



Greening R. Thomas' Framework with Environment Variables: A Divide and Conquer Approach

Laetitia Gibart^(✉), H el ene Collavizza, and Jean-Paul Comet

University C ote d'Azur, I3S Laboratory, UMR CNRS 7271, CS 40121,
06903 Sophia Antipolis Cedex, France

{laetitia.gibart,helene.collavizza,jean-paul.comet}@univ-cotedazur.fr

When we model a complex biological system, we try to understand the causality chains that explain the different behaviours observed. However, these observations are often made under experimental conditions which are not necessarily comparable since they depend on the culture medium for example. The construction of a right modelisation therefore depends on our ability to take into account all this information in a single framework.

In this article, we show that well-known R. Thomas' modelling framework allows the simulations of successive environmental situations in a unique global network at the expense of the use of artefacts. Therefore, it becomes possible to search for parameter settings compatible with biological knowledge for all environments by just enumerating the parameter settings. Another option we recommend here, is a green extension of R. Thomas' framework with the notion of *environments*. For each environment, the regulatory network is adapted and parameter settings compatible with the associated biological knowledge are searched on a smaller search space. Then, these sets of settings are intersected to obtain those which yield the traces consistent with observations of all environments. This "divide and conquer" approach is amazingly more efficient than the global approach.

1 Introduction

Modelling a biological system aims at understanding the underlying chains of causalities which leads the system behave as observed. Biological systems are called complex because the underlying causalities are difficult to be extracted from global observation. Thus systems biology can be seen as the study of the interactions between the components of biological systems, and of the consequences of these interactions on functions and behaviours of these systems. In order to complicate the portrait of this research field, observations are often made under experimental conditions which are not necessarily comparable (constant supply of glucose, and reduced supply of oxygen for example).

Moreover, even in a given modelisation framework, several modelling choices are possible because different instantiations of dynamical parameters which pilot the behaviour of the model can lead to traces consistent with all observations.

If the modeller chooses a particular setting, when new information is known, the parameter identification step must be restarted from the beginning. The systematic approach would then consist in characterizing, at each step, all of the parameter settings consistent with current knowledge: when a new observation becomes available, the modeller just refines the previous set of consistent parameter settings by selecting only those that are also consistent with this new information.

In the 70's, qualitative models based on discrete mathematics [10,19] have proved useful to understand the main causalities that govern observed phenotypes [20,21], and the multivalued framework of R. Thomas and H. Snoussi has become a classic for biological regulatory networks. It aroused new interest when, in the early 2000s, formal methods came to complete this formalism [2], as well as that of signaling networks [5]. For example, we developed a genetically modified Hoare Logic [1] for characterizing the set of parameter settings making possible a particular trace (if known). If only global temporal properties are available, these properties are translated into a formal temporal logic and the right parameter settings are selected via a model checking decision procedure [12]. These two approaches are combined in *TotemBioNet*, a tool which enumerates all parameter settings and selects those that are consistent with the biological properties [4].

The issue addressed in this paper is how to mix up the search for all parameter settings that are consistent with the temporal properties of multiple environments. The difficulty arises from the very wide diversity of behaviours due to diverse environments. We consider in this paper that behavioural biological knowledge have already been translated into a formal temporal logic (here CTL, Computational Tree Logic) and first show how the classic R. Thomas' framework allows the design of a unique regulatory network that mimics the different environments. That requires the duplication of the states of the internal variables to allow different behaviours according to the environments. A global property encompassing all temporal properties in all environments is built and verified for each parameter setting.

But we recommend another option based on a "divide and conquer" approach: a green extension of R. Thomas' framework with the notion of *environments*. During the divide step, a specific (and thus smaller) regulatory network is built for each environment, and the sets of settings consistent with the associated temporal property are searched. During the combine step, the intersection of these sets is compute to obtained the settings which satisfy the properties for all environments.

Extending R. Thomas' framework with environments is more efficient than using a single network which takes into account all environments. When applied to a network modelling the main regulations of cellular metabolic pathways, the method based on a single network would only give the result after an unreasonable time (estimated to 49,1 years), whereas the parameter settings are computed in 44.6 min in this new framework.

Running Example. *Pseudomonas aeruginosa* is an opportunistic bacteria that can secrete mucus. Mucus production is due to the presence of the protein *mucB* which is activated through a genetic element called an *operon*. Moreover *mucB* inhibits the operon and the operon activates itself through several molecules. These individual influences are summarized in the influence graph of *Pseudomonas aeruginosa* framed in blue in Fig. 1.

Mucus produced by mucoid *Pseudomonas aeruginosa* is composed by alginate and its accumulation can lead to the creation of bacteria biofilms. When these bacteria affect the lung, they cause serious infections, particularly for Cystic Fibrosis (CF) diagnosed patient [16]. Microbiologists discovered that the transition of bacteria from non-mucoid state to mucoid one is due to a very high concentration of calcium-ion in the cell environment in lung of CF patients. This led us to add in Fig. 1 the environment variable *Calcium* (in green).

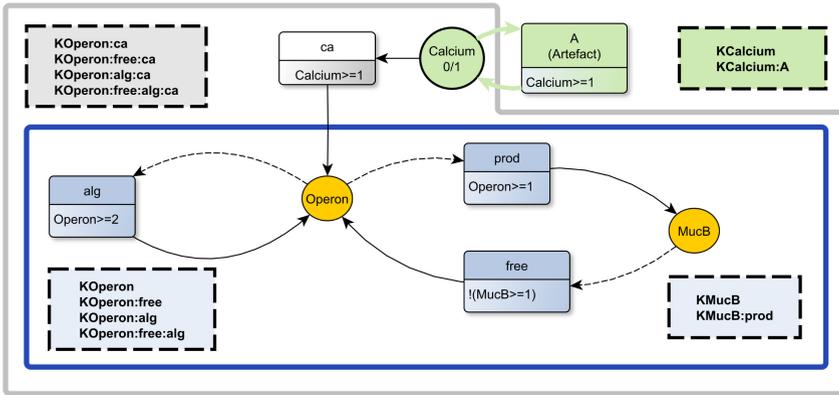


Fig. 1. Influence graph and parameters for *Pseudomonas aeruginosa* mucus production system. Blue frame: initial influence graph. Grey frame: influence graph with environment variable *Calcium* (green) playing a role in the lungs of Cystic Fibrosis patients. (Color figure online)

This article is organised as follows. Section 2 sketches the framework of regulatory networks. Section 3 is dedicated to the approach based on a unique regulation network encompassing all environments. Section 4 defines the regulatory networks with environments and presents the environment by environment approach. Section 5 then compares these two approaches from a theoretical and efficiency point of view. Finally, Sect. 6 presents the case study based on an abstract model of the regulation of the cellular metabolism.

2 Adding Environment Variables to Thomas' Framework

Given a biological system, and some hypothesis on the dynamics of this system, our modelling approach is to:

- Create an *influence graph* IG that defines the *individual* influences between variables,
- Deduce from IG the set of *parameters* which express the relative strength of influences on their common target. These parameters are used to define the *global dynamics* of the system, via a parameter setting P ,
- Find an appropriate translation of the biological knowledge in terms of a *temporal logic formula*,
- and at the end, use the *TotemBioNet* tool to find the parameter settings which make the dynamics based on IG and P consistent with this formula.

