in the control of the initial metabolic process.

To figure out how gene expression triggers specific phenotypes depending on the environmental constraints [3], several constraint-based approaches for integrating metabolic and regulatory networks have been developed that combine Boolean dynamics for the regulatory layer with quasi-steady-state approximations of the metabolic layer (see [17] for an overview), one of them being FlexFlux [18], which implements the rFBA framework [9]. A major limitation when using such frameworks to analyse regulated metabolic models is that they require a precise description of the regulatory and signaling layers in the form of Boolean rules. A noticeable exception is [24], where RBA is used to deduce regulations according to perturbations of the environment. However, to induce regulations, the authors assume that no feedback from metabolism to regulation occurs, which does not correspond to the functioning of most systems. In practice, these rules are manually curated from the literature or experimental data. This has been done for example in the case of *E. coli* [8,7] and a few other organisms. But, the need for a manual curation of Boolean rules of regulated metabolism is a strong limitation to the use of these frameworks.

Signaling and regulatory rules can be identified from transcriptomic or phosphoproteomics data by solving combinatorial or MILP problems in order to optimize data-fitting and parsimony hypotheses [23,20,26,22,25]. In this direction, the caspoTS and the BoNesis approaches [22,20,26,6] were developed for inferring Boolean rules to model the response of regulatory and signaling networks from multiple time-series data. The goal of this paper is to lay foundation for the extension of these approaches to the inference of regulatory rules driving metabolism. This is done by discretizing both the rFBA framework (especially the QSS approximation) and the metabolic data used as input of the inference procedure.

This paper is structured as follows. Sect. 2 gives the background on the dynamic rFBA framework for the simulation of coupled metabolic and regulatory networks. In Sect. 3, we define a formal Boolean abstraction of dynamic rFBA simulations. Then, in Sect. 4, we build on this Boolean abstraction to express the inference of the logic of metabolic regulations as a satisfiability problem. Finally, in Sect. 5, we apply the obtained inference framework on a case study of simplified core carbon metabolism.

Notations The cardinality of a finite set X is denoted by $|X|$. Given a vector $x \in D^n$ and a set of indices $I \subseteq \{1, \cdots, n\}$, x_I denotes the vector of dimension $|I|$ equal to $(x_i)_{i \in I}$. The Boolean domain is denoted by $\mathbb{B} = \{0, 1\}$. Given two Boolean vectors $x, y \in \mathbb{B}^n$, we write $x \preceq y$ iff $\forall i \in \{1, \cdots, n\}, x_i \leq y_i$. Finally, given a non-negative real vector $s \in \mathbb{R}^n_{\geq 0}$, we denote by $\beta(s) \in \mathbb{B}^n$ its *binarization*, *i.e.* $\forall i \in \{1, \ldots, n\}, \beta(s)_i = 1$, if $s_i > 0$, and $\beta(s)_i = 0$, if $s_i = 0$.

2 Background: regulated metabolic networks

2.1 Coupling metabolic and regulatory networks

A *regulated metabolic network* consists of two layers. The regulatory layer is modelled by a Boolean network, which controls the metabolites and fluxes of the metabolic layer, which is characterized by linear equations. Feedbacks are provided by the components of the metabolic network, which are involved in the Boolean functions associated with the regulatory layer.

Formally, a metabolic network is given by a set of biochemical reactions linked together by the metabolites that they consume and produce.

Definition 1. *A* metabolic network *is a tuple* $\mathcal{N} = (\mathrm{Int}, \mathrm{Ext}, \mathcal{R}, S)$ *with a set of internal metabolites* Int*, a set of external metabolites* Ext*, a set* $\mathcal{R}$ *of irreversible reactions, and a stoichiometric matrix* $S \in \mathbb{R}^{(|\mathrm{Int}|+|\mathrm{Ext}|)\times|\mathcal{R}|}$.

Given flux bounds $l_r, u_r \in \mathbb{R}, 0 \leq l_r \leq u_r$*, for each* $r \in \mathcal{R}$*, a* metabolic steady state *is a flux vector* $v \in \mathbb{R}^{|\mathcal{R}|}$ *with* $S_{\mathrm{Int},\mathcal{R}} \cdot v = 0$ *and* $l_r \leq v_r \leq u_r$*, for all* $r \in \mathcal{R}$*. Here* $S_{\mathrm{Int},\mathcal{R}}$ *denotes the submatrix of* S *whose rows correspond to the internal metabolites.*

For the sake of simplicity, we assume that all reactions are irreversible. Reversible reactions may be split into a forward and backward reaction if necessary.

Definition 2 (Input and output metabolites). *For an external metabolite* $m \in \mathrm{Ext}$*, we denote by* $w_m = w_m(t) \in \mathbb{R}_{\geq 0}$ *the concentration of* m *at time* $t \geq 0$*.*

An external metabolite $m \in \mathrm{Ext}$ *is called an* input (resp. output) metabolite *if there exists a reaction* $r \in \mathcal{R}$ *with* $S_{mr} < 0$ *(resp.* $S_{mr} > 0$*). Here* S_{mr} *denotes the stoichiometric coefficient of metabolite* m *in reaction* r*. The set of all input metabolites is denoted by* $\mathrm{Inp} \subseteq \mathrm{Ext}$*.*

A regulatory network is a set of biological entities (e.g. genes, reactions, metabolites) or even abiotic entities (e.g. temperature, pH) that are linked by causal effects: the activity of some nodes can affect positively or negatively the activity of other nodes. This activity can be represented by a Boolean network.

Definition 3. *A* Boolean network *(BN) of dimension* n *is a function* $f : \mathbb{B}^n \to \mathbb{B}^n$*. For each* $i \in \{1, \ldots, n\}$*, the* i*-th component* $f_i : \mathbb{B}^n \to \mathbb{B}$ *is called the* local function *of* i*.*

The influence graph $G(f)$ *of* f *is a signed digraph* (V, E) *with* $V = \{1, \ldots, n\}$ *and* $E \subseteq V \times \{-, +\} \times V$ *such that* $(i, s, j) \in E$ *if and only if there exists* $x \in \mathbb{B}^n$ *with* $x_i = 0$ *such that* $s \cdot f_j(x) < s \cdot f_j(x_1, \cdots, x_{i-1}, 1, x_{i+1}, \cdots, x_n)$*. In the following we will slightly abuse notation by identifying* $G(f)$ *with its edge set,* i.e. $G(f) = E$*.*

A BN f *is* locally monotone *whenever for each influence* $(i, s, j) \in G(f)$*, there is no influence with opposite sign,* i.e. $(i, -s, j) \notin G(f)$*.*

We assume here that the fluxes of a metabolic network can be controlled by the activity of the input metabolites and additional regulatory proteins. More

precisely, the activity of some reactions can be blocked (forced to have a zero flux) whenever certain conditions on the activity of input metabolites and regulatory proteins are met. Moreover, we assume that the activity of regulatory proteins is mediated by the metabolic network only. The resulting model is then supposed to run on two time scales: the metabolic network is a fast system, which, depending on the activity of input metabolites and regulatory proteins will converge to a steady state of the reactions fluxes; the regulatory network is a slow system, which gets updated once the metabolic network is in steady state.

Definition 4 (Regulated metabolic network). *A* regulated metabolic network *is a triplet* $(\mathcal{N}, \mathcal{P}, f)$ *composed of:*

- *a metabolic network* $\mathcal{N} = (\mathrm{Int}, \mathrm{Ext}, \mathcal{R}, S)$ *with* k *input metabolites* $\mathrm{Inp} = \{e_1, \cdots, e_k\} \subseteq \mathrm{Ext}$ *and* m *reactions* $\mathcal{R} = \{r_1, \cdots, r_m\}$*;*
- *a set of* d *regulatory proteins* $\mathcal{P} = \{p_1, \ldots, p_d\}$
- *a BN* f *of dimension* $n = |\mathrm{Inp}| + |\mathcal{R}| + |\mathcal{P}|$ *where* $\{1, \ldots, n\} = \mathrm{Inp} \cup \mathcal{R} \cup \mathcal{P}$ *such that* $G(f)$ *is a bipartite graph between* $\mathcal{P}$ *and* $\mathrm{Inp} \cup \mathcal{R}$*.*

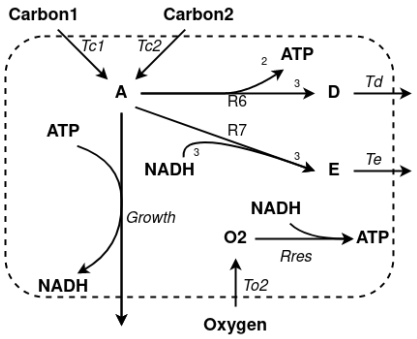
In this work, local functions for input metabolites in the BN f are never used (although the local functions of reactions may depend on them). Therefore we set arbitrarily $f_e = 0, \forall e \in \mathrm{Inp}$.

The BN f models the regulation of the fluxes in the metabolic network $\mathcal{N}$. This regulation is always in one direction: either a flux v_r is only restricted by the flux bounds $l_r \leq v_r \leq u_r$, whenever $f_r(x) = 1$, or it is blocked, $v_r = 0$, whenever $f_r(x) = 0$. Following this convention, a reaction $r \in \mathcal{R}$ is never regulated whenever $f_r(x) = 1$. As we will define formally in the next section, the regulations impact the steady states of the metabolic network.

