



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

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

### ► To cite this version:

Kerian Thuillier, Caroline Baroukh, Alexander Bockmayr, Ludovic Cottret, Loïc Paulevé, et al..  
Learning Boolean controls in regulated metabolic networks: a case-study. 2021. hal-03207589v1

**HAL Id: hal-03207589**

**<https://hal.science/hal-03207589v1>**

Preprint submitted on 25 Apr 2021 (v1), last revised 2 Sep 2021 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Kerian Thuillier<sup>1</sup>, Caroline Baroukh<sup>2</sup>, Alexander Bockmayr<sup>3</sup>, Ludovic Cottret<sup>2</sup>, Loïc Paulevé<sup>4</sup>, and Anne Siegel<sup>1</sup>

<sup>1</sup> Univ Rennes, Inria, CNRS, IRISA, F-35000 Rennes, France

<sup>2</sup> LIPME, INRAE, CNRS, Université de Toulouse, Castanet-Tolosan, France

<sup>3</sup> Freie Universität Berlin, Institute of Mathematics, D-14195 Berlin, Germany

<sup>4</sup> 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 [14,2,5]. Similarly, the study of metabolism at steady-state has lead to various constraint-based approaches [18,15], 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) [18]. 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 [4,26,20,8]. 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 feed-back 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 [15] for an overview), one of them being FlexFlux [17], which implements the rFBA framework [8].

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 [23], where RBA is used to deduce regulations according to perturbations of the environment. However, to induce regulations, they 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* [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 [22,19,25,21,24]. In this direction, the caspoTS and the BoNesis approaches [21,19,25,6] 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.

*Notations* The cardinality of a finite set  $X$  is denoted by  $|X|$ . Given a set  $D$ , the concatenation of two vectors  $x \in D^k$ ,  $y \in D^m$  is denoted by  $(x, y)$  resulting in a vector of dimension  $k + m$  equals to  $(x_1, \dots, x_k, y_1, \dots, y_m)$ . Given a vector  $x \in D^n$  and a set of indices  $I \subseteq \{1, \dots, n\}$ ,  $x_I$  denotes the vector of dimension  $|I|$  equals to  $(x_i)_{i \in I}$ . The Boolean domain is denoted by  $\mathbb{B} = \{0, 1\}$ . Given two Boolean vectors  $x, y \in \mathbb{B}^n$  of dimension  $n$ ,  $x \preceq y \iff \forall i \in \{1, \dots, n\}, x_i \leq y_i$ .

## 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 nodes and fluxes of the

metabolic layer, associated with linear equations. Feedbacks are ensured by components of the metabolic networks which are themselves involved in the Boolean functions associated with the regulatory layer.

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

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

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  $S_{\text{Int}, \mathcal{R}} \cdot v = 0$  and  $l_r \leq v_r \leq u_r$ , for all  $r \in \mathcal{R}$ . Here  $S_{\text{Int}, \mathcal{R}}$  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. The external metabolite “biomass” does not have a true existence but it is only used to mimic the production of all the precursors needed to the cellular growth.

**Definition 2 (Input and output metabolites).** For an external metabolite  $m \in \text{Ext}$ , 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  $m \in \text{Ext}$  is called an input (resp. output) metabolite if there exists a reaction  $r \in \mathcal{R}$  with  $S_{mr} < 0$  (resp.  $S_{mr} > 0$ ). The set of all input metabolites is denoted by  $\text{Inp} \subseteq \text{Ext}$ .

A regulatory network is a set of biological entities (e.g. genes, reactions, metabolites) or even abiotic entities (e.g. temperature, pH) which 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  $f : \mathbb{B}^n \rightarrow \mathbb{B}^n$ . For each  $i \in \{1, \dots, n\}$ ,  $f_i : \mathbb{B}^n \rightarrow \mathbb{B}$  is the  $i$ -th component of this function called the local function of  $i$ .

Its influence graph  $G(f)$  is a signed digraph  $(V, E)$  with  $V = \{1, \dots, n\}$  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  $s \cdot f_j(x) < s \cdot f_j(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_n)$ . In the following we will slightly abuse notations by identifying  $G(f)$  with its set of edges, 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:  $(i, -s, j) \notin G(f)$ .

We consider 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 null flux) whenever certain conditions on the activity of input metabolites and regulatory proteins are met. Moreover, we consider that the activity of regulatory proteins is mediated by the metabolic network only. The obtained model is then assumed 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 for the fluxes of its reactions; 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  $\mathcal{N} = (\text{Int}, \text{Ext}, \mathcal{R}, S)$  with  $k$  input metabolites  $\text{Inp} = \{e_1, \dots, e_k\} \subseteq \text{Ext}$  and  $m$  reactions  $\mathcal{R} = \{r_1, \dots, r_m\}$ ;
- a set of  $d$  regulatory proteins  $\mathcal{P} = \{p_1, \dots, p_d\}$
- a BN  $f$  of dimension  $n = |\text{Inp}| + |\mathcal{R}| + |\mathcal{P}|$  where  $\{1, \dots, n\} = \text{Inp} \cup \mathcal{R} \cup \mathcal{P}$  such that  $G(f)$  is a bipartite graph between  $\mathcal{P}$  and  $\text{Inp} \cup \mathcal{R}$ .

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

The BN  $f$  models the regulation of the fluxes of the metabolic network  $\mathcal{N}$ . This regulation is always in one direction: either a flux is free whenever the reaction has value 1 in the BN, or it is blocked (forced to null) whenever the reaction has value 0 in the BN. Following this convention, a reaction  $r \in \mathcal{R}$  is never regulated whenever  $f_r(x) = 1$ . As we will defined formally in the next section, these regulation impact the steady states of the metabolic network.