In this section, we first define the influence graph, and then introduce parameters. These two pieces of information describe the biological system and constitute a *regulatory network*. The next definitions concern the biological property: the temporal logic language (CTL) and the models of a CTL formula.

2.1 Regulatory Network with Multiplexes

Multiplexes were introduced in [11], as an extension of R. Thomas' modelling framework. They express, via a logic formula, some conditions under which an influence occurs. For example, if both variables a and b influence the variable c only when they form a protein complex, then, this coordinated influence can be represented in a multiplex which allows the effective influence only when both a and b are simultaneously present. A multiplex then combines in a unique predecessor some conditions on many variables. As the number of parameters to be identified depends directly on the number of predecessors, the introduction of multiplexes reduces the number of parameters (see Definition 2).

Definition 1 (Influence Graph with multiplexes). *An influence graph with multiplexes $IG = (V, M, A)$ is a directed graph such that:*

- Vertices are variables in V or multiplexes in M ($V \cap M = \emptyset$),
- With each variable $v \in V$ is associated a discrete domain $D_v = \llbracket v_l, v_u \rrbracket$ where $0 \leq v_l \leq v_u$,
- Arcs in A go from multiplexes to variables ($A \subset M \times V$),
- With each multiplex $m \in M$ is associated a formula φ_m which expresses the condition under which m influences its target variable(s). The language of multiplex formulas is defined by:
 - Atoms are atomic formulas ($v \geq n$) with $v \in V$ and $n \in D_v$
 - if φ, φ_1 and φ_2 are multiplex formulas, then $\neg\varphi, \varphi_1 \square \varphi_2$ are also multiplex formulas, where \square is either \wedge, \vee or \Rightarrow .

Given an influence graph, *parameters* represent the relative strength of influences on a variable, and by allowing to build the global behaviour of the system.

Definition 2 (Parameters). *The parameters of a variable v are denoted $K_{v,\omega}$ where ω is a subset of the predecessors of v in IG . $K_{v,\omega}$ represents the direction of evolution of variable v when it is controlled according to the multiplexes in ω .*

- A variable v has $2^{d^-(v)}$ parameters where $d^-(v)$ is the number of predecessors of v in IG .
- $\mathcal{K}(v)$ denotes the set of parameters of variable v .
- \mathcal{K} is the set of all parameters: $\mathcal{K} = \cup_{v \in V} \mathcal{K}(v)$.

Definition 3 (Regulatory network). A Regulatory network is a couple $\mathcal{N} = (IG, \mathcal{K})$.

Pseudomonas aeruginosa Regulatory Network. In Fig. 1, variables of IG are represented by circles and multiplexes (with their associated formulas) by round rectangles. Parameters are listed in dashed line rectangles, written in the concrete syntax of *TotemBioNet*. For example, $K_MucB:prod$ represents the case where $MucB$ is influenced by $prod$ while $K_Operon:alg:free$ represents the case where $Operon$ is influenced by both alg and $free$.

2.2 Formulas of Biological Properties and Their Models

We focus here on biological properties which express some global behaviour such that the existence of an attraction basin or a sustained oscillation. For such properties, we need to talk about future, and about the successive transitions of the system, in other words, to talk about the paths. Among the different temporal logics we have chosen the *Computational Tree Logic (CTL)*. Formulas of *CTL* are inductively built over variables in V in the usual way, using boolean operators and using modalities on time X ($neXt$), F (Future), U (Until), G (Generally), and modalities on paths E (Exists), A (All), see [7,9] where the semantics of these formulas is formally defined.

Definition 4 (CTL Language). Given an influence graph IG , the language \mathcal{L}_V of *CTL* formulas over V is defined by :

- an atom is either a boolean constant (\top, \perp), or a comparison $v \Delta n$ with $v \in V$, $n \in D_V$, and Δ an operator among $<, >, \leq, \geq, =$
- a *CTL* formula is either an atom or :
 - $\neg\varphi_1, \varphi_1 \nabla \varphi_2$ where ∇ is a boolean operator among $\wedge, \vee, \Rightarrow, \Leftrightarrow$
 - $E(\varphi_1 U \varphi_2)$ or $A(\varphi_1 U \varphi_2)$
 - $\odot \varphi_1$ where \odot is in EX, EF, EG, AX, AF, AG
 with φ_1 and φ_2 *CTL* formulas

Operon Stable States. From biologists we know that a non-mucoid *Pseudomonas aeruginosa* will never create mucus, and that when *Pseudomonas aeruginosa* is mucoid, it cannot turn off again. This can be translated by the following *CTL* formula:

$$\varphi_{bacteria}^0 \equiv ((Operon = 0) \Rightarrow AG(\neg(Operon = 2))) \wedge ((Operon = 2) \Rightarrow AG(\neg(Operon = 0)))$$

It expresses that starting from a state where $Operon = 0$ (non-mucoid) then there is no path leading to a state where $Operon = 2$ (mucoid), and vice versa.

The dynamics of a regulatory network $\mathcal{N} = (IG, \mathcal{K})$ is determined by a valuation of its parameters in \mathcal{K} , which is called a *parameter setting*. Our aim is therefore to find the parameter settings which lead to dynamics consistent with the *CTL* formula expressing a temporal biological knowledge.

Definition 5 (Parameter Setting). *Given a regulatory network $\mathcal{N} = (IG, \mathcal{K})$, a parameter setting $P \in \mathcal{PN}$ assigns a value in D_v to each parameter $K_{v,\omega}$ associated with variable v :*

$$\begin{aligned} P : \mathcal{K} &\rightarrow \llbracket 0, m \rrbracket \\ K_{v,\omega} &\mapsto k_{v,\omega} \in D_v \end{aligned}$$

where $m = \max\{v_u \mid v \in V\}$ and \mathcal{PN} denotes the set of parameter settings of \mathcal{N} .

A parameter setting $P \in \mathcal{PN}$ sets the global dynamics of the system. On each state, the possible transitions of the system depend on the *applicable parameters* of the variables. Indeed, the applicable parameter of a variable v in a particular state is determined by the combination of multiplexes targeting v whose formulas are true in that state. The value given by P to the applicable parameter of a variable defines how this variable can evolve.

Definition 6 (States). *The set of states of the system is the Cartesian product $S = \prod_{v \in V} D_v$. We denote $s \in S$ a state.*

Definition 7 (Applicable parameter on a state). *The applicable parameter of variable v on a state $s \in S$, is the only parameter $K_{v,\omega}$ s.t.*

- $\forall m \in \omega$, φ_m is true on state s ,
- $\forall m \in \text{pred}(v)$, $m \notin \omega$, φ_m is false on state s (where $\text{pred}(v)$ is the set of predecessors of v).

The applicable parameter of v in state s is denoted $K_v[s]$. Each multiplex m , predecessor of v , such that φ_m is true on s is called a resource of v in s .

Applicable Parameters for Pseudomonas aeruginosa. The applicable parameter of **Operon** for state *01* (where 0 is the value of *Operon* and 1 the value of *Mucus*) is **K_Operon** because neither $\varphi_{alg} = (\text{Operon} \geq 2)$ nor $\varphi_{free} = \neg(\text{Mucus} \geq 1)$ are true. The applicable parameter of **MucB** for state *21* is **K_MucB:prod** because $\varphi_{prod} = \neg(\text{Operon} \geq 1)$ is true.