An example of regulated metabolic network is shown in Fig. 1. This example is based on a highly simplified model of core carbon metabolism, originally proposed in [9]. At the metabolic level (Fig. 1a), there are 9 metabolites and $m = 9$ reactions. The internal metabolites are Int = {A, D, E, O2, ATP, NADH}, the external metabolites are Ext = {Carbon1, Carbon2, Oxygen}. All the $k = 3$ external metabolites are input metabolites, Ext = Inp. The set of irreversible reactions is $\mathcal{R}$ = {Tc1, Tc2, To2, Td, Te, Growth, Rres, R6, R7}. The stoichiometric coefficients are also given in Fig. 1a. By default, they are set to 1, except for the reactions $R6$ and $R7$.

The regulatory level (Fig. 1b) of the regulated metabolism introduces $d = 2$ regulatory proteins: $\mathcal{P}$ = {RPcl, RPO2}. Thus, the Boolean network f is of dimension $n = k+m+d = 14$. It consists of 14 functions (see Fig. 1b) which map a Boolean vector $x = (x_{\mathrm{Carbon1}}, x_{\mathrm{Carbon2}}, x_{\mathrm{Oxygen}}, x_{\mathrm{RPcl}}, x_{\mathrm{RPO2}}, x_{\mathrm{Tc1}}, x_{\mathrm{Tc2}}, x_{\mathrm{To2}}, x_{\mathrm{Td}}, x_{\mathrm{Te}}, x_{\mathrm{Growth}}, x_{\mathrm{Rres}}, x_{\mathrm{R6}}, x_{\mathrm{R7}}) \in \mathbb{B}^n$ to a Boolean value in $\mathbb{B}$. The local functions associated with regulatory proteins in $\mathcal{P}$ involve only external metabolite variables. Among the 9 functions associated with reactions, only two (Tc2, Rres) are non-constant: they involve the two regulatory proteins.

The influence graph of the network is shown in Fig. 1c. Only the shown nodes (RPcl, RPO2, Tc2, Rres) have a non-constant local function or are used in the local function of another node (Carbon1, Oxygen). The influence graph



(a) *Metabolic Network*

(c) Influence graph $G(f)$ of the regulatory Boolean network f. Nodes without in-going or out-going edges are not represented. Positive edges are drawn in green with a regular tipping arrow, negative edges are drawn in red with a bar arrow.

	Regulatory proteins		**Input metabolites**		
Local function	$f_{\mathrm{RPO2}}(x)$	$f_{\mathrm{RPcl}}(x)$	$f_{\mathrm{Carbon1}}(x)$	$f_{\mathrm{Carbon2}}(x)$	$f_{\mathrm{Oxygen}}(x)$
Boolean rule	$\neg x_{\mathrm{Oxygen}}$	x_{Carbon1}	0	0	0

	Reactions								
Local function	$f_{\mathrm{Tc1}}(x)$	$f_{\mathrm{Tc2}}(x)$	$f_{\mathrm{To2}}(x)$	$f_{\mathrm{Td}}(x)$	$f_{\mathrm{Te}}(x)$	$f_{\mathrm{Growth}}(x)$	$f_{\mathrm{Rres}}(x)$	$f_{\mathrm{R6}}(x)$	$f_{\mathrm{R7}}(x)$
Boolean rule	1	$\neg x_{\mathrm{RPcl}}$	1	1	1	1	$\neg x_{\mathrm{RPO2}}$	1	1

(b) *Boolean Network*. All Boolean functions equal to 1 correspond to reactions which are not regulated by the Boolean network.

Fig. 1: **Example of regulated metabolic network**. In the metabolic network **(a)**, each node represents a metabolite and each hyperedge a reaction. For instance, the hyperedge R7 linking {A; NADH} to {E} models the reaction A + 3 NADH → 3 E. Integer values over hyperedges are stoichiometric coefficients, the default value is 1. **(b)** defines the Boolean network regulating the metabolic network in (a), with $x \in \mathbb{B}^n$ and $n = 14$. **(c)** shows the influence (or regulatory) graph of the Boolean network in (b), with square nodes denoting the regulatory proteins.

shows the multi-layered regulations of the network: external input metabolites (Carbon1, Oxygen) regulate regulatory proteins (RPcl, RPO2), which regulate reactions (Tc2, Rres).

2.2 Dynamic rFBA

Flux Balance Analysis (FBA) [19] returns an *optimal* metabolic steady state, according to a given linear objective function in the reaction fluxes. In the following, we assume that the objective function is to maximize the flux through a reaction *Growth*. For regulated metabolic networks, the rFBA framework [9] allows defining a discrete time series of optimal steady states, where regulatory variables can force reaction fluxes to be zero and input metabolite concentrations define upper bounds on uptake fluxes.

Definition 5. *Let $(\mathcal{N}, \mathcal{P}, f)$ be a regulated metabolic network with flux bounds $l_r, u_r \in \mathbb{R}, 0 \leq l_r \leq u_r$, for $r \in \mathcal{R}$. A* metabolic-regulatory steady state *is a triple $(v, w, x) \in \mathbb{R}^{|\mathcal{R}|} \times \mathbb{R}^{|\mathrm{Ext}|} \times \mathbb{B}^{|\mathrm{Inp}|+|\mathcal{R}|+|\mathcal{P}|}$ such that*

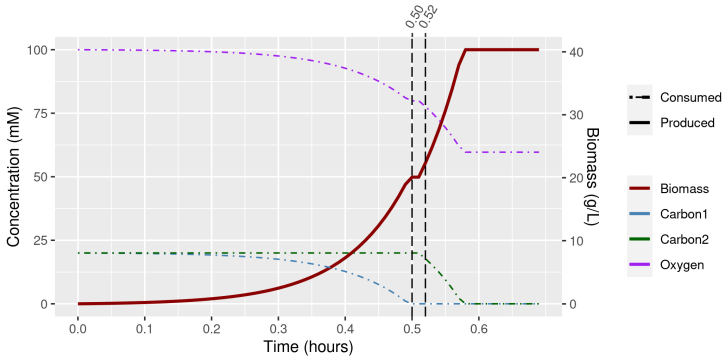
- $S_{\mathrm{Int},\mathcal{R}} \cdot v = 0$,
- *for each reaction $r \in \mathcal{R}$, $l_r \cdot x_r \leq v_r \leq u_r \cdot x_r$,*
- *for each input metabolite $m \in \mathrm{Inp}$ and each reaction $r \in \mathcal{R}$ with $S_{mr} < 0$, $v_r \leq$ uptake_bound(w_m), where uptake_bound(w_m) denotes the maximum flux through uptake reaction r, given the input metabolite concentration w_m.*

Two successive metabolic-regulatory steady states (v^k, w^k, x^k) at time t^k, and $(v^{k+1}, w^{k+1}, x^{k+1})$ at time t^{k+1}, are linked by the following relations:

1. The external metabolite concentrations w^{k+1} are obtained from the previous concentrations w^k by assuming the constant uptake/secretion fluxes v^k for the whole time period $[t^k, t^{k+1}]$.
2. The Boolean state x^{k+1} is obtained by applying the regulatory function f to the binarized input metabolites concentrations $x'_{\mathrm{Inp}} = \beta(w^{k+1}_{\mathrm{Inp}})$ at time t^{k+1}, together with the binarized reaction fluxes $x'_{\mathcal{R}} = \beta(v^k)$ and the Boolean values $x'_{\mathcal{P}} = x^k_{\mathcal{P}}$ of the regulatory proteins at time t^k, *i.e.*,
$$x^{k+1} = f(x')$$
3. $(v^{k+1}, w^{k+1}, x^{k+1})$ is a metabolic-regulatory steady state maximizing the flux through the *Growth* reaction, i.e., there is no metabolic-regulatory steady state (v', w^{k+1}, x^{k+1}) such that $v'_{Growth} > v^{k+1}_{Growth}$.

In this paper, we rely on the FlexFlux implementation of rFBA [18], which assumes a fixed time step τ between successive metabolic-regulatory steady states ($t^{k+1} - t^k = \tau$ for any k). The *Growth* reaction is assumed to reflect the growth of the cell. FlexFlux computes the evolution of the total biomass of the cell as $\mathrm{biomass}^{k+1} = \mathrm{biomass}^k \cdot e^{v^k_{Growth} \cdot \tau}$ (from a given initial $\mathrm{biomass}^0$). The maximum uptake fluxes of input metabolites $m \in \mathrm{Inp}$ at step k are defined as
$$\textit{uptake\_bound}(w_m) = w_m / (\mathrm{biomass}^k \cdot \tau).$$
Finally, the update of the external metabolite concentrations is computed as
$$w^{k+1}_m = w^k_m - (S_{mr} v^k_r / v^k_{Growth}) \cdot (\mathrm{biomass}^k - \mathrm{biomass}^{k+1}),$$
where $r \in \mathcal{R}$ is the uptake/secretion reaction for the external metabolite m ($S_{mr} < 0$ or $S_{mr} > 0$), which is assumed to be unique.