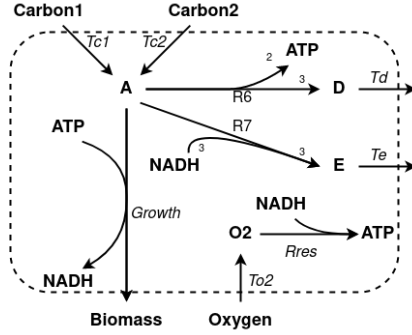
Finally, given a real-valued vector  $s \in \mathbb{R}_{\geq 0}^n$ , we write  $\beta(s) \in \mathbb{B}^n$  its *binarization*:

$$\forall i \in \{1, \dots, n\}, \quad \beta(s)_i = \begin{cases} 1 & \text{if } s_i > 0 \\ 0 & \text{if } s_i = 0 \end{cases}.$$

An example of a regulated metabolic network is shown in Fig. 1. This example is based on the simplified core carbon metabolic network, modelling diauxic shift of carbon, proposed in [8].

At the metabolic level (Fig. 1a), there are 4 external metabolites and  $m = 9$  reactions. Internal metabolites are  $\text{Int} = \{\text{A}, \text{D}, \text{E}, \text{O}_2, \text{ATP}, \text{NADH}\}$  and external metabolites are  $\text{Ext} = \{\text{Carbon1}, \text{Carbon2}, \text{Oxygen}, \text{Biomass}\}$ . Among them, the  $k = 3$  input metabolites are  $\text{Inp} = \{\text{Carbon1}, \text{Carbon2}, \text{Oxygen}\}$ . The set of irreversible reactions is  $\mathcal{R} = \{\text{Tc1}, \text{Tc2}, \text{To2}, \text{Td}, \text{Te}, \text{Growth}, \text{Rres}, \text{R6}, \text{R7}\}$ . The stoichiometric coefficients are shown in the Figure. By default, they are set to 1, except for the reactions  $\text{R6}$  and  $\text{R7}$ .

The regulatory level (Fig. 1c) of the regulated metabolism introduces  $d = 2$  regulatory proteins:  $\mathcal{P} = \{\text{RPcl}, \text{RPO2}\}$ . Thus, the Boolean network  $f$  is of dimension  $n = k + m + d = 14$ . It consists of 15 functions (see Fig. 1c) which map a Boolean vector  $x = (x_{\text{Carbon1}}, x_{\text{Carbon2}}, x_{\text{Oxygen}}, x_{\text{RPcl}}, x_{\text{RPO2}}, x_{\text{Tc1}}, x_{\text{Tc2}}, x_{\text{To2}}, x_{\text{Td}}, x_{\text{Te}}, x_{\text{Rres}}, x_{\text{R6}}, x_{\text{R7}}) \in \mathbb{B}^n$  to a Boolean value in  $\mathbb{B}$ . The functions associated with regulatory proteins in  $\mathcal{P}$  involve only metabolite variables. Among the nine functions associated with reactions, only two ( $\text{Tc2}$ ,  $\text{Rres}$ ) are non-constant: they involve the two regulatory proteins. The influence graph of the network is shown in Fig. 1b. Only the shown nodes have a non-constant function ( $\text{RPcl}$ ,

(a) *Metabolic Network*(b) 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	$f_{RPO2}(x)$	$f_{RPl}(x)$	$f_{Carbon1}(x)$	$f_{Carbon2}(x)$	$f_{Oxygen}(x)$
Boolean rule	$\neg x_{Oxygen}$	$x_{Carbon1}$	0	0	0

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

(c) *Boolean Network*

Fig. 1: Example of regulated metabolic network with its metabolic network (a) and its Boolean network (c). In (a), each node is a metabolite and each hyperedge is a reaction. For instance, the hyperedge R7 linking  $\{A; NADH\}$  to  $\{E\}$  models the reaction  $A + 3 NADH \rightarrow 3 E$ . Integer values over hyperedges are stoichiometric coefficients, default value is 1. (c) shows the Boolean network regulating the metabolic network (a) where  $x \in \mathbb{B}^n$  and  $n = 14$ . All Boolean functions set to 1 correspond to reactions which are not regulated by the Boolean network. (b) shows the influence (or regulatory) graph of the Boolean network described in (c), green and red edges are positive and negative edges, respectively. Square nodes are regulatory proteins.

RPO2, Tc2, Rres) or are used in the local function of another node (Carbon1, Oxygen). The influence graph 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) [18] returns an *optimal* metabolic steady state, according to a given objective function over reaction fluxes. In the following we assume the objective function is to maximize the flux of a reaction *Growth*. For regulated metabolic networks, the rFBA framework [8] 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 triplet  $(v, w, x) \in \mathbb{R}^{|\mathcal{R}|} \times \mathbb{R}^{|\text{Ext}|} \times \mathbb{B}^{|\text{Inp}|+|\mathcal{R}|+|\mathcal{P}|}$  such that