Given a parameter setting P , a variable v and a state s , the value of the applicable parameter for v on s indicates if, on state s , v tends to increase, decrease or stay stable. This allows the definition of the (asynchronous) state transition graph which sums up the global behaviour of the system for this chosen parameter setting.

Definition 8 (ASTG for a parameter setting P). *Given a regulatory network $\mathcal{N} = (IG, \mathcal{K})$ whose variables are denoted v_1, \dots, v_n and $P \in \mathcal{PN}$ a parameter setting for \mathcal{N} , the associated Asynchronous State Transition Graph $ASTG_P$ is defined as follow:*

- **Vertices** are states $s \in S$,
- **Loops:** there is an arc from s to itself if $P(K_{v_i}[s]) = s_i, \forall i = 1 \dots n$. This expresses that each variable v_i has reached its focal value toward which it tends.
- **Arcs:** there is an arc from $s^p = (s_1^p, \dots, s_n^p)$ to $s^q = (s_1^q, \dots, s_n^q)$ if there exists one and only one index i s.t. $s_i^p \neq s_i^q$ with either : $s_i^q = s_i^p + 1$ and $P(K_{v_i}[s^p]) > s_i^p$ or $s_i^q = s_i^p - 1$ and $P(K_{v_i}[s^p]) < s_i^p$. This expresses that v_i has not reached its focal value and will increase (or decrease) by one level.

Finally, the two last definitions characterize what is a model of a property on the dynamics of the system. They are based on a decision procedure (such as a model checking procedure).

Definition 9 (Decision procedure for properties on \mathcal{N}). Given a regulatory network \mathcal{N} and the language \mathcal{L}_V of CTL formulas over variables of \mathcal{N} , $check_{\mathcal{N}}$ is a decision procedure for formulas in \mathcal{L}_V :

$$\begin{aligned} check_{\mathcal{N}} : \mathcal{PN} \times \mathcal{L}_v &\rightarrow Bool \\ (P, \varphi) &\mapsto true/false \end{aligned}$$

such that $check_{\mathcal{N}}(P, \varphi) = true$ iff φ is true on $ASTG_P$.

Definition 10 (Model of $\varphi \in \mathcal{L}_V$). $P \in \mathcal{PN}$ is a model of φ in the regulatory network \mathcal{N} iff $check_{\mathcal{N}}(P, \varphi) = true$. The set of models of φ is denoted $\mathcal{M}(\varphi) = \{P \in \mathcal{PN} \mid check_{\mathcal{N}}(P, \varphi) = true\}$.

2.3 Environmental Regulatory Networks

We are now ready to introduce *environments*, the new concept proposed in this article. In an influence graph with *environments*, some variables can be used to set the system into a specific configuration.

Definition 11 (Influence graph with environment variables). An influence graph with environment variables $IG_{EV} = (V, EV, M, A)$ is an influence graph s.t.:

- (V, M, A) is an influence graph,
- $EV \subsetneq V$ is an ordered set of environment variables,
- The environment variables have no predecessors in IG : $\forall v \in EV, d^-(v) = 0$,
- Environment variables appear in atoms of multiplex formulas.

Calcium Environment Variable. In the grey frame of Fig. 1, the **Calcium** variable is an environment variable which can be set to 0 or 1. It does not have any associated parameter, but the presence of **Calcium** adds new parameters to **Operon** variable which is targeted by **Calcium** through the multiplex **ca** (in the grey dashed-line rectangle).

Definition 12 (Environments). Given $IG_{EV} = (V, EV, M, A)$, an environment assigns a value to each environment variable of EV . It is a tuple of values (e_1, \dots, e_p) s.t. for all k in $1 \dots p$, $e_k \in D_{v_k}$, where $p = |EV|$ is the number of environment variables. The set of environments is denoted E .

Definition 13 (Environmental property). An environmental property is a couple (φ, e) where φ is a CTL formula in $\mathcal{L}_{V \setminus EV}$, the language of CTL formulas over systemic variables in $V \setminus EV$ and $e \in E$ is an environment. φ expresses a biological temporal property which occurs under the environmental condition represented by e .

Environmental Properties for Pseudomonas aeruginosa. Because **Calcium** variable can be set to 0 or 1, there are two environments: e^0 when *Calcium* = 0 and e^1 when *Calcium* = 1. In e^1 the bacteria will become mucoid, expressed by the CTL formula:

$$\varphi_{bacteria}^1 \equiv AG(AF(mucB = 1))$$

In e^0 , the bacteria does not change its phenotype: a mucoid (resp. non mucoid) bacterium will remain mucoid (resp. non mucoid); this is the formula $\varphi_{bacteria}^0$ used to illustrate Definition 4. Thus, there are two environmental properties $(\varphi_{bacteria}^0, e^0)$ and $(\varphi_{bacteria}^1, e^1)$.

Our main objective is now to study a system for which different dynamic behaviours are known under different environment contexts. So, the temporal biological knowledge that we search exhaustively for all models is a list of environmental properties $\Psi = [(\varphi^i, e^i)]_{i=1}^n$. We present in the next sections two ways to compute these models. The first one is a global approach which uses a modelling artefact to simulate the different environments in a single network. The second approach begins by defining as many regulatory networks with environment as there are environments : In each network, the number of parameters is reduced. It then computes the global models by intersection of the sets of models found for each of the environments.

3 All Environments' Coexistence in Thomas' Framework

In order to handle the biological knowledge on the behaviours of the system in different environments, the first idea is to design a unique regulatory network that takes into account the different environments. In this section, we present the formalisation of such regulatory network.

3.1 Regulatory Network

Because each environment can make the system behave in a peculiar way, the *ASTG* under construction has to handle “copies” of the useful states, one copy per environment. An environment variable is then considered as an inner variable which can take as many different qualitative levels as imposed by the environment. In the influence graph with environments presented in the previous section, environment variables have no predecessors. Then a unique parameter is associated with them, and each environment variable is attracted towards the value to its unique parameter, whatever its initial value, leading to an implicit change of value. To guarantee the stability of environment variables, we complete the influence graph by adding auto-regulations on each environment variable.

Auto-regulations for Simulating the Stability of Environment Variables. Let us suppose that the environment variable $v \in EV$ takes its value in $\llbracket v_l, v_u \rrbracket$. Then, for each value $n \in \llbracket v_l + 1, v_u \rrbracket$, one adds a multiplex \mathbf{A}_v^n whose formula is simply $v \geq n$. Thus when $v < n$ (resp. $v \geq n$), the multiplex \mathbf{A}_v^n is not a resource (resp. is a resource) of v . In Fig. 1, multiplex \mathbf{A} allows the variable *Calcium* to stay at level 0 (resp. 1) when initialised at 0 (resp. 1).

Multiplexes \mathbf{A}_v^n play a particular role inside this formalisation, because they do not represent any particular aspect of the biological system: they do allow the *ASTG* to simulate the existence of different stable values for the environment variables. In that meaning, they are artefactual.