An example of a dynamic rFBA simulation using FlexFlux of the regulated metabolic network of Fig. 1 is shown in Fig. 2. It uses a time step of $0.01h$ and is initialized with 100 mM of Oxygen, 20 mM of Carbon1 and 20 mM of Carbon2. The simulation shown in Fig. 2a is composed of 70 metabolic steady states. By applying the binarization β, these 70 metabolic steady states correspond to 5



(a) Simulation showing the evolution of the concentrations of the external metabolites (Oxygen, Carbon1, Carbon2) and the production of biomass by the Growth reaction.

		External metabolites			Regulatory proteins		Reaction flows								
Time	w_{biomass}	w_{Carbon1}	w_{Carbon2}	w_{Oxygen}	x_{RPO2}	x_{RPcl}	v_{Tc1}	v_{Tc2}	v_{To2}	v_{Td}	v_{Te}	v_{Growth}	v_{Rres}	v_{R6}	v_{R7}
0.49	17.05	2.95	20.0	82.95	0	1	10.5	0.0	10.5	0.0	0.0	10.5	10.5	0.0	0.0
0.50	18.95	1.05	20.0	81.05	0	1	6.15	0.0	6.15	0.0	0.0	6.15	6.15	0.0	0.0
0.51	20.10	0.0	20.0	79.90	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.52	20.10	0.0	20.0	79.90	0	0	0.0	10.5	10.5	0.0	0.0	10.5	10.5	0.0	0.0
0.53	22.35	0.0	17.76	77.65	0	0	0.0	10.5	10.5	0.0	0.0	10.5	10.5	0.0	0.0

(b) Focus on the times from $0.49h$ to $0.53h$ in the simulation, showing the switch from Carbon1 to Carbon2 for biomass production.

Fig. 2: Dynamic rFBA simulation of the regulated metabolic network in Fig. 1. The simulation is done with FlexFlux and is initialized with 100mM of Oxygen, 20 mM of Carbon1, and 20 mM Carbon2. Tke time step is set to $0.01h$. The flux bounds are $\forall r \in \{\text{Tc1, Tc2}\}$, $(l_r, u_r) = (0, 10.5)$, $\forall r \in \{\text{Td, Te}\}$, $(l_r, u_r) = (0, 12.0)$, $\forall r \in \{\text{R6, R7, Rres, Growth}\}$, $(l_r, u_r) = (0, 9999)$ and for Oxygen, $(l_r, u_r) = (0, 15.0)$.

different binarized metabolic steady states, which are shown in Tab. 1. These binarized metabolic steady states capture the main features of the simulation.

More precisely, the simulation shows that until $0.5h$ only Carbon1 and Oxygen are consumed to produce biomass. This corresponds to a first time period where the behavior of the system is monotone: the binarized metabolic steady states are equal on this time range. The presence of Carbon1 activates the regulatory protein RPcl inhibiting the reaction Tc2 according to the regulatory rules. At $0.5h$, Carbon1 is depleted and the current Boolean state $x \in \mathbb{B}^{15}$ is such that $x_{\text{Carbon1}} = 0$, $x_{\text{RPcl}} = 1$, $x_{\text{Tc2}} = 0$ (second qualitative behavior with equal binarization of the metabolic steady states). At $0.51h$, as shown in Fig. 2b, the Boolean state x is updated to x' so that the Boolean state of RPcl becomes $x'_{\text{RPcl}} = f_{\text{RPcl}}(x) = x_{\text{Carbon1}} = 0$. The Boolean state of *Tc2* remains unchanged because $x_{\text{RPcl}} = 1$. No biomass is produced at $0.51h$. This corresponds to a third qualitative behavior. At $0.52h$, the Boolean state x' is updated to x'': all the node states remain unchanged except for $x''_{\text{Tc2}} = f_{\text{Tc2}}(x') = \neg x'_{\text{RPcl}} = 1$.

	External metabolites				Regulatory proteins		Reactions								
Time	$\overline{w_{\text{Biomass}}}$	$\overline{w_{\text{Carbon1}}}$	$\overline{w_{\text{Carbon2}}}$	$\overline{w_{\text{Oxygen}}}$	$\overline{x_{\text{RPO2}}}$	$\overline{x_{\text{RPcl}}}$	$\overline{v_{\text{Tc1}}}$	$\overline{v_{\text{Tc2}}}$	$\overline{v_{\text{To2}}}$	$\overline{v_{\text{Td}}}$	$\overline{v_{\text{Te}}}$	$\overline{v_{\text{Growth}}}$	$\overline{v_{\text{Rres}}}$	$\overline{v_{\text{R6}}}$	$\overline{v_{\text{R7}}}$
0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0
0.1	1	1	1	1	0	1	1	0	1	0	0	1	1	0	0
0.51	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0
0.52	1	0	1	1	0	0	0	1	1	0	0	1	1	0	0
0.59	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Table 1: Binarization of the metabolic steady states of simulation in Fig. 2. It contains the binarized values of the metabolic steady state computed by the rFBA simulation. A timepoint t appears in the table if and only if the binarization of the simulated steady state is different from the binarized metabolic steady state of time $t-1$.

This corresponds to a fourth qualitative behavior. The reaction *Tc2* is not inhibited anymore, and biomass is produced due to the uptake of Carbon2 and Oxygen (through *Tc2*, *Growth* and *Rres*) until Carbon2 depletion at $t = 0.59h$ (fifth qualitative behavior).

3 Boolean abstraction of dynamic rFBA

In the previous example, we illustrated how the simulation of a regulated metabolic network may generate time-periods for which the qualitative behavior is similar, meaning that the variation of all the metabolic variables is monotone and the Boolean values of the regulatory proteins are constant. In this section, we introduce a discrete definition of steady states to capture the monotone behaviors observed in rFBA simulations. This allows introducing a discretized form of rFBA, which will be used in the next section for the reverse-engineering framework.

3.1 Boolean metabolic steady states

Given a metabolic network $\mathcal{N} = (\text{Int}, \text{Ext}, \mathcal{R}, S)$, we derive a logical characterization of the notion of steady state, considering that reactions are either inactive or active, and metabolites either absent or present. This will result in a set of *Boolean* metabolic steady states that form an over-approximation of the continuous steady states.

We associate all reactions with propositional variables $\mathcal{V} = \{\overline{v_r}\}_{r\in\mathcal{R}}$. For each metabolite $m \in \text{Int} \uplus \text{Ext}$, we introduce a variable $\overline{z_m}^+$ as a Boolean abstraction of the production of m and a variable $\overline{z_m}^-$ as a Boolean abstraction of the consumption of m:

$$\forall m \in \text{Int} \uplus \text{Ext}, \quad \overline{z_m}^+ \stackrel{\text{def}}{=} \bigvee_{\substack{r\in\mathcal{R},\\ S_{mr}>0}} \overline{v_r}, \qquad \overline{z_m}^- \stackrel{\text{def}}{=} \bigvee_{\substack{r\in\mathcal{R},\\ S_{mr}<0}} \overline{v_r},$$

(where an empty disjunction is considered to be false).

For each internal metabolite m, we introduce a variable $\widehat{z_m}$ which is equal to 1 iff m is in a logical steady state:

$$\forall m \in \mathrm{Int}, \quad \widehat{z_m} \stackrel{\text{def}}{=} (\overline{z_m}^{+} \Leftrightarrow \overline{z_m}^{-}).$$

For the external metabolites, we introduce propositional variables $\mathcal{V}_{ext} = \{\overline{z_m}\}_{m\in\mathrm{Ext}}$ indicating whether or not m is present in the environment. The formula

$$\widehat{\mathcal{N}}_{\mathrm{Ext}} \stackrel{\text{def}}{=} \bigwedge_{m\in\mathrm{Ext}} (\overline{z_m}^{-} \Rightarrow \overline{z_m})$$

then states that an external metabolite can only be consumed if it is present in the environment.

Definition 6 (Boolean metabolic steady state). *A* Boolean metabolic steady state *of a metabolic network* $\mathcal{N} = (\mathrm{Int}, \mathrm{Ext}, \mathcal{R}, S)$ *is a Boolean vector* $\hat{\nu} \in \mathbb{B}^{|\mathrm{Ext}|+|\mathcal{R}|}$ *which is a satisfying assignment of the following logical steady state formula:*

$$\widehat{\mathcal{N}} \stackrel{def}{=} \widehat{\mathcal{N}}_{\mathrm{Ext}} \wedge \bigwedge_{m\in\mathrm{Int}} \widehat{z_m}$$

We denote by $\mathrm{MSS}^{\mathbb{B}}(\mathcal{N}) \subseteq \mathbb{B}^{|\mathrm{Ext}|+|\mathcal{R}|}$ *the set of all the Boolean metabolic steady states of the metabolic network* $\mathcal{N}$*.*

As an immediate consequence of this definition, we get the following property:

Property 1. For each metabolic-regulatory steady state (v, w, x) of the regulated metabolic network $(\mathcal{N}, \mathcal{P}, f)$, the binarized value $\beta(w, v)$ of the external metabolite concentrations w and the reaction fluxes v is a Boolean metabolic steady state, i.e., $\beta(w, v) \in \mathrm{MSS}^{\mathbb{B}}(\mathcal{N})$.

Note that the converse is not true: since the logical characterization neglects the stoichiometry, Boolean metabolic steady states may have no real-valued counterpart.