- $S_{\text{Int}, \mathcal{R}} \cdot v = 0$ ,
- 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  $m \in \text{Inp}$  and each reaction  $r \in \mathcal{R}$  with  $S_{mr} < 0$ ,  $v_r \leq \text{uptake\_bound}(w_m)$ , where  $\text{uptake\_bound}(w_m)$  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 input metabolite concentrations  $w^{k+1}$  are obtained from the previous concentrations  $w^k$  by assuming constant uptake fluxes  $v^k$  in the time period  $[t^k, t^{k+1}]$ .
2. Let the Boolean state  $x' \in \mathbb{B}^{|\text{Inp}|+|\mathcal{R}|+|\mathcal{P}|}$  be defined by the binarized input metabolites concentrations  $x'_{\text{Inp}} = \beta(w_{\text{Inp}}^{k+1})$  at  $t^{k+1}$ , the binarized reaction fluxes  $x'_{\mathcal{R}} = \beta(v^k)$ , and the Boolean values  $x'_{\mathcal{P}} = x_{\mathcal{P}}^k$  of the regulatory proteins at  $t^k$ . Then,

$$x^{k+1} = f(x')$$

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  $v'_{\text{Growth}} > v_{\text{Growth}}^{k+1}$ .

In this paper, we rely on the FLEXFLUX implementation of rFBA [17], which assumes a fixed time step  $\tau$  between successive metabolic-regulatory steady states ( $t^{k+1} - t^k = \tau$  for any  $k$ ). The *Growth* reaction is assumed to reflect the growth the cell. FLEXFLUX computes the evolution of the total biomass of the cell as  $\text{biomass}^{k+1} = \text{biomass}^k \cdot e^{v_{\text{Growth}}^k \cdot \tau}$  (from a given initial biomass<sup>0</sup>). The maximum uptake fluxes of input metabolites  $m \in \text{Inp}$  are defined as  $\text{uptake\_bound}(w_m) =$

Time	External metabolites				Regulatory proteins		Reactions								
	$w_{\text{Biomass}}$	$w_{\text{Carbon1}}$	$w_{\text{Carbon2}}$	$w_{\text{Oxygen}}$	$x_{\text{RPO2}}$	$x_{\text{RPcl}}$	$v_{\text{Tc1}}$	$v_{\text{Tc2}}$	$v_{\text{Tc2}}$	$v_{\text{Td}}$	$v_{\text{Te}}$	$v_{\text{Growth}}$	$v_{\text{Rres}}$	$v_{\text{R6}}$	$v_{\text{R7}}$
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 shown Fig. 2. It contains the binarized values of the metabolic steady state computed by the r-FBA 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  $t - 1$ .

$w_m/(\text{biomass} \cdot \tau)$ . Finally, the update of the external metabolite concentrations is computed as follows:

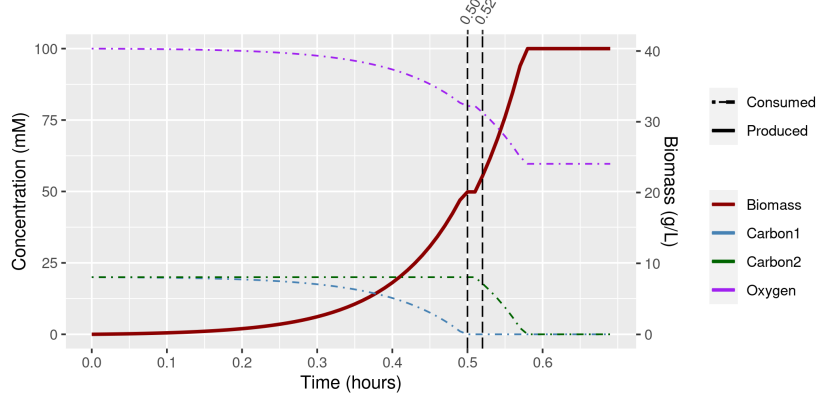
$$w_m^{k+1} = w_m^k - S_{mr} v_r^k / (v_{\text{Growth}}^k \cdot (\text{biomass}^k - \text{biomass}^{k+1}))$$

where  $r \in \mathcal{R}$  is the uptake/export 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 introduced Fig. 1 is shown in Fig. 2. It uses a time step of 0.01 hour and is initialized with 100 mM of Oxygen, 20 mM of Carbon1 and Carbon2. The simulation shown in Fig. 2a was composed of 70 metabolic steady states. By applying the binarization  $\beta$  introduced in the previous section, these 70 metabolic steady states correspond to 5 different binarized metabolic steady states. They are shown in Table. 1. These binarized metabolic steady states capture the main features of the simulation.

More precisely, the simulation shows that until 0.5 hour only Carbon1 and Oxygen are consumed to produce Biomass. This corresponds to a first time period where the behavior of the system is monotonous: 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 0.5 hour, 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 corresponding to equal binarization of metabolic steady-states). At 0.51, as shown 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 0.51. This corresponds to a third qualitative behavior of the system. At 0.52, the Boolean state  $x'$  is updated to  $x''$ : all the node states remain unchanged except for  $x''_{\text{Tc2}} = f_{\text{Tc2}}(x') = \neg x'_{\text{RPcl}} = 1$ . This corresponds to a fourth qualitative behavior of the system. The reaction Tc2 is not inhibited anymore, and the Biomass is produced thanks to the inputs of Carbon2 and Oxygen (thanks to Tc2, Growth and Rres until Carbon2 depletion at  $t = 0.59$  (fifth qualitative behavior of the system).





(a) Simulation graph showing the evolution of the quantity of external metabolites (Oxygen, Carbon1, Carbon2) and the production of Biomass.