Parameters Controlling the Stability of Each Environment. For each environment variable $v \in EV$, the set of parameters controlling its dynamics is (see Definition 2)

$$\mathcal{K}(v) = \{K_{v,\omega} \mid \omega \text{ is a subset of the predecessors of } v\}.$$

Since v is an environment variable, the only predecessors are the $n_v = v_u - v_l$ artefactual multiplexes \mathbf{A}_v^k , $k \in \llbracket v_l + 1, v_u \rrbracket$. Thus, there are 2^{n_v} parameters. Fortunately some of them are structurally inoperable (not at all useful for determining dynamics) and have

not to be instantiated: Indeed when $v = k$ ($k \in \llbracket v_l, v_u \rrbracket$), the only resources of v are the multiplexes \mathbf{A}_v^i with $i \leq k$. Thus parameters $K_{v,\omega}$ such that $\mathbf{A}_v^i \in \omega$ and $\mathbf{A}_v^j \notin \omega$ with $j < i$ are inoperable. All in all, when $v \in EV$ can take its value in $\llbracket v_l, v_u \rrbracket$, there exists exactly $v_u - v_l + 1$ operable parameters: $K_v, K_{v,\{A_v^{v_l+1}\}}, K_{v,\{A_v^{v_l+1}, A_v^{v_l+2}\}}, \dots, K_{v,\{A_v^{v_l+1}, A_v^{v_l+2}, \dots, A_v^{v_u}\}}$.

Moreover the stability of environment variables are strong properties that impose the values of previously defined operable parameters. Indeed, when environment variable v is set to k ($k \in \llbracket v_l, v_u \rrbracket$), it cannot change, leading to deduce that the value of the effective parameter is k . Thus $K_v = v_l, K_{v,\{A_v^{v_l+1}\}} = v_l + 1, K_{v,\{A_v^{v_l+1}, A_v^{v_l+2}\}} = v_l + 2, \dots, K_{v,\{A_v^{v_l+1}, A_v^{v_l+2}, \dots, A_v^{v_u}\}} = v_u$.

This regulatory network, including auto-regulations of environment variables and values of associated parameters that guarantee their stability, is denoted \mathcal{N}_{global} in the sequel.

3.2 Formula Summing Up all Behavioural Properties

Last but not the least, the different behavioural properties have to be expressed for this regulatory network containing all environments. We consider here that the biological knowledge has been summed up in a list of environmental properties $[(\varphi^i, e^i)]_{i=1}^n$. Obviously, the characterisation of the states corresponding to environment e^i is given by the formula: $\varepsilon^i \equiv \bigwedge_{v_k \in EV} (v_k = e_k^i)$ where e_k^i represents the value of environment variable v_k in the environment e^i . Naturally, the list of environmental properties can be transcribed in CTL formula:

$$\Phi_{global} \equiv \bigwedge_{i \in \llbracket 1, n \rrbracket} (\varepsilon^i \Rightarrow \varphi^i)$$

where ε^i characterises initialisation of environment variables. The values of parameters of environment variables used to build \mathcal{N}_{global} guarantee the stability of these variables. Finally one can use the decision procedure of *TotemBioNet* using the modified influence graph and Formula Φ_{global} : it enumerates all possible parameter settings, and selects only those that are consistent with Φ_{global} .

3.3 Application to *Pseudomonas aeruginosa*

We apply now this global approach for determining all parameter settings consistent with known behavioural properties of the *Pseudomonas aeruginosa* system. The auto-regulation of *Calcium* variable (multiplex and parameters) governing the *Calcium* environment variable is represented outside the grey frame in Fig. 1. Variable *operon* can take 3 different levels ($operon \in \llbracket 0, 2 \rrbracket$) and has 3 predecessors (itself, *mucB* and *Calcium*) and thus 3^3 parameters. Variable *mucB* can take 2 different levels ($mucB \in \llbracket 0, 1 \rrbracket$) and has a unique predecessor (*operon*) and thus 2^1 parameters. Finally, the number of parameter settings to consider is: $3^{2^3} \times 2^{2^1} = 26,244$.

Finally, the formula Φ_{global} is defined from $\varphi_{bacteria}^0$ and $\varphi_{bacteria}^1$ of Subsect. 2.3:

$$\Phi_{global} \equiv ((Calcium = 0) \Rightarrow \varphi_{bacteria}^0) \quad \wedge \quad ((Calcium = 1) \Rightarrow \varphi_{bacteria}^1)$$

4 Divide with Environments, Combine with Intersection

This “divide and conquer” approach works environment by environment. For each environment, a smaller regulatory network is used, and only the environmental property associated with the environment is checked. Afterwards, the models of the global system are built by abstracting and then intersecting the models found in each specific environment.

4.1 Regulatory Networks with Environments

Setting a value for an environment variable reduces the state space to the hyperplane defined by this value. This has a major impact on the size of the search space: some parameters of the targets of the environment variables become inoperable, leading to a drastic reduction of the number of parameter settings to consider.

Definition 14 (State space for an environment). *Given an influence graph with environment variables $IG_{EV} = (V, EV, M, A)$ and an environment $e \in E$, the state space of the system for e is $S_e = \prod_{v \notin EV} D_v \times \prod_{v_k \in EV} \{e_k\}$.*

Definition 15 (Operable parameters for an environment). *Given IG_{EV} , and an environment $e \in E$, a parameter $K_{v,\omega}$ is operable if there exists at least a state $s \in S_e$ where $K_{v,\omega}$ is applicable.*

Pseudomonas aeruginosa's Operable Parameters. For e^0 environment, the operable parameters are the original parameters in dashed rectangles in the blue frame of Fig. 1. Since *Calcium* targets variable *Operon*, the parameters associated with *Operon* change for e^1 environment, they are `K_Operon:ca`, `K_Operon:alg:ca`, `K_Operon:free:ca`, `K_Operon:alg:free:ca` listed in the grey frame.

Definition 16 (Regulatory network with environment). *A regulation network for environment $e \in E$ is the couple $\mathcal{N}_e = (IG_{EV}, \mathcal{K}_e)$ where $\mathcal{K}_e \subset \mathcal{K}$ is the subset of operable parameters for e . A parameter setting P_e assigns to each $K_v \in \mathcal{K}_e$ a value in D_v . The set of all parameter settings is denoted \mathcal{PN}_e .*

Definition 17 (ASTG for a parameter setting P_e in environment e). *Given $\mathcal{N}_e = (IG_{EV}, \mathcal{K}_e)$ a regulatory network for environment e and $P_e \in \mathcal{PN}_e$ a parameter setting for \mathcal{N}_e , the associated $ASTG_{P_e}$ is defined as follow:*

- **Vertices** are states $s \in S_e$,
- **Loops:** there is an arc from s to itself if $P_e(K_{v_i}[s]) = s_i, \forall v_i \in V \setminus EV$.
- **Arcs:** there is an arc from $s^p = (s_1^p, \dots, s_n^p)$ to $s^q = (s_1^q, \dots, s_n^q)$ if there exists one and only one index i s.t. $v_i \in V \setminus EV$, and $s_i^p \neq s_i^q$ with either : $s_i^q = s_i^p + 1$ and $P_e(K_{v_i}[s^p]) > s_i^p$ or $s_i^q = s_i^p - 1$ and $P_e(K_{v_i}[s^p]) < s_i^p$.

4.2 Formulas and Abstraction of Models

In this approach, for each environmental property (φ, e) , we successively search the models of the formulas φ associated with e . From a regulatory network with environment \mathcal{N}_e and a particular parameter setting $P_e \in \mathcal{PN}_e$, the associated transition graph is built and the decision procedure for formula φ is launched on this reduced graph.