Applied to the example, the internal metabolic constraints are the following:
$\overline{z_{\mathrm{A}}}^{+} = \overline{v_{\mathrm{Tc1}}} \vee \overline{v_{\mathrm{Tc2}}}, \quad \overline{z_{\mathrm{A}}}^{-} = \overline{v_{\mathrm{R6}}} \vee \overline{v_{\mathrm{R7}}} \vee \overline{v_{\mathrm{Growth}}}$
$\overline{z_{\mathrm{D}}}^{+} = \overline{v_{\mathrm{R6}}}, \quad \overline{z_{\mathrm{D}}}^{-} = \overline{v_{\mathrm{Td}}}, \quad \overline{z_{\mathrm{E}}}^{+} = \overline{v_{\mathrm{R7}}}, \quad \overline{z_{\mathrm{E}}}^{-} = \overline{v_{\mathrm{Te}}}$
$\overline{z_{\mathrm{O2}}}^{+} = \overline{v_{\mathrm{To2}}}, \quad \overline{z_{\mathrm{O2}}}^{-} = \overline{v_{\mathrm{Rres}}}$
$\overline{z_{\mathrm{ATP}}}^{+} = \overline{v_{\mathrm{R6}}} \vee \overline{v_{\mathrm{Rres}}}, \quad \overline{z_{\mathrm{ATP}}}^{-} = \overline{v_{\mathrm{Growth}}}$
$\overline{z_{\mathrm{NADH}}}^{+} = \overline{v_{\mathrm{Growth}}}, \quad \overline{z_{\mathrm{NADH}}}^{-} = \overline{v_{\mathrm{R7}}} \vee \overline{v_{\mathrm{Rres}}}$
The logical steady state constraints equivalent to $\widehat{\mathcal{N}} = 1$ are obtained by gathering contraints on internal and external metabolites:
$\overline{v_{\mathrm{Tc1}}} \vee \overline{v_{\mathrm{Tc2}}} = \overline{v_{\mathrm{R6}}} \vee \overline{v_{\mathrm{R7}}} \vee \overline{v_{\mathrm{Growth}}}$
$\overline{v_{\mathrm{R6}}} = \overline{v_{\mathrm{Td}}} \qquad \overline{v_{\mathrm{R7}}} = \overline{v_{\mathrm{Te}}} \qquad \overline{v_{\mathrm{To2}}} = \overline{v_{\mathrm{Rres}}}$
$\overline{v_{\mathrm{R6}}} \vee \overline{v_{\mathrm{Rres}}} = \overline{v_{\mathrm{Growth}}} \qquad \overline{v_{\mathrm{R7}}} \vee \overline{v_{\mathrm{Rres}}} = \overline{v_{\mathrm{Growth}}}$
$\overline{v_{\mathrm{Tc1}}} \Rightarrow \overline{z_{\mathrm{Carbon1}}} \qquad \overline{v_{\mathrm{Tc2}}} \Rightarrow \overline{z_{\mathrm{Carbon2}}} \qquad \overline{v_{\mathrm{To2}}} \Rightarrow \overline{z_{\mathrm{Oxygen}}}$

From these equations, we deduce that there are 38 Boolean metabolic steady states for the example shown in Fig. 1. These Boolean metabolic steady states are detailed in Appendix A. Among them, we recover the five binarized metabolic-regulatory steady states (Table 1) appearing in the rFBA simulations of Fig.2.

3.2 Boolean dynamics

Using the logical characterization of metabolic steady states, we define a Boolean counterpart of dynamic rFBA (Sect. 2.2). A Boolean state of the regulated metabolic network $(\mathcal{N}, \mathcal{P}, f)$ assigns a Boolean value to external metabolites, reactions, and regulatory proteins, which gives a Boolean vector of dimension $n = k + m + d$. Such a Boolean state $x \in \mathbb{B}^n$ should match with a Boolean metabolic steady state. Denoting by $\mathcal{M} = \text{Ext} \cup \mathcal{R}$ the external metabolites and reactions, $x_{\mathcal{M}}$ should verify the Boolean metabolic steady state constraints described in the previous section ($x_{\mathcal{M}} \in \text{MSS}^{\mathbb{B}}(\mathcal{N})$). The general idea is then to capture the possible successions of such Boolean states, subject to the regulations through the regulatory proteins specified by the Boolean network f.

A key ingredient of dynamic rFBA is the objective function to maximize, typically the fluxes of reactions producing biomass. However, at the Boolean level, it is not possible to directly rank metabolic steady states according to their biomass production, as this will be either absent or present. Thus, a specific *Boolean objective function* has to be provided to score a Boolean metabolic steady state. This takes the form of a function $\hat{o}$ mapping Boolean metabolic steady states to natural numbers: $\hat{o} : \mathbb{B}^{k+m} \to \mathbb{N}$. The Boolean dynamics will only select Boolean metabolic steady states maximizing this supplied objective.

When considering possible next states, it is crucial to account for those where the input metabolites change their value. Hereafter, we consider any possible change.

The Boolean dynamic rFBA is formalized by a function $\text{next}^{\mathbb{B}}_{(\mathcal{N},\mathcal{P},f,\hat{o})}$ which associates any Boolean state of the regulated metabolic network to a set of admissible next states:

Definition 7 (Boolean dynamic rFBA: $\text{next}^{\mathbb{B}}_{(\mathcal{N},\mathcal{P},f,\hat{o})} : \mathbb{B}^n \to 2^{\mathbb{B}^n}$). *For any Boolean states $x, y \in \mathbb{B}^n$, $y \in \text{next}^{\mathbb{B}}_{(\mathcal{N},\mathcal{P},f,\hat{o})}(x)$ if and only if for $x' = (y_{\text{Inp}}, x_{\mathcal{R}\cup\mathcal{P}}) \in \mathbb{B}^n$,*

1. *the values of the regulatory proteins are computed synchronously from x' according to f: $y_{\mathcal{P}} = f_{\mathcal{P}}(x')$,*
2. *y matches with a Boolean metabolic steady state: $y_{\mathcal{M}} \in Z(x')$, and*
3. *the matching Boolean metabolic steady state maximizes the supplied objective function: $\forall y'_{\mathcal{M}} \in Z(x'), \hat{o}(y_{\mathcal{M}}) \geq \hat{o}(y'_{\mathcal{M}})$.*

Here $Z(x') = \{z \in \text{MSS}^{\mathbb{B}}(\mathcal{N}) \mid z_{\text{Inp}} = x'_{\text{Inp}}, z_{\mathcal{R}} \preceq f_{\mathcal{R}}(x')\}$ is the set of Boolean metabolic steady states that match with the value of external metabolites and with the regulations from x'.

Let us consider the regulated metabolic network from Fig. 1. It appears that the steady states maximizing the growth maximize the input fluxes. Thus, we set the Boolean objective function $\hat{o}$ as the sum of input reactions:

$$\hat{o}(x) = x_{\mathrm{Tc1}} + x_{\mathrm{Tc2}} + x_{\mathrm{To2}} \ .$$

Consider the Boolean state from Table 1 at time 0, which we name x, and the next Boolean state at time 0.51, which we name y, with the same input metabolite values ($x_{\mathrm{Inp}} = y_{\mathrm{Inp}}$). Using the notation from the above definition, we set $x' = x$. Imagine the case where no reactions is regulated, i.e., the regulatory BN is of the form $f'_r(x) = 1$ for every $r \in \mathcal{R}$. Among the Boolean metabolic steady states z matching the input values ($z_{\mathrm{Inp}} = x'_{\mathrm{Inp}}$), the ones that maximize $\hat{o}$ always verify $z_{\mathrm{Tc2}} = 1$ (Boolean metabolic steady states 26, 29, 32, 38 in the Table 3 in Appendix A), which does not match with y. Thus y would not be an admissible next state.

Considering now the regulatory BN f of Fig. 1, we obtain $f_{\mathrm{Tc2}}(x') = \neg x'_{\mathrm{RPcl}} = 0$ and for each other reaction $r \in \mathcal{R} \setminus \{\mathrm{Tc2}\}$, $f_r(x') = 1$. The set $Z(x')$ contains 4 matching optimal Boolean steady states (rows 25, 28, 31, 37 of Table A.3), among them the one matching with y. Thus $y \in \mathrm{next}^{\mathbb{B}}_{(\mathcal{N},\mathcal{P},f,\hat{o})}(x)$.

Let x be now the Boolean state at time 0.1, and y the next Boolean state at time 0.51, where the input metabolites have a different state (Carbon1 switched to 0). Let x' be equal to x except for the input metabolites, which are equal to y_{Inp}. We obtain that $f_{\mathrm{RPO2,RPcl}}(x') = (\neg x'_{\mathrm{Oxygen}}, x'_{\mathrm{Carbon1}}) = (0,0) = y_{\mathrm{RPO2,RPcl}}$. Moreover, $f_{\mathrm{Tc2}}(x') = \neg x'_{\mathrm{RPcl}} = 0$ and for each other reaction $r \in \mathcal{R}$, $r \neq \mathrm{Tc2}$, $f_r(x') = 1$. In this case, there is only one Boolean metabolic steady state z such that $z_{\mathrm{Inp}} = x'_{\mathrm{Inp}}$ and $z_{\mathcal{R}} \preceq f_{\mathcal{R}}(x')$. It appears that it matches with y, i.e., $z = y_{\mathcal{M}}$; thus $y \in \mathrm{next}^{\mathbb{B}}_{(\mathcal{N},\mathcal{P},f,\hat{o})}(x)$.