Time	External metabolites				Regulatory proteins		Reaction flows								
	$w_{\text{Biomass}}$	$w_{\text{Carbon1}}$	$w_{\text{Carbon2}}$	$w_{\text{Oxygen}}$	$x_{\text{RPO2}}$	$x_{\text{RPcl}}$	$v_{\text{Tc1}}$	$v_{\text{Tc2}}$	$v_{\text{Tc3}}$	$v_{\text{Td}}$	$v_{\text{Te}}$	$v_{\text{Growth}}$	$v_{\text{Rres}}$	$v_{\text{R6}}$	$v_{\text{R7}}$
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 times 0.49 to 0.53 hour of the simulation. These times correspond to a set of regulatory updates allowing switching from Carbon1 to Carbon2 for Biomass production.

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

### 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 of the system is similar, meaning that the variation of all the metabolic variables is monotonous and the Boolean values of the regulatory proteins are constant. In this section, we introduce a discrete definition of steady-states which allows to captures the monotonous behaviors observed in rFBA simulations. This allows to introduce a discretized rFBA framework, 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  $m$

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

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

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

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

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

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

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  $\mathcal{N} = (\text{Int}, \text{Ext}, \mathcal{R}, S)$  is a Boolean vector  $\hat{v} \in \mathbb{B}^{|\text{Ext}| + |\mathcal{R}|}$  which is a satisfying assignment of the following logical steady-state formula:

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

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

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

Remark that the converse property is not true: because the logical characterization omit stoichiometry, Boolean metabolic steady states may have no real-valued counterpart.

Applied to the example, the internal metabolic constraints are the following:

$$\begin{aligned} \overline{z_A}^+ &= \overline{v_{Tc1}} \vee \overline{v_{Tc2}}, & \overline{z_A}^- &= \overline{v_{R6}} \vee \overline{v_{R7}} \vee \overline{v_{Growth}} \\ \overline{z_D}^+ &= \overline{v_{R6}}, & \overline{z_D}^- &= \overline{v_{Td}}, & \overline{z_E}^+ &= \overline{v_{R7}}, & \overline{z_E}^- &= \overline{v_{Te}} \\ \overline{z_{O2}}^+ &= \overline{v_{To2}}, & \overline{z_{O2}}^- &= \overline{v_{Rres}} \\ \overline{z_{ATP}}^+ &= \overline{v_{R6}} \vee \overline{v_{Rres}}, & \overline{z_{ATP}}^- &= \overline{v_{Growth}} \\ \overline{z_{NADH}}^+ &= \overline{v_{Growth}}, & \overline{z_{NADH}}^- &= \overline{v_{R7}} \vee \overline{v_{Rres}} \end{aligned}$$

The logical steady-state constraints equivalent to  $\hat{\mathcal{N}} = 1$  are obtained by gathering constraints on internal and external metabolites:

$$\begin{aligned} \overline{v_{Tc1}} \vee \overline{v_{Tc2}} &= \overline{v_{R6}} \vee \overline{v_{R7}} \vee \overline{v_{Growth}} \\ \overline{v_{R6}} &= \overline{v_{Td}} & \overline{v_{R7}} &= \overline{v_{Te}} & \overline{v_{To2}} &= \overline{v_{Rres}} \\ \overline{v_{R6}} \vee \overline{v_{Rres}} &= \overline{v_{Growth}} & \overline{v_{R7}} \vee \overline{v_{Rres}} &= \overline{v_{Growth}} \\ \overline{v_{Tc1}} &\Rightarrow \overline{z_{Carbon1}} & \overline{v_{Tc2}} &\Rightarrow \overline{z_{Carbon2}} & \overline{v_{To2}} &\Rightarrow \overline{z_{Oxygen}} \end{aligned}$$

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 regulated-metabolic steady-states (Table 1) appearing in the r-FBA simulations of Fig.2.

### 3.2 Boolean dynamics

From 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: it is 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: with  $\mathcal{M} = \text{Ext} \cup \mathcal{R}$  referring to external metabolites and fluxes of 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 along their biomass production, as it 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 a natural numbers:  $\hat{o} : \mathbb{B}^{k+m} \rightarrow \mathbb{N}$ . The Boolean dynamics will select among Boolean metabolic steady states maximizing this supplied objective only.

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

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

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

1. the value of 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:  $\forall y'_{\mathcal{M}} \in Z(x'), \hat{o}(y_{\mathcal{M}}) \geq \hat{o}(y'_{\mathcal{M}})$ ,

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

## 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 in different conditions, our objective is to infer all the regulatory Boolean networks that can reproduce the behaviors. Besides the ground-truth model, the inference may suggest alternative regulatory logic.

The domain of search for BNs, we denote by  $\mathbb{F}$ , is delimited by an influence graph  $\mathcal{G}$ : any candidate  $f \in \mathbb{F}$  should satisfy that  $G(f) \subseteq \mathcal{G}$ , 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 augmented from the ground-truth regulatory model  $f^\circ$  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.

However, such an 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. We rely on the Boolean abstraction of dynamic rFBA presented in the previous section to express the inference problem.

### 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.

The inputs of the relaxed inference problem are:

- a metabolic network  $\mathcal{N}$  and a set of regulatory proteins  $\mathcal{P}$ ,
- sequences of metabolic-regulatory steady states, represented by sets of pairs  $(s^t, s^{t+1})$ , with  $s^t = (v^t, w^t, x^t)$   $s^{t+1} = (v^{t+1}, w^{t+1}, x^{t+1})$  following notations from Def. 5: the observed changed of metabolic-regulatory steady states are given as  $T \subseteq \mathbb{S} \times \mathbb{S}$  with  $\mathbb{S} = \mathbb{R}^{|\text{Inp}|+|\mathcal{R}|} \times \mathbb{B}^{|\mathcal{RPs}|}$ ,

- a domain of putative regulatory BNs  $\mathbb{F}$  of dimension  $n = |\text{Inp}| + |\mathcal{R}| + |\mathcal{P}|$ ,
- a Boolean state objective score  $\hat{o} : \mathbb{B}^n \rightarrow \mathbb{N}$ .