Definition 18 (Decision procedure for properties on \mathcal{N}_e). Given a regulatory network \mathcal{N}_e , a parameter setting for this network $P_e \in \mathcal{PN}_e$, the language $\mathcal{L}_{V \setminus EV}$ of CTL formulas over systemic variables in $V \setminus EV$, $check_{\mathcal{N}_e}$ is defined by:

$$\begin{aligned} check_{\mathcal{N}_e} : \mathcal{PN}_e \times \mathcal{L}_{V \setminus EV} &\rightarrow Bool \\ (P_e, \varphi) &\mapsto true/false \end{aligned}$$

such that $check_{\mathcal{N}_e}(P_e, \varphi) = true$ iff φ is true on $ASTG_{P_e}$.

Definition 19 (Model of an environmental property). A model of an environmental property (φ, e) is the set of parameter settings which validate φ in \mathcal{N}_e : $\mathcal{M}_e(\varphi) = \{P_e \in \mathcal{PN}_e \mid check_{\mathcal{N}_e}(P_e, \varphi) = true\}$.

To be able to combine the models $\mathcal{M}_{e^i}(\varphi^i)$ for several e^i , which relate on different operable parameter sets, one needs to *abstract* the set of operable parameter settings to a common superset. Since $\mathcal{K}_{e^i} \subset \mathcal{K}$ for all e^i , each parameter setting P_{e^i} are abstracted by a subset of parameter settings in \mathcal{PN} .

Definition 20 (Abstraction of a parameter setting). Let \mathcal{PN} be the set of parameter settings for the regulation network with environment variables $\mathcal{N} = (IG_{EV}, \mathcal{K})$. Let $\mathcal{N}_e = (IG_{EV}, \mathcal{K}_e)$ the regulation network for the particular environment $e \in E$, and \mathcal{PN}_e its set of parameter settings.

An abstraction of a parameter setting $P_e \in \mathcal{PN}_e$ to \mathcal{PN} is the subset $\mathcal{AP}_e \subset \mathcal{PN}$ such that : $\forall P \in \mathcal{AP}_e, \forall K_{v,\omega} \in \mathcal{K}_e, P(K_{v,\omega}) = P_e(K_{v,\omega})$, and $\forall K_{v,\omega} \notin \mathcal{K}_e, P(K_{v,\omega}) \in D_v$.

In other words, P_e is the *projection* of \mathcal{AP}_e on parameters of \mathcal{K}_e .

Definition 21 (Abstraction of a model). The abstraction of a model of an environmental property (φ, e) is the union of abstractions of the parameter settings in $\mathcal{M}_e(\varphi)$: $\mathcal{AM}_e(\varphi) = \bigcup_{P_e \in \mathcal{M}_e(\varphi)} \mathcal{AP}_e$.

Given a list of environmental properties $[(\varphi^i, e^i)]_{i=1}^n$, the parameter settings satisfying all these properties is the intersection of the abstractions of the models of each (φ^i, e^i) .

Definition 22 (Model of environmental properties). Let $\Psi = [(\varphi^i, e^i)]_{i=1}^n$ a list of environmental properties. The model of Ψ is the set : $\mathcal{M}(\Psi) = \bigcap_{i=1}^n \mathcal{AM}_{e^i}(\varphi^i)$.

Example of abstraction. Let $\mathcal{K} = \{K_1, K_2, K_3, K_4\}$ be the parameters of \mathcal{N} with $D_1 = \llbracket 0, 1 \rrbracket$, $D_2 = \llbracket 0, 2 \rrbracket$, $D_3 = \llbracket 0, 3 \rrbracket$ and $D_4 = \llbracket 0, 1 \rrbracket$ the domains of their associated variables. Assume that $\mathcal{K}_{e^1} = \{K_1, K_4\}$ and $\mathcal{K}_{e^2} = \{K_1, K_2\}$ are the operable parameters for environments e^1 and e^2 .

Let $P_{e^1} \in \mathcal{PN}_{e^1}$ be a parameter setting of \mathcal{N}_{e^1} which assigns 0 to K_1 and 1 to K_4 (denoted $P_{e^1} = (0, -, -, 1)$). Let $P_{e^2} \in \mathcal{PN}_{e^2}$, $P_{e^2} = (0, 1, -, -)$ be a parameter setting of \mathcal{N}_{e^2} . Then $AP = (0, 1, 3, 1) \in \mathcal{PN}$ which assigns 0 to K_1 , 1 to K_2 , 3 to K_3 and 1 to K_4 belongs to \mathcal{AP}_{e^1} and to \mathcal{AP}_{e^2} . $AP' = (0, 0, 2, 1)$ also belongs to \mathcal{AP}_{e^1} . Furthermore, if P_{e^1} is a model of φ^1 and P_{e^2} is a model of φ^2 , then AP is a model of the list of environmental properties $[(\varphi^i, e^i)]_{i=1}^2$.

4.3 Application to *Pseudomonas aeruginosa*

We illustrate here the environment by environment approach on *Pseudomonas aeruginosa*. According to the values the environment variable *Calcium* can take, two environmental regulatory networks have to be constructed. Formula $\varphi_{bacteria}^0$ must be checked on the first one, and $\varphi_{bacteria}^1$ on the second one.

Some parameters become *inoperable* for some environments (see Def. 15). For example here, *Operon*, has three predecessors in the global approach. But in e^0 , the multiplex *ca* is not a resource anymore, and all parameters containing *ca* are in that case inoperable. This divides by two their number of parameter of *Operon*.

In e^0 , *Operon* has 2 predecessors (so has 2^2 parameters) and can take its value in $D_{Operon} = \llbracket 0, 2 \rrbracket$, thus it has 3^{2^2} parameter settings. *mucB* is not directly regulated by a environment variable, so the number of parameters which is 2^{2^1} , does not change compared to a not extended Thomas' Framework. Thus, for environment e^0 , the formula $\varphi_{bacteria}^0$ must be checked on $3^{2^2} \times 2^{2^1} = 324$ parameter settings. Similarly, there is the same amount of parameter settings to handle for checking $\varphi_{bacteria}^1$ in e^1 . Finally, there are $324 + 324 = 648$ parameter settings to consider.

5 Comparing the Two Approaches

5.1 Theoretical Point of View

The question which naturally arises, is to know if the two approaches compute the same models. The answer relies firstly on the link between the transition graph obtained for an environment and the subgraph of the global *ASTG* induced by an environment, and secondly on the link between the global formula and the environmental formulas.

Lemma 1 (Isomorphism between $ASTG_{P_e}$ and a subgraph of $ASTG_P$). *For each environment e , each parameter setting P_e of \mathcal{PN}_e and each $P \in \mathcal{AP}_e$, there exists a canonical isomorphism between $ASTG_{P_e}$ and the subgraph of $ASTG_P \in \mathcal{N}_{global}$ reduced to S_e .*

Proof of the lemma is given in Appendices. Following notations of Subject. 2.2 and Definition 22, we denote $\mathcal{M}(\Phi_{global})$ the set of models of Φ_{global} on the regulatory network completed with auto-regulations on environment variables and associated parameters.