4 Inference of regulations from rFBA time series

Given sequences of metabolic-regulatory steady states obtained by dynamic rFBA from a ground-truth regulated metabolic network under different conditions, our objective is to infer all the regulatory Boolean networks that can reproduce the observed behaviors. Besides the ground-truth model, the inference may suggest alternative regulatory logics.

Definition 8 (Search domain for BNs). *The search domain for BNs, denoted by $\mathbb{F}$, is constrained by an influence graph $\mathcal{G}$: any candidate $f \in \mathbb{F}$ should satisfy $G(f) \subseteq \mathcal{G}$,* i.e. *uses at most the influences allowed in $\mathcal{G}$. Moreover, we assume that f is locally monotone.*

Typically, $\mathcal{G}$ contains the putative influences from and to regulatory proteins. In our case study, $\mathcal{G}$ is obtained from the ground-truth regulatory model f° by "forgetting" the sign of influences (for each $(i,s,j) \in G(f^\circ)$, $\{(i,+,j),(i,-,j)\} \subseteq \mathcal{G}$), and adding putative influences.

Our inference problem mixes both linear constraints for characterizing the optimal steady states of the metabolic network with Boolean constraints for

characterizing the value changes of regulatory proteins. To express the inference problem, we rely on the Boolean abstraction of dynamic rFBA presented in the previous section .

4.1 Approximation as a Boolean satisfiability problem

We propose a relaxation of the inference problem by the means of the Boolean dynamic rFBA interpretation given in Sect. 3.

Inputs of the relaxed inference problem. The inputs of the problem are **(i)** a metabolic network $\mathcal{N}$ and a set of regulatory proteins $\mathcal{P}$, **(ii)** sequences of metabolic-regulatory steady states, represented by sets of pairs (s^t, s^{t+1}), with $s^t = (v^t, w^t, x^t)$ and $s^{t+1} = (v^{t+1}, w^{t+1}, x^{t+1})$ following the notation from Def. 5: the observed changes of metabolic-regulatory steady states are given as $T \subseteq \mathbb{S}\times\mathbb{S}$ with $\mathbb{S} = \mathbb{R}^{|\text{Inp}|+|\mathcal{R}|} \times \mathbb{B}^{|RPs|}$, **(iii)** a domain of putative regulatory BNs $\mathbb{F}$ of dimension $n = |\text{Inp}| + |\mathcal{R}| + |\mathcal{P}|$, **(iv)** a Boolean state objective score $\hat{o} : \mathbb{B}^n \to \mathbb{N}$.

Relaxed inference problem The relaxed inference problem consists then in identifying the $f \in \mathbb{F}$ such that for each $(s, s') \in T$,

$$\beta(s') \in \text{next}^{\mathbb{B}}_{(\mathcal{N},\mathcal{P},f,\hat{o})}(\beta(s)).$$

Formulation as a satisfiability problem. Relying on the Boolean dynamic rFBA abstraction, the inference problem boils down to a satisfiability problem in propositional Boolean logic using two levels of quantifiers (2-QBF):

$$\begin{aligned}\exists f \in \mathbb{F}, \forall (s,s') \in T, \exists y \in \text{MSS}^{\mathbb{B}}(\mathcal{N}),\ & y_{\text{Inp}} = x'_{\text{Inp}}, y_{\mathcal{P}} = f_{\mathcal{P}}(x'),\ y_{\mathcal{R}} \preceq f_{\mathcal{R}}(x'),\\ & \forall z \in \text{MSS}^{\mathbb{B}}(\mathcal{N}), (z_{\text{Inp}} \neq x'_{\text{Inp}} \vee z_{\mathcal{P}} \neq f_{\mathcal{P}}(x') \vee z_{\mathcal{R}} \not\preceq f_{\mathcal{R}}(x') \vee \hat{o}(z) \leq \hat{o}(y))\end{aligned}$$

with $x' \in \mathbb{B}^n$ defined as $x'_{\text{Inp}} = \beta(s')_{\text{Inp}}$ and $x'_{\mathcal{R}\cup\mathcal{P}} = \beta(s)_{\mathcal{R}\cup\mathcal{P}}$.

Note that without the Boolean optimization criteria $\hat{o}$ (equivalently $\hat{o}(z) = c$), the problem reduces to a SAT problem where the only constraints relate to the local functions of the regulatory proteins:

$$\exists f \in \mathbb{F}, \exists y \in \text{MSS}^{\mathbb{B}}(\mathcal{N}),\ y_{\text{Inp}} = x'_{\text{Inp}},\ y_{\mathcal{P}} = f_{\mathcal{P}}(x')$$

Indeed, $y_{\mathcal{R}} \preceq f_{\mathcal{R}}(x')$ is always verified whenever $f_r(x) = 1$ for each $r \in \mathcal{R}$.

Since the Boolean dynamic rFBA gives an over-approximation of metabolic steady states, and even assuming that the Boolean objective function $\hat{o}$ matches with the optimal metabolic steady states, our formulation leads to an approximation of admissible regulatory BN f: it may happen that a spurious Boolean metabolic steady state (having no real counter part) has a strictly higher value with $\hat{o}$ than non-spurious ones.

4.2 Implementation in Answer-Set Programming

Answer-Set Programming (ASP) [1,12] is a declarative framework allowing solving combinatorial satisfaction problems. It relies on the stable model semantics [10]. The basic idea of ASP is to express a problem in a logical format so that the (logic) models of its representation provide the solutions to the original problem. Problems are expressed as logic programs (first order logic predicates expressed with rules with the shape `<head> :- <body> .`). Stable models of the logic programs are referred to as *answer sets*. Although determining whether a program has an answer set is the fundamental decision problem in ASP, modern ASP solvers like clingo [13] support various combinations of reasoning modes, among them, regular and projective enumeration, intersection and union, multi-criteria optimization and subset minimal and maximal model enumeration [15].

The stable model semantics of ASP combined with disjunctive programming are the key ingredients that enable expressing two quantification levels Boolean formulas (2-QBF problem), *i.e.* $\exists x$, $\forall y$, $\phi(x, y)$ where $\phi(x, y)$ is a quantifier-free propositional formula (Σ_2^{P}-complete) [10]. The encoding of 2-QBF relies on the so-called *saturation technique* [11,14]. Essentially, for fixed x and y, the encoding ensures that a maximal (saturated) answer-set is returned if and only if $\phi(x, y)$. Thus, whenever there exists y such that $\phi(x, y)$ does not hold (counter-example), a smaller answer-set is returned. Following the subset-minimal stable semantics, the 2-QBF problem is satisfiable if and only if only saturated answer-set are subset-minimal.

5 Case study

As a proof of concept, we apply our inference framework to the simplified core carbon metabolism described in Fig. 1. First, from this ground-truth model, we generate sample dynamic rFBA simulations for different input conditions, reproducing existing biological observations [9]. Next we take these simulations as input for our method, together with an influence graph extending the one from the ground truth model with additional putative regulations. Using our inference method, we then enumerate BNs that are compatible with both the simulations and the influence graph. The results show that the ground truth model is well recovered, together with some alternative BNs. In particular, a simpler BN matching the data is identified, which uses fewer regulations. It turns out that the missing regulation is not needed to reproduce the expected biological behaviour. Our implementation relying on the ASP solver CLINGO [13] together with the case study is available at `https://github.com/bioasp/boolean-caspo-flux`. They can be reproduced using the notebooks and docker image at `https://doi.org/10.5281/zenodo.5060984`.

Input simulations We designed six dynamic r-FBA simulations of the BN of Fig. 1(b) to mimic the studies of the core carbon metabolism in [9]. They correspond to different sets of initially available input metabolites and regulatory proteins (Table 3a, and Fig. 4 in Appendix B). For instance, Experiment 1

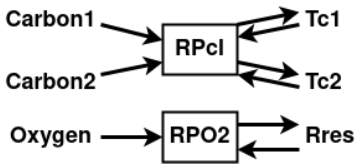
assumes that all input metabolites (Carbon1, Carbon2, Oxygen) are available. Experiment 2 assumes that Carbon1 and Carbon2 are present at initialization but not Oxygen.

For each case, we use FlexFlux with an initial biomass value of 0.1 and a time step of 0.01 to simulate the system. Each of the 6 simulations involves 200 metabolic steady states. For initial external metabolite values ($\overline{z_{\text{Carbon1}}}$, $\overline{z_{\text{Carbon2}}}$, $\overline{z_{\text{Oxygen}}}$), the regulatory proteins are initialized such that $x_{\text{RPcl}} = \overline{z_{\text{Carbon1}}}$ and $x_{\text{RPO2}} = \neg\overline{z_{\text{Oxygen}}}$ (Table 3a). Each simulation $S = \{(v, w, x)_0, ..., (v, w, x)_{200}\}$ includes 201 continuous metabolic-regulatory steady states (1 for the initialization and 200 for the simulation). The simulations are then binarized with $S^{\mathbb{B}} = \{(\overline{v_t}, \overline{z_t}) = \beta((v_t, w_t)) \mid \forall v_t \in S\}$, and consecutive identical Boolean states are removed. Table 1 shows the binarized metabolic-regulatory steady states from the simulation of the first experiment. From the 201 continuous metabolic steady states, 5 Boolean metabolic-regulatory steady states remain, corresponding to the time steps {0, 1, 51, 52, 59} (see Table 4 in Appendix B for the resulting states in each simulation).