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

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

Thanks to the formulation in the Boolean dynamic rFBA framework, the inference problem boils down to a satisfiability problem in propositional Boolean logic using two levels of quantifiers (2-QBF), expressed as follows.

Given  $(s, s') \in T$ , with  $x' \in \mathbb{B}^n$  such that  $x'_{\text{Inp}} = \beta(s')_{\text{Inp}}$  and  $x'_{\mathcal{R} \cup \mathcal{P}} = \beta(s)_{\mathcal{R} \cup \mathcal{P}}$ ,

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

The overall problem is then the conjunction over all pairs  $(s, s') \in T$ .

Remark that with no Boolean optimization criteria (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, for each  $r \in \mathcal{R}$ , have  $f_r(x) = 1$  always ensures that  $y_{\mathcal{R}} \preceq f_{\mathcal{R}}(x')$ .

## 4.2 Implementation in Answer-Set Programming

Answer-Set Programming (ASP) [1,11] is a declarative framework allowing solving combinatorial satisfaction problems. It relies on the stable model semantics, which enables expressing 2-QBF problems using disjunctive programming ( $\Sigma_2^P$ -complete) [9].

An ASP program is a logic program modelled with a set of logical rules. The logical rules contain first order logic predicates. Each rule has the following form:

$$a_0 \leftarrow l_1, \dots, l_n$$

where  $a_0$  is a variable free atom and each  $\{l\}_{\{1, n\}}$  is a literal which is either an atom  $a$  or its negation  $\text{not } a$ . This rule states that when all the literals  $\{l\}_{\{1, n\}}$  are true then  $a_0$  is true too. If  $a_0$  is  $\perp$  (*false*), then the rule is an *integrity constraint*:

$$\leftarrow l_1, \dots, l_n$$

Integrity constraints are satisfied only if the conjunction of all the right-hand side literals is false. In other words, they are satisfied if and only if at least one literal is  $\{l\}_{\{1, n\}}$  is false. In contrast, a rule of the form  $a_0 \leftarrow$  states that  $a_0$  is true.

ASP allows to define *disjunctive logic program* by adding disjunctive declaration in rule left-hand side. To declare a disjunction, each left-hand side atom of a rule must be separated by a semicolon (;).

$$a_0; \dots; a_m \leftarrow l_1, \dots, l_n$$

Disjunction rules are subset-minimal. In other words, an answer set, containing a set of disjunctive variables  $\mathbb{S}$ , is a solution if and only if there is not any solution answer set of which the set of disjunctive variables is a subset of  $\mathbb{S}$ . For instance, let's consider the following example:

$$\begin{aligned} a &\leftarrow \\ b &\leftarrow a \\ a; b; c &\leftarrow \end{aligned}$$

where  $a$ ,  $b$  and  $c$  are atoms. There are two solution answer sets:  $\{a, b\}$  and  $\{a, b, c\}$ . However, the atoms  $a$ ,  $b$  and  $c$  are declared with a disjunction rule. As the solution  $\{a, b, c\}$  is not subset minimal (it contains  $\{a, b\}$ ), it is discarded. The only subset minimal solution of this logic program is  $\{a, b\}$ .

Disjunction rules allow to model two quantification levels Boolean formulas (2-QBF) with ASP, *i.e.*  $\exists x, \forall y, \phi(x, y)$  where  $\phi(x, y)$  is a quantifier-free propositional formula. The encoding relies on the so-called *saturation technique* [10,13]. Essentially, for a fixed  $x$  and  $y$ , the encoding ensures 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 the smallest.

Our implementation of the Boolean dynamic rFBA inference problem is available at <https://github.com/bioasp/boolean-caspo-flux>. We rely on the solver CLINGO [12], which enables an efficient enumeration of compatible BNs  $f$ . Moreover, it gives the possibility to project the solution on each node of the BN, *i.e.*, enumerating only the local functions of a given node. This is particularly handy to capture a large set of solutions: by enumerating the solutions node per node, we can approximate the solution space by their cartesian product.

## 5 Case study

The approach was applied on an example inspired by the regulated metabolic network shown Fig. 1. For this regulated metabolic network, six experiments were designed in order to mimic the simulations introduced in [?]. Each experiment is based on a different set  $A \subseteq E = \{\text{Carbon1}, \text{Carbon2}, \text{Oxygen}\}$  of initially available input metabolites. The initialisation of each experiment is detailed Table. 2. For instance, experiment 1 is initialised by assuming that all input metabolites ( $\{\text{Carbon1}, \text{Carbon2}, \text{Oxygen}\}$ ) are available. Experiment 2 assumes that  $\{\text{Carbon1}, \text{Carbon2}\}$  are present at initialization but Oxygen is not.

The search space of regulated metabolic networks on which the inference problem was applied consisted in all regulated metabolic networks which have the same metabolic networks and regulated proteins as the example of Fig. 1 and which Boolean rules had the same the influence graph as the the example of Fig. 1. This corresponds to a search space of 1,944,320 BNs.