Theorem 1 ($\mathcal{M}(\Psi) = \mathcal{M}(\Phi_{global})$). *Given a list $\Psi \equiv [(\varphi^i, e^i)]_{i=1}^n$ of environmental properties on an influence graph with environment variables $IG_{EV} = (V, EV, M, A)$, the set $\mathcal{M}(\Psi)$ of models of Ψ (computed environment by environment) is equal to the set of models of Φ_{global} on \mathcal{N}_{global} .*

Proof. (1) Let us consider a parameter setting P selected by the environment by environment approach. $P \in \bigcap_{i=1}^n \mathcal{AM}_{e^i}(\varphi^i)$ (Def. 22). For all $i \in \llbracket 1, n \rrbracket$, φ^i is satisfied in all states of $ASTG_{P_{e^i}}$, then, by Lemma 1, φ^i is satisfied in states corresponding to e^i in $ASTG_P$. The formula $(\varepsilon^i \wedge AG(\varepsilon^i)) \Rightarrow \varphi^i$ is then satisfied in all states of $ASTG_P$. We conclude that Φ_{global} is satisfied ($\mathcal{M}(\Psi) \subset \mathcal{M}(\Phi_{global})$).

(2) Conversely, let us consider now a parameter setting P selected by the classical approach: $P \in \mathcal{M}(\Phi_{global})$. Since $\Phi_{global} \equiv \bigwedge_{i=1}^n ((\varepsilon^i \wedge AG(\varepsilon^i)) \Rightarrow \varphi^i)$, for all $i \in$

$\llbracket 1, n \rrbracket$, φ^i is satisfied in all states defining the environment e^i . Thus, by Lemma 1, $P \in \bigcap_{i=1}^n \mathcal{M}_{e^i}(\varphi^i)$, in other words, $\mathcal{M}(\Phi_{global}) \subset \mathcal{M}(\Psi)$. \square

Let us just remark that this proof does not suppose that all e^i are different in the list Ψ . Then, if two behaviours (φ^1 and φ^2) are known in a common environment e , one can represent this information either by a unique environmental property ($\varphi^1 \wedge \varphi^2, e$) or by the list of environmental properties $[(\varphi^1, e), (\varphi^2, e)]$.

5.2 Practical Results

In this subsection, we illustrate the benefit of the environment by environment approach in terms of efficiency. We first give a short description of the tools used to search the models and to compute the intersection.

TotemBioNet: A Tool to Compute the Models of a Formula. The computation of the models of a formula is implemented in our tool *TotemBioNet* [4, 14], which is dedicated to the identification of parameters in R. Thomas' Modelling framework. Its inputs are an influence graph with environment variables, an environment e , the values of known parameters, and formalised behaviours, expressed either as Hoare triples for trace properties or as *CTL* formulas for temporal properties. *TotemBioNet* enumerates the parameter settings and for each $P_e \in \mathcal{PN}_e$ builds the *ASTG* $_{P_e}$ associated with P_e , and calls the model-checker *NuSMV* [6] as a decision procedure for a *CTL* property φ . The final *TotemBioNet* output is therefore $\mathcal{M}_e(\varphi)$. Successive environments are treated by calling *TotemBioNet* as many times as the number of environments, after having set the environment variables to their specific values. The outputs are written in .csv files which are the inputs of the intersection module.

MDDs to Compute the Intersection of Models. We choose to use *Multi Valued Decision Diagrams* (MDDs) to compute the intersection of sets of models from different environments. MDDs provide a compact representation for discrete data sets, with efficient set operations: intersection, union and set complement. Furthermore, it is easy (and efficient) to express that inside a common set of variables, certain variables can take any value. In this way, it is very convenient to abstract and combine sets of parameter settings which do not relate to the same variables because of inoperable parameters. *TotemBioNet* calls the *Colomoto mddlib* library¹ developed by A. Naldi to compute the intersection of the models obtained in successive environments. This library was designed for modelling biological systems, and notably to find stable states and analyse circuits [17].

Pseudomonas aeruginosa Execution Time. The 26, 244 parameter settings for the global approach (see Sect. 3.3) are enumerated and checked against the global formula Φ_{global} in 147.238 seconds on a personal computer.²

For the environment by environment approach, consistent parameter settings are computed by *TotemBioNet* in about 700 ms for each environment. Intersection of sets of models (using MDD to find the final result) needs 9.61 ms and thus the total time to

¹ <https://github.com/colomoto/mddlib>.

² All the given execution times are means over 20 *TotemBioNet* runs on an Intel[®] Core[™] i7-7700 CPU/3.60 GHz \times 8, RAM: 32 Go, under Linux. Interested readers can get the input files and results for the examples presented in this paper: <https://gitlab.com/totembionet/totembionet/-/tree/master/examples/CMSB2021>.

compute the models compatible for the two environments is 1.41 s, which is 105 times faster.

This example is a very small regulation network which involves only few variables. It already shows that the second approach is less time consuming. With a bigger regulatory network involving many environment variables, the relevance of the second approach is even more visible. The next section chooses a cell metabolism regulatory network as case study to show the scaling of the approach.

6 Case Study: Cell Metabolism

Cellular metabolism is a set of chemical reactions that occur in living cells. It involves intertwined biochemical reaction series, better known as *metabolic pathways*, which are thinly regulated. These processes allow cells to grow, multiply and maintain their structures [3]. Starting from classical biological knowledge of a healthy cell, we already proposed a qualitative regulatory network of the metabolism regulation [8] based on a precursor model [13]. Blue frame in Fig. 2 in Appendices represents the initial influence graph when the cell is in a healthy context. Unfortunately, when the context changes (nutrient lacks for example) some of these regulations are affected.

6.1 Metabolism Regulations According to Environments

Dependence of the metabolism on nutrient availability has been largely studied, and we decided to incorporate these dependences at a coarse grained level by adding four kinds of nutrients: sugar, amino acids, lipids and also oxygen. The level of their availability affects the regulation: For example, with oxygen supply the cell uses the oxidative phosphorylation pathway while without oxygen supply it uses fermentative pathways [15]. Grey frame in Fig. 2 incorporates environment variables representing nutrient supplies ($GLC \in \llbracket 0, 2 \rrbracket$ for sugar, $AA \in \llbracket 0, 2 \rrbracket$ for Amino Acids, $exO_2 \in \llbracket 0, 1 \rrbracket$ for oxygen and $FA \in \llbracket 0, 1 \rrbracket$ for Fatty Acids) and Table 1 (Appendices) lists all operable parameters of this regulation network. The parameters of variables which do not depend on environment variables have been all determined from biological knowledge, only parameters concerning NCD , O_2 and $GLYC$ remain to be identified (see below).

Globally, one has to consider $3 \times 3 \times 2 \times 2 = 36$ different environments, and for each environment, some properties are associated. The list $\Psi = [(\varphi^i, e^i)]_{i=1}^{36}$ of environmental properties is represented in Tables 2 and 3 (Appendices).

6.2 All Environments Coexistence in Thomas' Framework

In this approach, the time necessary for extracting all the models of Φ_{global} essentially depends on the number of parameter settings which *exponentially* depends on the number of unknown parameters. Thus, let us first count the number of unidentified parameters.

Taking into account FA does not add new multiplex (FA participates to an already present multiplex). So, it does not increase the number of parameters but the three others do. GLC acts on $GLYC$ through two distinct multiplexes: $g1c1$ and $g12$ which respectively specify the external sugar at level 1 and 2. Thus, with the 2 previously existing predecessors, the $GLYC$ in-degree becomes 4, leading to a number of parameters for $GLYC$ equal to $2^4 = 16$ (instead of $2^2 = 4$). Among them, 4 parameters are

structurally inoperable because when the formula of `glc2` is true, the formula of `glc1` is also true. Consequently, a parameter $K_{GLYC,\omega}$ where `glc2` $\in \omega$ and `glc1` $\notin \omega$ is not operable. Finally, GLYC has 12 parameters.