Candidate models The search domain $\mathbb{F}$ for the candidate BNs is delimited by the influence graph $\mathcal{G}$ of Fig. 3b, which extends the influence graph from the ground-truth model by additional putative regulations, and by relaxing the sign constraints. Since the influence graph $G(f)$ of the ground-truth BN f is included in $\mathcal{G}$, we have $f \in \mathbb{F}$. In addition, $\mathbb{F}$ contains all the BNs such that $f_i(x) = 1$, for all $i \in \text{Inp} \cup \mathcal{R} \setminus \{\text{Tc1, Tc2, Rres}\}$. Furthermore, f_{RPcl} can depend on Carbon1, Carbon2, Tc1, and Tc2, f_{RPO2} can depend on Oxygen, Rres, f_{Tc1} and f_{Tc2} can depend on RPcl, and f_{Rres} can depend on Rres. Overall, $\mathbb{F}$ contains 1 944 320 BNs.

	Input Metabolite			Regulatory Protein	
Experiment	$\overline{z_{\text{Carbon1}}}$	$\overline{z_{\text{Carbon2}}}$	$\overline{z_{\text{Oxygen}}}$	x_{RPcl}	x_{RPO2}
1	1	1	1	1	0
2	1	1	0	1	1
3	0	1	0	0	1
4	1	0	0	1	1
5	1	0	1	1	0
6	0	1	1	0	0

(a) Initial states of the six rFBA simulations used to create the dataset for the case study.



(b) Influence graph $\mathcal{G}$ delimiting the domain of putative regulatory BNs $\mathbb{F}$. Nodes without in-going or out-going edges are not represented. Black regular tipping arrows are unsigned edges, *i.e.* both positive and negative edges.

Fig. 3: Input data for the case study. Table (a) summarizes the experimental conditions used to generate the input simulations. Figure (b) shows the influence graph delimiting the search domain for the inference problem.

	$f_{\mathrm{RPO2}}(x)$	$f_{\mathrm{RPcl}}(x)$	$f_{\mathrm{Tc1}}(x)$	$f_{\mathrm{Tc2}}(x)$	$f_{\mathrm{Rres}}(x)$	Subset minimal	Ground truth
Model 1	$\neg x_{\mathrm{Oxygen}}$	x_{Carbon1}	1	$\neg x_{\mathrm{RPcl}}$	1	✓	
Model 2	$\neg x_{\mathrm{Oxygen}}$	x_{Carbon1}	1	$\neg x_{\mathrm{RPcl}}$	$\neg x_{\mathrm{RPO2}}$		✓
Model 3	$\neg x_{\mathrm{Oxygen}}$	x_{Carbon1}	x_{RPcl}	$\neg x_{\mathrm{RPcl}}$	1		
Model 4	$\neg x_{\mathrm{Oxygen}}$	x_{Carbon1}	x_{RPcl}	$\neg x_{\mathrm{RPcl}}$	$\neg x_{\mathrm{RPO2}}$		

Table 2: Inferred models having subset minimal local functions. The not shown local functions $f_{\mathrm{Carbon1}}(x)$, $f_{\mathrm{Carbon2}}(x)$, $f_{\mathrm{Oxygen}}(x)$, $f_{\mathrm{To2}}(x)$, $f_{\mathrm{Td}}(x)$, $f_{\mathrm{Te}}(x)$, $f_{\mathrm{Growth}}(x)$, $f_{\mathrm{R6}}(x)$, $f_{\mathrm{R7}}(x)$ are set to 1.

Boolean objective function Our inference framework requires defining an objective function $\hat{o}$ over the Boolean metabolic steady states. Given the set of input metabolites Inp = {Carbon1, Carbon2, Oxygen}, the objective function is defined as $\hat{o}(x) = \sum_{e \in \mathrm{Inp}} x_{\mathrm{e}}, \forall x \in \mathrm{MSS}^{\mathbb{B}}(\mathcal{N})$. This is motivated by the observation that maximizing biomass production often corresponds to maximizing the uptake of inputs according to the QSS constraints. Therefore, if an available input metabolite is not used in the observed Boolean metabolic network, then this must be explained by at least one regulation. This objective function allows capturing more refined behaviors at the discrete level than a standard biomass optimization function, which may be too rough when considering discretized values.

Results Applying the constraints from above allows inferring 40 models. All these models share 3 local functions whose value is not constantly 1 ($f_{\mathrm{RPO2}}(x)$, $f_{\mathrm{RPcl}}(x)$, $f_{\mathrm{Tc2}}(x)$). They also share 9 local functions equal to 1 ($f_{\mathrm{Carbon1}}(x)$, $f_{\mathrm{Carbon2}}(x)$, $f_{\mathrm{Oxygen}}(x)$, $f_{\mathrm{To2}}(x)$, $f_{\mathrm{Td}}(x)$, $f_{\mathrm{Te}}(x)$, $f_{\mathrm{Growth}}(x)$, $f_{\mathrm{R6}}(x)$, $f_{\mathrm{R7}}(x)$). Finally, 2 functions can be set both to 1 or different from 1 according to the model. The 4 *smallest* inferred models are described in Table 2. They can be considered as the smallest because each local function f_i of these 4 models is contained in the local function f_i of the 36 other models. Note that the ground truth, *i.e.* the model used to generate the input data, is correctly inferred (Model 2).

As we represent the local Boolean functions using their disjunctive normal form (DNF), we can focus on the *simplest* models by looking at the *subset-minimal* ones: a Boolean function f_i is smaller than a Boolean function g_i if each of the clauses of f_i is a subset of a clause of g_i. In this case study, there is a single subset-minimal model: the BN 1 of Table 2. The two functions $f_{\mathrm{Rres}}(x)$, $f_{\mathrm{Tc1}}(x)$ are set to 1 due to the subset-minimal constraint. The inferred model is thus $f_{\mathrm{RPO2}}(x) = \neg x_{\mathrm{Oxygen}}$, $f_{\mathrm{RPcl}}(x) = x_{\mathrm{Carbon1}}$, $f_{\mathrm{Tc2}}(x) = \neg x_{\mathrm{RPcl}}$ and all the others local functions are set to 1. Note that only $f_{\mathrm{Rres}}(x)$ differs between the inferred subset-minimal model and the ground truth model.

In order to check whether this subset-minimal model could be considered as an alternative to the ground truth one, we performed dynamic rFBA simulations with the six experimental conditions described in Table 3a. We observe that the resulting time series are strictly identical to the simulations of the ground truth model used to generate the dataset. This suggests that the regulation on *Rres* is not necessary to reproduce the observed behaviours. The proposed subset-

minimal model allows inferring all the needed regulations and can be considered as the simplest regulated metabolic model matching the experimental conditions of Table 3a. Already in [9], the authors recognize that unlike others regulations, *Rres* "regulation is not necessary for the solution". Biologically, this regulation is only present to ensure that unnecessary enzymes decay. However, since enzyme amounts are not explicitly represented in the rFBA framework, the dataset does not reflect this biologic behavior, making it impossible to infer properly the regulation. Taking into account enzymatic resources using methods such as r-deFBA [17], should allow solving this issue. However, the inference approach will also have to be adapted to this kind of extended metabolic modeling.

6 Discussion

We proposed a formal framework to infer Boolean rules for the regulation of a metabolic network. The formulation of dynamic rFBA as sequence of steady states of the regulated metabolic network enables inferring the Boolean rules from time series under multiple conditions. A proof of concept was performed on the simulation of the diauxic shift in carbon metabolism on a small model.

Our method builds on a Boolean abstraction of the dynamic rFBA framework. It enables a formulation of the inference problem as a pure Boolean satisfiability problem using two levels of quantifiers, which can be efficiently solved using Answer Set Programming. One important parameter is the Boolean objective function, which aims at identifying Boolean metabolic steady states that match the optimal real-valued ones. This function is currently specified manually, based on biological expertise. Future work may explore how to derive an objective function automatically. An alternative direction is to solve directly the inference problem by mixing linear programming and Boolean constraints. Future work will investigate the scalability of solving these different inference problems.

Several other perspectives are to be explored. First, all regulations were considered as synchronous, which may not be the case *in vivo*, where regulations can have different time scales. This choice was actually imposed by the use of the FlexFlux implementation. Nevertheless, our method can be easily adapted to support fully-asynchronous and asynchronous updating modes, enabling potential alternative solutions. Second, the production and degradation times of regulatory proteins and enzymes were not taken into account. Moreover, the regulations were considered to be binary. However, we know that metabolism proceeds by finer regulations than the abstraction proposed here, as captured for instance by regulatory dynamic enzyme-cost FBA [17].