For each experiment, we used the rFBA implementation FLEXFLUX with an initial *biomass* value of 0.1, a time step of 0.01 to simulate the system behavior. Each of the 6 simulations is composed of 200 metabolic steady states. For a given experiment initialised with the external metabolite values  $(\bar{z}_{\text{Carbon1}}, \bar{z}_{\text{Carbon2}}, \bar{z}_{\text{Oxygen}})$ , regulatory proteins were initialised such that  $x_{\text{RPcl}} = \bar{z}_{\text{Carbon1}}$  and  $x_{\text{RPO2}} = \neg \bar{z}_{\text{Oxygen}}$ . The initial values of regulatory proteins for each experiment are shown Table. 2. Each simulation  $S = \{(v, w, x)_0, \dots, (v, w, x)_{200}\}$  contained 201 continuous metabolic-regulatory steady states (1 for the initialisation and 200 for the simulation). Simulation graphs of the 6 experiments are shown Fig. 3 in appendix.

The six simulations were processed to be used by the model-inference algorithm. They were binarized:  $S^{\mathbb{B}} = \{(\bar{v}_t, \bar{z}_t) = \beta((v_t, w_t)) \mid \forall v_t \in S\}$ . Then, these binarized metabolic-regulatory steady states were filtered to remove redundant time step. Thus, only the binarized metabolic-regulatory steady states different from the previous time step were conserved. In other words, for each binarized simulation  $S^{\mathbb{B}} = \{(v^0, z^0), \dots, (v^{200}, z^{200})\}$ ,  $\forall t \in [0; 200]$ ,  $t = 0 \vee v^t \neq v^{t-1}$  are conserved. Table. 1 shows the binarized metabolic-regulatory steady states representing the simulation of the first experiment. From the 201 initial continuous metabolic steady states, we conserved 5 binarized metabolic-regulatory steady states corresponding to the time steps  $\{0, 1, 51, 52, 59\}$ . The set of all the remaining time steps for each binarized simulation is shown in appendix Table. 5.

To solve the inference problem, one must supply an objective function  $\hat{o}$ . Given  $\text{Inp} = \{\text{Carbon1}, \text{Carbon2}, \text{Oxygen}\}$  the set of external input metabolites, the objective function was defined as  $\forall x \in \text{MSS}^{\mathbb{B}}(\mathcal{N})$ ,  $\hat{o}(x) = \sum_{e \in \text{Inp}} x_e$ . This function was motivated by the fact the maximization of the biomass production often corresponds to the maximisation of inputs according to the QSS

Experiment	Input Metabolite			Regulatory Protein	
	$\bar{z}_{\text{Carbon1}}$	$\bar{z}_{\text{Carbon2}}$	$\bar{z}_{\text{Oxygen}}$	$x_{\text{RPcl}}$	$x_{\text{RPO2}}$
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

Table 2: Initialisation states of the six rFBA simulations used to create the dataset used as input of application of the reverse-engineering method to a case-study.

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

Table 3: Inferred models having subset minimal local functions. Not shown local functions ( $f_{\text{Carbon1}}(x)$ ,  $f_{\text{Carbon2}}(x)$ ,  $f_{\text{Oxygen}}(x)$ ,  $f_{\text{To2}}(x)$ ,  $f_{\text{Td}}(x)$ ,  $f_{\text{Te}}(x)$ ,  $f_{\text{Growth}}(x)$ ,  $f_{\text{R6}}(x)$ ,  $f_{\text{R7}}(x)$ ) are set to 1.

constraints. Therefore, if an available external input metabolite is not used in the observed Boolean metabolic network, then it must be explained by at least one regulation. This objective function allows to capture more refined behaviors at the discrete level than a usual biomass optimization function which may be too rough when we consider discretized values.

Applying the satisfaction constraints describe previously allows inferring 40 models. All these models share have 3 fixed local functions ( $f_{\text{RPO2}}(x)$ ,  $f_{\text{RPcl}}(x)$ ,  $f_{\text{Tc2}}(x)$ ), 9 local functions set to 1 ( $f_{\text{Carbon1}}(x)$ ,  $f_{\text{Carbon2}}(x)$ ,  $f_{\text{Oxygen}}(x)$ ,  $f_{\text{To2}}(x)$ ,  $f_{\text{Td}}(x)$ ,  $f_{\text{Te}}(x)$ ,  $f_{\text{Growth}}(x)$ ,  $f_{\text{R6}}(x)$ ,  $f_{\text{R7}}(x)$ ) and 2 functions which can be both set to 1 or fixed. The 4 smallest inferred models are described in Table. 3. Each local functions  $f_i$  of these 4 models are contained in the local function  $f_i$  of the 36 others models. Notice that the ground truth, *i.e.* the model used to generate the input data, is correctly inferred (model 2).

Adding the subset minimal constraint infers the model 1. The two unfixed functions ( $f_{\text{Rres}}(x)$ ,  $f_{\text{Tc1}}(x)$ ) are set to 1 due to the subset minimal constraint. The inferred model is thus  $f_{\text{RPO2}}(x) = \neg x_{\text{Oxygen}}$ ,  $f_{\text{RPcl}}(x) = x_{\text{Carbon1}}$ ,  $f_{\text{Tc2}}(x) = \neg x_{\text{RPcl}}$  and all the others local functions are set to 1. Notice that only  $f_{\text{Rres}}(x)$  differs between the inferred subset minimal model and the ground truth model.