The target of AA, called NCD, gains also 2 predecessors (because AA has three levels) and has now 3 predecessors: $|\mathcal{K}(NCD)| = 2^3 = 8$. Two parameters are also structurally inoperable for NCD, because AA cannot reach level 2 without passing level 1, decreasing the number of parameters to 6. Finally, the target of `exO2`, that is O_2 , has now 2 predecessors leading to $|\mathcal{K}(O_2)| = 2^2 = 4$ parameters.

The number of parameter settings is equal to the product of the numbers of values that each of the parameters can take: the 12 (resp. 6, 4) parameters associated to GLYC (resp. NCD, O_2) can take 3 (resp. 3, 2) different values ($D_{GLYC} = D_{NCD} = \llbracket 0, 2 \rrbracket$ and $D_{O_2} = \llbracket 0, 1 \rrbracket$). This gives rise to a number of parameter settings equal to: $3^{12} \times 3^6 \times 2^4 = 6,198,727,824$.

Knowing that *TotemBioNet* performs 4.2 decisions $check_{\mathcal{N}}(P, \Phi_{global})$ per second (for this regulatory network \mathcal{N} , the formula Φ_{global} and any parameter setting $P \in \mathcal{NP}$), enumeration of all parameter settings would take approximatively 49,1 years.³

6.3 Divide with Environments, Combine with Intersection

The second option treats in an independent way each of the 36 environments on exponentially smaller search spaces, drastically decreasing the number of calls to the decision procedure.

Indeed, for each environment where `exO2`= 0 (resp. = 1), the number of operable parameters associated with O_2 is only two: $K_{O_2,\emptyset}$, $K_{O_2,\{\text{PHOX}\}}$ (resp. $K_{O_2,\{\text{exO}_2\}}$, $K_{O_2,\{\text{exO}_2,\text{PHOX}\}}$). It is the same for the other targets of environment variables. For each environment where AA= 0 (resp. = 1, = 2), the number of operable parameters associated with NCD is only two: $K_{NCD,\emptyset}$ and $K_{NCD,\{\text{KREBS}\}}$ (resp. $K_{NCD,\{\text{AA1}\}}$ and $K_{NCD,\{\text{AA1},\text{KREBS}\}}$, $K_{NCD,\{\text{AA1},\text{AA2}\}}$ and $K_{NCD,\{\text{AA1},\text{AA2},\text{KREBS}\}}$). In a similar manner, for each environment where GLC= 0 (resp. = 1, = 2), the number of operable parameters associated with GLYC is only 4. All in all, in each environment, *TotemBioNet* has to consider is $3^4 \times 3^2 \times 2^2 = 2916$ parameter settings.

For each environment, the 2916 decisions are computed by *TotemBioNet* in approximatively 74,365s. For the 36 environments, *TotemBioNet* needs 2677.142s. Adding 789ms for the computation of the intersection between all sets of models, the total execution time for extracting the exhaustive set of models is $(2677.142 + 0.789)/60 = 44.6$ min.

This proves the usefulness of the second approach when modelling larger influence graphs for which the first approach is unable to compute the models in an acceptable time. This second option is 579,103 times faster.

7 Conclusion

Our “divide and conquer” approach allows to reduce the time necessary for searching all models of a list of environmental properties in an unthinkable way. In fact, for a given influence graph, the global execution time quasi-linearly depends on the number of parameter settings which *exponentially* depends on the number of unknown parameters. The *environment by environment* approach seeks to reduce as much as possible

³ $49.1 = 6,198,727,824/(4.2 \times 3600 \times 24 \times 365.25)$, where 365.25 is for leap years.

the number of parameters to be identified by taking advantage of the fact that each environment (and by the way each environmental property) does not involve the whole set of parameters. This allows us to process large examples that were not yet accessible.

To process even more complex networks, it becomes manifest to parallelise the whole process: A coarse-grained parallelization is very easy because the searches of the models for each environment are completely independent.

From a modelling point of view, the environment variables are very powerful, because they constitute a good tool for exploring hypotheses. In particular, the consequences of Knock-Outs can be studied *via* such variables: If a Knock-Out leads to stopping a metabolic pathway, one can add an environment variable regulating this metabolic pathway, and impose *via* parameters of its target the decrease in the activity of its target.

In a longer term perspective, these environment variables constitute a first step towards a coupling of several sub-systems: Before embarking on the coupling, we can consider each of the studied subsystems with environment variables which control them differently in different part of the global phase space.

Appendices

Proof of the lemma. Each state of $ASTG_{P_e}$ is trivially unequivocally associated to a state of $ASTG_P$ (see Definitions 6 and 14). Let us show that transitions are the same. Let us consider a common state s . For determining the applicable parameters at s in $ASTG_{P_e}$, one has to evaluate the formulas of multiplexes controlling each non environment variable. Atoms concerning either environment or non environment variables are evaluated in the same way in $ASTG_{P_e}$ and $ASTG_P$ (the tuple representing s in $ASTG_{P_e}$ equals the one representing s in $ASTG_P$). Thus applicable parameters of non-environment variables at s are the same, leading to the same transitions that do not change the environment variables.

Moreover, because of our choice of parameter values for controlling the evolution of environment variables, there does not exist any transition in $ASTG_P$ that change the values of environment variables.

Thus, when all environment variables are fixed, the subgraph of $ASTG_P$ reduced to the states corresponding to environment e , and $ASTG_{P_e}$ are isomorphic. \square

Table 1. Operable Parameters associated with the metabolism regulation Influence Graph. The majority of parameters have been identified from biological knowledge, only parameters concerning NCD, O_2 and GLYC remain to be identified.