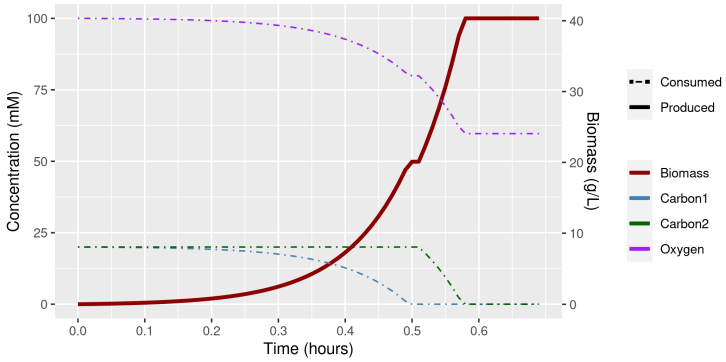
Acknowledgments Work of LC and CB is supported by the French Laboratory of Excellence project "TULIP" (grant number ANR-10-LABX-41; ANR-11-IDEX-0002-02). Work of LP is supported by the French Agence Nationale pour la Recherche (ANR) in the scope of the project "BNeDiction" (grant number ANR-20-CE45-0001).

A Binarized metabolic steady state

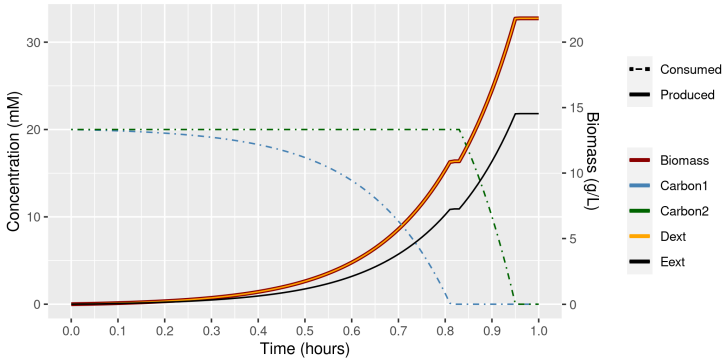
	External metabolites			Reactions									
	$\overline{z_{\text{Carbon1}}}$	$\overline{z_{\text{Carbon2}}}$	$\overline{z_{\text{Oxygen}}}$	$\overline{v_{\text{Tc1}}}$	$\overline{v_{\text{Tc2}}}$	$\overline{v_{\text{To2}}}$	$\overline{v_{\text{Td}}}$	$\overline{v_{\text{Te}}}$	$\overline{v_{\text{Growth}}}$	$\overline{v_{\text{Rres}}}$	$\overline{v_{\text{R6}}}$	$\overline{v_{\text{R7}}}$	**Experimentation**
1	0	0	0	0	0	0	0	0	0	0	0	0	2, 3, 4
2	0	0	1	0	0	0	0	0	0	0	0	0	1, 5, 6
3	0	1	0	0	0	0	0	0	0	0	0	0	2, 3
4	0	1	0	0	1	0	1	1	1	0	1	1	2, 3
5	0	1	1	0	0	0	0	0	0	0	0	0	1, 6
6	0	1	1	0	1	1	0	0	1	1	0	0	1, 6
7	0	1	1	0	1	1	0	1	1	1	0	1	
8	0	1	1	0	1	1	1	0	1	1	1	0	
9	0	1	1	0	1	0	1	1	1	0	1	1	
10	0	1	1	0	1	1	1	1	1	1	1	1	
11	1	0	0	0	0	0	0	0	0	0	0	0	4
12	1	0	0	1	0	0	1	1	1	0	1	1	4
13	1	0	1	0	0	0	0	0	0	0	0	0	5
14	1	0	1	1	0	1	0	0	1	1	0	0	5
15	1	0	1	1	0	1	0	1	1	1	0	1	
16	1	0	1	1	0	1	1	0	1	1	1	0	
17	1	0	1	1	0	0	1	1	1	0	1	1	
18	1	0	1	1	0	1	1	1	1	1	1	1	
19	1	1	0	0	0	0	0	0	0	0	0	0	2
20	1	1	0	0	1	0	1	1	1	0	1	1	
21	1	1	0	1	0	0	1	1	1	0	1	1	2
22	1	1	0	1	1	0	1	1	1	0	1	1	
23	1	1	1	0	0	0	0	0	0	0	0	0	1
24	1	1	1	0	1	1	0	0	1	1	0	0	
25	1	1	1	1	0	1	0	0	1	1	0	0	1
26	1	1	1	1	1	1	0	0	1	1	0	0	
27	1	1	1	0	1	1	0	1	1	1	0	1	
28	1	1	1	1	0	1	0	1	1	1	0	1	
29	1	1	1	1	1	1	0	1	1	1	0	1	
30	1	1	1	0	1	1	1	0	1	1	1	0	
31	1	1	1	1	0	1	1	0	1	1	1	0	
32	1	1	1	1	1	1	1	0	1	1	1	0	
33	1	1	1	0	1	0	1	1	1	0	1	1	
34	1	1	1	1	0	0	1	1	1	0	1	1	
35	1	1	1	1	1	0	1	1	1	0	1	1	
36	1	1	1	0	1	1	1	1	1	1	1	1	
37	1	1	1	1	0	1	1	1	1	1	1	1	
38	1	1	1	1	1	1	1	1	1	1	1	1	

Table 3: All the Boolean metabolic steady states admissible for the metabolic network $\mathcal{N}$ show Fig. 1a. The external metabolite *Biomass* is not shown since its value can be both 0 and 1 for each Boolean metabolic steady state. The experimentation column indicates the numbers of the experiments where the Boolean metabolic steady states occurs.

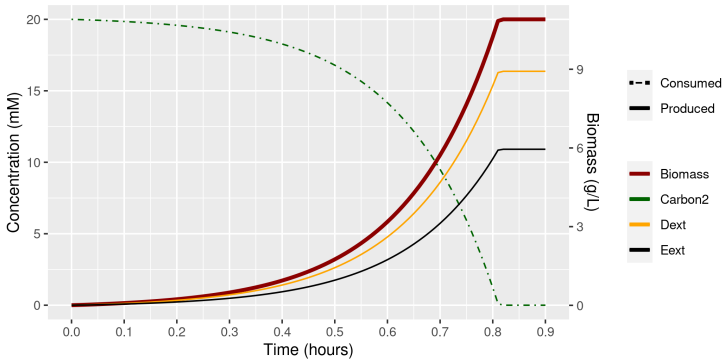
B Experiments and simulations



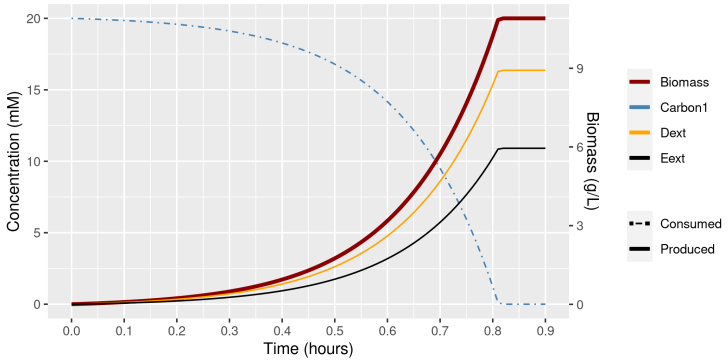
(a) Simulation of experiment 1.



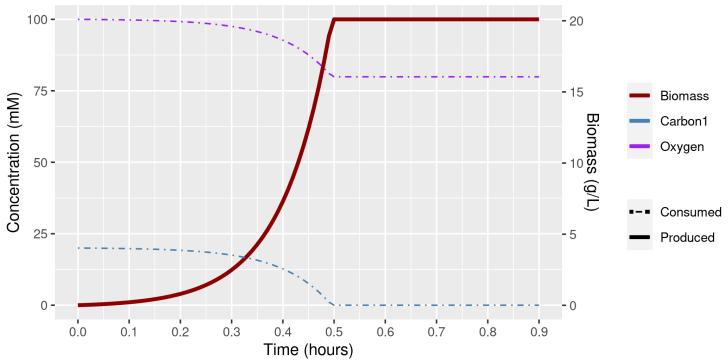
(b) Simulation of experiment 2.



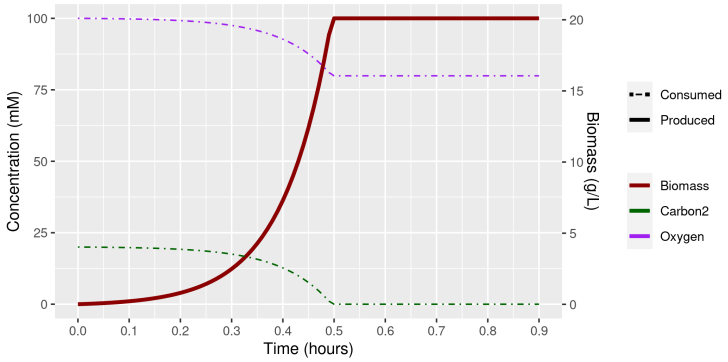
(c) Simulation of experiment 3.



(d) Simulation of experiment 4.



(e) Simulation of experiment 5.



(f) Simulation of experiment 6.

Fig. 4: Simulation made with FlexFlux of the regulated metabolic network in Fig. 1 for each experiment (Table 3a). Time step is set to 0.01. Reaction domains are $\forall r \in \{\text{Tc1, Tc2}\}$, $(l_r, u_r) = (0, 10.5)$, $\forall r \in \{\text{Td, Te}\}$, $(l_r, u_r) = (0, 12.0)$, $\forall r \in \{\text{R6, R7, Rres, Growth}\}$, $(l_r, u_r) = (0, 9999)$ and for Oxygen, $(l_r, u_r) = (0, 15.0)$.
The same simulation graphs are obtained using the local function $f_{\text{Rres}} = \neg x_{\text{RPO2}}$ and $f_{\text{Rres}} = 1$.