In order to check whether the regulated metabolism model inferred by our approach (model 1 in Table. 3) could be considered as an alternative to the case-study described in Fig. 1, we performed a r-FBA simulation of the inferred regulated metabolism model (*i.e.* without the regulation on the reaction *Rres*) according to the six experimental conditions described in Table 2. We observed that these quantitative time-series simulations were strictly identical to the simulations of the toy example used to generate the dataset as shown Fig. ???. This suggests that the regulation on *Rres* is not necessary to explain the dataset. The proposed model allows to fully infer all the needed regulations and can be considered as the simplest regulated metabolic model matching with the experimental conditions of Table 2. Covert himself recognized in his article that, unlike others regulations, *Rres* “regulation is not necessary for the solution” [8]. The regulation is biologically present only to ensure that unnecessary enzymes decay. However, since enzymes concentrations are not represented explicitly in r-FBA methodology, the dataset does not reflect this biologic behavior, making it impossible to infer properly the regulation. Taking into account enzymes, with methods such as r-deFBA [16], should allow to prevent this issue. Nevertheless, the inference approach will also need to be adapted to work with these kind of advance metabolic modeling approaches.



## 6 Discussion

We propose here a framework to infer Boolean rules describing the regulation of a metabolic network. The mathematical formulation of the dynamic rFBA in the form of successions of steady states of the regulated metabolic network has allowed us to develop an algorithm to infer Boolean rules from time series in multiple conditions. A proof of concept was performed on the simulation of the diauxic shift on a small scale model.

Our method relies on a Boolean abstraction of the dynamic rFBA framework. This enables a formulation of the inference problem as a pure Boolean satisfiability problem using 2 levels of quantifiers, which can be efficiently solved using logic programming, such as Answer-Set Programming. One important parameter of the abstraction is the Boolean objective function which aims at identifying Boolean metabolic steady states which match with the optimal real-valued ones. This function is currently specified manually, from expertise on the network. Future work may explore how to derive it automatically. An alternative direction is to solve directly the inference problem mixing linear programming and Boolean constraints.

Several perspectives are available after this first approach. First, here all regulations are considered as synchronous, what is obviously not the case *in vivo* where regulations are activated in a stochastic way. Second, the activation and degradation times of the regulatory and enzymatic proteins are not taken into account. Moreover, the regulations are considered here as binary but we already know that the metabolism proceeds by finer regulations than what we propose here, for instance as captured by the regulatory dynamic enzyme-cost FBA [16].

## 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., Lecoq, 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., Dijn, 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., 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>
8. Covert, M.W., 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>
9. 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>
10. 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](https://doi.org/10.1007/978-3-642-03754-2_2)
11. 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)
12. Gebser, M., Kaminski, R., Kaufmann, B., Schaub, T.: Clingo = ASP + control: Preliminary report. *CoRR* **abs/1405.3694** (2014)
13. 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>
14. 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>
15. Liu, L., Bockmayr, A.: Regulatory dynamic enzyme-cost flux balance analysis: A unifying framework for constraint-based modeling. *J Theor Biol* **501**, 110317 (2020). <https://doi.org/10.1016/j.jtbi.2020.110317>
16. 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>
17. 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>
18. 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>
19. 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>
20. 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>

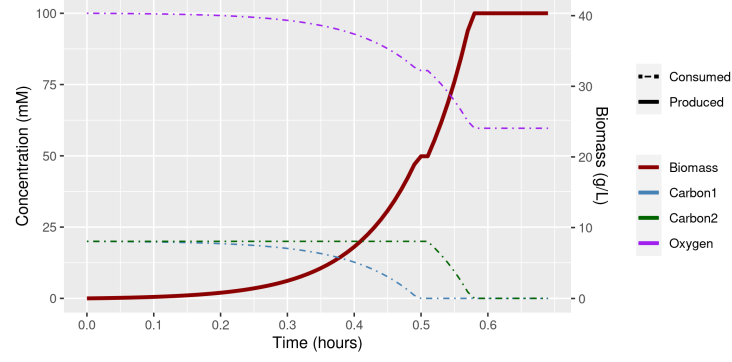
21. Razzaq, M., Paulevé, L., Siegel, A., Saez-Rodriguez, J., Bourdon, J., Guziolowski, C.: Computational discovery of dynamic cell line specific boolean networks from multiplex time-course data. *PLOS Computational Biology* **14**(10), e1006538 (2018). <https://doi.org/10.1371/journal.pcbi.1006538>
22. 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>
23. 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>
24. 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>
25. 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>
26. 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>

## A Binarized metabolic steady state

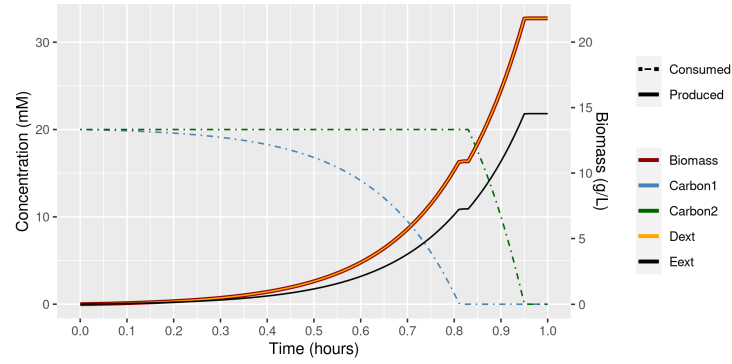
	External metabolites			Reactions									Experimentation
	$z_{\text{Carbon1}}$	$z_{\text{Carbon2}}$	$z_{\text{Oxygen}}$	$v_{\text{Tc1}}$	$v_{\text{Tc2}}$	$v_{\text{To2}}$	$v_{\text{Td}}$	$v_{\text{Te}}$	$v_{\text{Growth}}$	$v_{\text{Rres}}$	$v_{\text{R6}}$	$v_{\text{R7}}$	
1	0	0	0	0	0	0	0	0	0	0	0	0	2, 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	1	0	
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 4: 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.