# Parameters for ATP	# Parameters for O_2	# Parameters for NADH
$K_{ATP} = 0$	K_{O_2}	$K_{NADH} = 0$
$K_{ATP, LBP} = 0$	$K_{O_2, PHOX}$	$K_{NADH, FERM} = 0$
$K_{ATP, nLBP} = 0$	K_{O_2, exO_2}	$K_{NADH, PHOX} = 0$
$K_{ATP, LBP nLBP} = 0$	$K_{O_2, exO_2 PHOX}$	$K_{NADH, AAS} = 0$
$K_{ATP, PHOX} = 1$		$K_{NADH, FERM PHOX} = 0$
$K_{ATP, nLBP PHOX} = 1$	# Parameters for GLYC	$K_{NADH, FERM AAS} = 0$
$K_{ATP, LBP nLBP PHOX} = 2$	K_{GLYC}	$K_{NADH, AAS PHOX} = 0$
$K_{ATP, GLYC1} = 0$	$K_{GLYC, GR}$	$K_{NADH, FERM AAS PHOX} = 0$
$K_{ATP, GLYC1 nLBP} = 0$	$K_{GLYC, GLC1}$	$K_{NADH, FERM KREBS AAS} = 0$
$K_{ATP, GLYC1 LBP nLBP} = 2$	$K_{GLYC, GLC1 GR}$	$K_{NADH, GLYC AAS PHOX} = 0$
$K_{ATP, GLYC1 GLYC2} = 1$	$K_{GLYC, GLC1 GLC2}$	$K_{NADH, GLYC FERM AAS PHOX} = 1$
$K_{ATP, GLYC1 GLYC2 nLBP} = 1$	$K_{GLYC, GLC1 GLC2 GR}$	$K_{NADH, KREBS FERM AAS PHOX} = 1$
$K_{ATP, GLYC1 GLYC2 LBP nLBP} = 2$	$K_{GLYC, COF}$	$K_{NADH, GLYC KREBS FERM AAS PHOX} = 1$
$K_{ATP, GLYC1 PHOX} = 1$	$K_{GLYC, COF GR}$	$K_{NADH, GLYC KREBS AAS} = 0$
$K_{ATP, GLYC1 nLBP PHOX} = 1$	$K_{GLYC, COF GLC1}$	$K_{NADH, FERM GLYC AAS} = 0$
$K_{ATP, GLYC1 LBP nLBP PHOX} = 2$	$K_{GLYC, COF GLC1 GR}$	$K_{NADH, GLYC KREBS} = 0$
$K_{ATP, GLYC1 GLYC2 PHOX} = 1$	$K_{GLYC, COF GLC1 GLC2}$	$K_{NADH, FERM KREBS} = 0$
$K_{ATP, GLYC1 GLYC2 nLBP PHOX} = 1$	$K_{GLYC, COF GLC1 GLC2 GR}$	$K_{NADH, KREBS AAS} = 0$
$K_{ATP, GLYC1 GLYC2 LBP nLBP PHOX} = 2$	# Parameters for nLBP	$K_{NADH, GLYC AAS} = 0$
	$K_{nLBP} = 0$	$K_{NADH, GLYC PHOX} = 0$
# Parameters for LBP	$K_{nLBP, PPP} = 1$	$K_{NADH, FERM GLYC} = 0$
$K_{LBP} = 0$	$K_{nLBP, AAS} = 1$	$K_{NADH, KREBS} = 0$
$K_{LBP, LS} = 1$	$K_{nLBP, AAS PPP} = 1$	$K_{NADH, GLYC} = 0$
$K_{LBP, BOX} = 0$		$K_{NADH, FERM GLYC KREBS AAS} = 1$
$K_{LBP, LS BOX} = 1$	# Parameters for KREBS	$K_{NADH, KREBS PHOX} = 1$
# Parameters for NCD	$K_{KREBS} = 0$	$K_{NADH, KREBS FERM PHOX} = 1$
K_{NCD}	$K_{KREBS, An0} = 1$	$K_{NADH, KREBS AAS PHOX} = 1$
$K_{NCD, KREBS}$	$K_{KREBS, An0 aKG} = 2$	$K_{NADH, GLYC KREBS PHOX} = 1$
$K_{NCD, AA1}$	$K_{KREBS, BOX} = 1$	$K_{NADH, FERM GLYC KREBS PHOX} = 1$
$K_{NCD, AA1 KREBS}$		$K_{NADH, GLYC KREBS AAS PHOX} = 1$
$K_{NCD, AA1 AA2}$	# Parameters for PHOX	# Parameters for FERM
$K_{NCD, AA1 AA2 KREBS}$	$K_{PHOX} = 0$	$K_{FERM} = 0$
	$K_{PHOX, PC} = 1$	$K_{FERM, EP} = 1$

References

1. Bernot, G., Comet, J.P., Khalis, Z., Richard, A., Roux, O.F.: A genetically modified Hoare logic. *Theoret. Comput. Sci.* **765**, 145–157 (2019)
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. *J. Theor. Biol.* **229**(3), 339–347 (2004)
3. Blanco, A., Blanco, G.: Chapter 13 - metabolism. In: Blanco, A., Blanco, G. (eds.) *Medical Biochemistry*, pp. 275–281. Academic Press, January 2017. <https://doi.org/10.1016/B978-0-12-803550-4.00013-6>
4. Boyenval, D., Bernot, G., Collavizza, H., Comet, J.P.: What is a cell cycle checkpoint? the TotemBioNet answer. In: CMSB, pp. 362–372 (2020)
5. Chabrier, N., Fages, F.: Symbolic model checking of biochemical networks. In: Priami, C. (ed.) CMSB 2003. LNCS, vol. 2602, pp. 149–162. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-36481-1_13
6. Cimatti, A., et al.: NuSMV 2: an OpenSource tool for symbolic model checking. In: Brinksma, E., Larsen, K.G. (eds.) CAV 2002. LNCS, vol. 2404, pp. 359–364. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-45657-0_29
7. Clarke, E.M., Emerson, E.A.: Design and synthesis of synchronization skeletons using branching time temporal logic. In: Kozen, D. (ed.) *Logic of Programs 1981*. LNCS, vol. 131, pp. 52–71. Springer, Heidelberg (1982). <https://doi.org/10.1007/BFb0025774>
8. Gibart, L., Khoodeeram, R., Bernot, G., Comet, J.P., Trosset, J.Y.: Regulation of eukaryote metabolism: an abstract model. submitted (2021)
9. Huth, M., Ryan, M.: *Logic in Computer Science: Modelling and reasoning about systems*. Cambridge University Press (2000)
10. Kauffman, S.A.: Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* **22**(3), 437–467 (1969)
11. Khalis, Z., Bernot, G., Comet, J.P.: Gene Regulatory Networks: Introduction of multiplexes into R. Thomas' modelling. In: *Proceedings of the Nice Spring school on Modelling complex biological systems in the context of genomics*, pp. 139–151. EDP Science, ISBN: 978-2-7598-0437-5 (2009)
12. Khalis, Z., Comet, J.P., Richard, A., Bernot, G.: The SMBioNet method for discovering models of gene regulatory networks. *Genes, Genomes Genomics* **3**(special issue 1), 15–22 (2009)
13. Khoodeeram, R., Bernot, G., Trosset, J.Y.: An Ockham Razor model of energy metabolism. In: Amar, P., Képès, F., Norris, V. (eds.) *Proceedings of the Thematic Research School on Advances in Systems and Synthetic Biology*, pp. 81–101. EDP Science (2017), ISBN: 978-2-7598-2116-7
14. Laetitia, G., Bernot, G., Collavizza, H., Comet, J.P.: TotemBioNet enrichment methodology: Application to the qualitative regulatory network of the cell metabolism. In: *BIOINFORMATICS 2021* (2021)
15. Liberti, M.V., Locasale, J.W.: The warburg effect: how does it benefit cancer cells? *Trends Biochem. Sci.* **41**(3), 211–218 (2016). <https://doi.org/10.1016/j.tibs.2015.12.001>
16. Malhotra, S., Hayes, D., Wozniak, D.J.: Cystic fibrosis and pseudomonas aeruginosa: the host-microbe interface. *Clin. Microbiol. Rev.* **32**(3), June 2019. <https://doi.org/10.1128/CMR.00138-18>

17. Naldi, A., Thieffry, D., Chaouiya, C.: Decision diagrams for the representation and analysis of logical models of genetic networks. In: Calder, M., Gilmore, S. (eds.) CMSB 2007. LNCS, vol. 4695, pp. 233–247. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75140-3_16
18. Richard, A.: Fair paths in CTL (2008), personal communication. <https://gitlab.com/totembionet/totembionet>
19. Thomas, R.: Boolean formalization of genetic control circuits. *J. Theor. Biol.* **42**(3), 563–585 (1973)
20. Thomas, R.: Logical analysis of systems comprising feedback loops. *J. Theor. Biol.* **73**(4), 631–56 (1978)
21. Thomas, R., Gathoye, A., Lambert, L.: A complex control circuit. Regulation of immunity in temperate bacteriophages. *Eur. J. Biochem.* **71**(1), 211–227 (1976)