		External metabolites				Regulatory proteins		Reactions								
Experiment	Time	$\overline{z_{\text{Biomass}}}$	$\overline{z_{\text{Carbon1}}}$	$\overline{z_{\text{Carbon2}}}$	$\overline{z_{\text{Oxygen}}}$	$\overline{x_{\text{RPO2}}}$	$\overline{x_{\text{RPcl}}}$	v_{Tc1}	v_{Tc2}	v_{To2}	v_{Td}	v_{Te}	v_{Growth}	v_{Rres}	v_{R6}	v_{R7}
1	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0
	1	1	1	1	1	0	1	1	0	1	0	0	1	1	0	0
	51	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0
	52	1	0	1	1	0	0	0	1	1	0	0	1	1	0	0
	59	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0
	1	1	1	1	0	1	1	1	0	0	1	1	1	0	1	1
	83	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0
	84	1	0	1	0	1	0	0	1	0	1	1	1	0	1	1
	97	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
	1	1	0	1	0	1	0	0	1	0	1	1	1	0	1	1
	83	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
4	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0
	1	1	1	0	0	1	1	1	0	0	1	1	1	0	1	1
	83	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
	1	1	1	0	1	0	1	1	0	1	0	0	1	1	0	0
	51	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
	1	1	0	1	1	0	0	0	1	1	0	0	1	1	0	0
	51	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Table 4: All the different binarized metabolic steady states of each experiment. They are the input data used to solve the inference problem.

References

1. Baral, C.: Knowledge Representation, Reasoning and Declarative Problem Solving. Cambridge University Press, New York, NY, USA (2003)
2. Bernot, G., Comet, J.P., Richard, A., Guespin, J.: Application of formal methods to biological regulatory networks: extending thomas' asynchronous logical approach with temporal logic. Journal of Theoretical Biology **229**(3), 339–347 (2004). https://doi.org/10.1016/j.jtbi.2004.04.003
3. Buescher, J.M., Liebermeister, W., Jules, M., Uhr, M., Muntel, J., Botella, E., Hessling, B., Kleijn, R.J., Chat, L.L., Lecointe, F., Mader, U., Nicolas, P., Piersma, S., Rugheimer, F., Becher, D., Bessieres, P., Bidnenko, E., Denham, E.L., Dervyn, E., Devine, K.M., Doherty, G., Drulhe, S., Felicori, L., Fogg, M.J., Goelzer, A., Hansen, A., Harwood, C.R., Hecker, M., Hubner, S., Hultschig, C., Jarmer, H., Klipp, E., Leduc, A., Lewis, P., Molina, F., Noirot, P., Peres, S., Pigeonneau, N., Pohl, S., Rasmussen, S., Rinn, B., Schaffer, M., Schnidder, J., Schwikowski, B., Dijl, J.M.V., Veiga, P., Walsh, S., Wilkinson, A.J., Stelling, J., Aymerich, S., Sauer, U.: Global network reorganization during dynamic adaptations of bacillus subtilis metabolism. Science **335**(6072), 1099–1103 (2012). https://doi.org/10.1126/science.1206871
4. Chaves, M., Oyarzún, D.A., Gouzé, J.L.: Analysis of a genetic-metabolic oscillator with piecewise linear models. Journal of Theoretical Biology **462**, 259–269 (2019). https://doi.org/10.1016/j.jtbi.2018.10.026
5. Chaves, M., Tournier, L., Gouzé, J.L.: Comparing boolean and piecewise affine differential models for genetic networks. Acta Biotheor **58**(2-3), 217–232 (2010). https://doi.org/10.1007/s10441-010-9097-6
6. Chevalier, S., Froidevaux, C., Pauleve, L., Zinovyev, A.: Synthesis of boolean networks from biological dynamical constraints using answer-set programming. In: 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI). IEEE (2019). https://doi.org/10.1109/ictai.2019.00014

7. Covert, M.W., Knight, E.M., Reed, J.L., Herrgard, M.J., Palsson, B.O.: Integrating high-throughput and computational data elucidates bacterial networks. Nature **429**(6987), 92–96 (2004). https://doi.org/10.1038/nature02456
8. Covert, M.W., Palsson, B.Ø.: Transcriptional regulation in constraints-based metabolic models of Escherichia coli. J Biol Chem **277**(31), 28058–28064 (2002). https://doi.org/10.1046/j.1462-2920.2002.00282.x
9. Covert, M.W., h. Schilling, C., Palsson, B.: Regulation of gene expression in flux balance models of metabolism. Journal of Theoretical Biology **213**(1), 73–88 (2001). https://doi.org/10.1006/jtbi.2001.2405
10. Eiter, T., Gottlob, G.: On the computational cost of disjunctive logic programming: Propositional case. Annals of Mathematics and Artificial Intelligence **15**(3-4), 289–323 (sep 1995). https://doi.org/10.1007/bf01536399
11. Eiter, T., Ianni, G., Krennwallner, T.: Answer Set Programming: A Primer, pp. 40–110. Springer Berlin Heidelberg, Berlin, Heidelberg (2009). https://doi.org/10.1007/978-3-642-03754-2_2
12. Gebser, M., Kaminski, R., Kaufmann, B., Schaub, T.: Answer Set Solving in Practice. Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan and Claypool Publishers (2012)
13. Gebser, M., Kaminski, R., Kaufmann, B., Schaub, T.: Clingo = ASP + control: Preliminary report. CoRR **abs/1405.3694** (2014)
14. Gebser, M., Kaminski, R., Schaub, T.: Complex optimization in answer set programming. Theory and Practice of Logic Programming **11**(4-5), 821–839 (2011). https://doi.org/10.1017/s1471068411000329
15. Gebser, M., Kaufmann, B., Romero, J., Otero, R., Schaub, T., Wanko, P.: Domain-specific heuristics in answer set programming. Proceedings of the AAAI Conference on Artificial Intelligence **27**(1) (Jun 2013), `https://ojs.aaai.org/index.php/AAAI/article/view/8585`
16. de Jong, H.: Modeling and simulation of genetic regulatory systems: A literature review. Journal of Computational Biology **9**, 67–103 (2002). https://doi.org/10.1089/10665270252833208
17. Liu, L., Bockmayr, A.: Regulatory dynamic enzyme-cost flux balance analysis: A unifying framework for constraint-based modeling. Journal of Theoretical Biology **501**, 110317 (2020). https://doi.org/10.1016/j.jtbi.2020.110317
18. Marmiesse, L., Peyraud, R., Cottret, L.: FlexFlux: combining metabolic flux and regulatory network analyses. BMC Systems Biology **9**(1) (2015). https://doi.org/10.1186/s12918-015-0238-z
19. Orth, J.D., Thiele, I., Palsson, B.Ø.: What is flux balance analysis? Nat Biotechnol **28**(3), 245–248 (2010). https://doi.org/10.1038/nbt.1614
20. Ostrowski, M., Paulevé, L., Schaub, T., Siegel, A., Guziolowski, C.: Boolean network identification from perturbation time series data combining dynamics abstraction and logic programming. Biosystems **149**, 139–153 (2016). https://doi.org/10.1016/j.biosystems.2016.07.009
21. Oyarzún, D.A., Chaves, M., Hoff-Hoffmeyer-Zlotnik, M.: Multistability and oscillations in genetic control of metabolism. Journal of Theoretical Biology **295**, 139–153 (2012). https://doi.org/10.1016/j.jtbi.2011.11.017
22. Razzaq, M., Paulevé, L., Siegel, A., Saez-Rodriguez, J., Bourdon, J., Guziolowski, C.: Computational discovery of dynamic cell line specific boolean networks from multiplex time-course data. PLOS Computational Biology **14**(10), e1006538 (2018). https://doi.org/10.1371/journal.pcbi.1006538

23. Saez-Rodriguez, J., Alexopoulos, L.G., Epperlein, J., Samaga, R., Lauffenburger, D.A., Klamt, S., Sorger, P.K.: Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. Mol Syst Biol **5**(1), 331 (2009). https://doi.org/10.1038/msb.2009.87
24. Tournier, L., Goelzer, A., Fromion, V.: Optimal resource allocation enables mathematical exploration of microbial metabolic configurations. J. Math. Biol. **75**(6-7), 1349–1380 (2017). https://doi.org/10.1007/s00285-017-1118-5
25. Tsiantis, N., Balsa-Canto, E., Banga, J.R.: Optimality and identification of dynamic models in systems biology: an inverse optimal control framework. Bioinformatics **34**(14), 2433–2440 (2018). https://doi.org/10.1093/bioinformatics/bty139
26. Videla, S., Saez-Rodriguez, J., Guziolowski, C., Siegel, A.: caspo: a toolbox for automated reasoning on the response of logical signaling networks families. Bioinformatics p. btw738 (2017). https://doi.org/10.1093/bioinformatics/btw738
27. Zañudo, J.G.T., Yang, G., Albert, R.: Structure-based control of complex networks with nonlinear dynamics. Proc Natl Acad Sci USA **114**(28), 7234–7239 (2017). https://doi.org/10.1073/pnas.1617387114