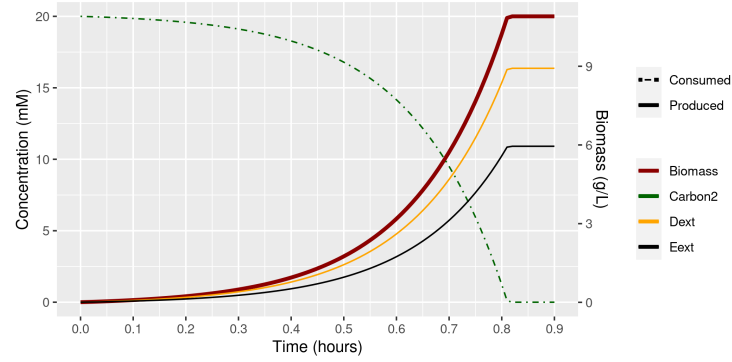
## B Experiments and simulations



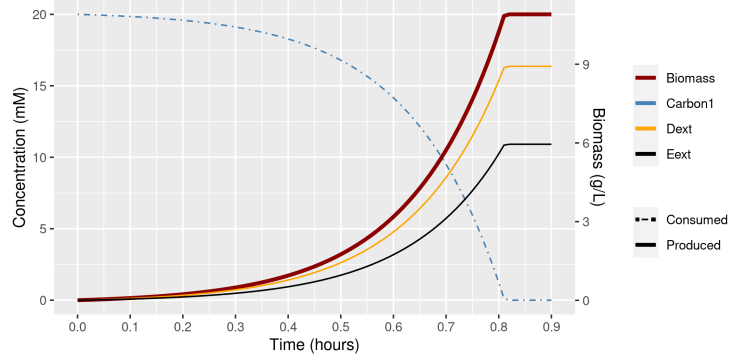
(a) Simulation of experiment 1.



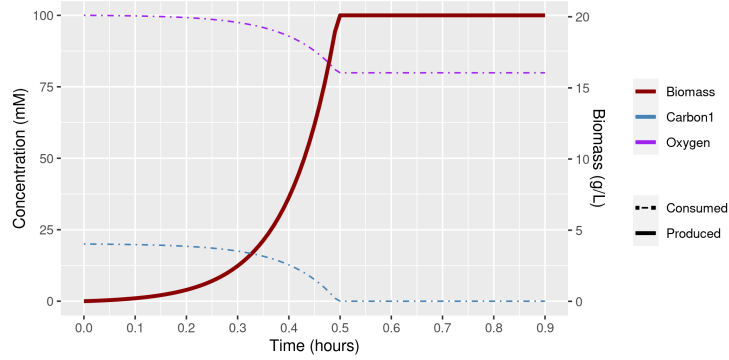
(b) Simulation of experiment 2.



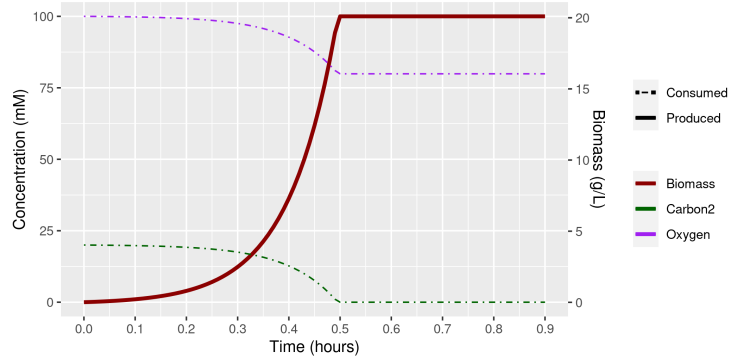
(c) Simulation of experiment 3.



(d) Simulation of experiment 4.



(e) Simulation of experiment 5.



(f) Simulation of experiment 6.

Fig. 3: Simulation made with FlexFlux of the regulated metabolic network shown Fig. 1 for each experiment (Table. 2). Time step is set to 0.01. Reaction domains are  $\forall r \in \{Tc1, Tc2\}, (l_r, u_r) = (0, 10.5)$ ,  $\forall r \in \{Td, Te\}, (l_r, u_r) = (0, 12.0)$ ,  $\forall r \in \{R6, R7, Rres, Growth\}, (l_r, u_r) = (0, 9999)$  and for Oxygen,  $(l_r, u_r) = (0, 15.0)$ .

The same simulation graphs are obtained using the local function  $f_{Rres} = \neg x_{RPO2}$  and  $f_{Rres} = 1$ .

		External metabolites				Regulatory proteins		Reactions								
Experiment	Time	$\bar{z}_{Biomass}$	$\bar{z}_{Carbon1}$	$\bar{z}_{Carbon2}$	$\bar{z}_{Oxygen}$	$\bar{x}_{RPO2}$	$\bar{x}_{RPF1}$	$\bar{v}_{Tc1}$	$\bar{v}_{Tc2}$	$\bar{v}_{Tc3}$	$\bar{v}_{Td}$	$\bar{v}_{Tt}$	$\bar{v}_{Growth}$	$\bar{v}_{Rres}$	$\bar{v}_{R6}$	$\bar{v}_{R7}$
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 5: All the different binarized metabolic steady states of each experiment. They are the input data used to solve the inference problem.