Contribution of Computational Tree Logic to Biological regulatory networks : example from Pseudomonas aeruginosa

Sabine PERES[1] and Jean-Paul COMET[2]

[1] Physiologie Mitochondriale, INSERM EMI 9929,
146, rue Léo Saignat 33076 Bordeaux cedex, France
sabine_peres@yahoo.fr
[2] LaMI, Tour Evry 2, 523 Place des terrasses de l'agora 91000 Evry Cedex France
comet@lami.univ-evry.fr

Abstract. The Computational Tree Logic allows us to express some properties of genetic regulatory networks. These systems are studied using the feedback circuits evolved by René Thomas which constitute the semantic of our formal approach. We illustrate this formal language with the system of mucus production in *pseudomonas aeruginosa*, which is a mucoid bacteria that plays an important role in the cystic fibrosis. With the Thomas' theory, we could wonder if the mucoid state could be a steady state alternative to the non-mucoid state. We would like to know whether it is possible to have a recurrent mucoid state. Model-checking allows us to prove that the formula which expresses this property is satisfied by certain models. Moreover, using this formal language we can propose scenarii for confronting the model to experimentation.

1 Introduction

Biological experiments are motivated by hypotheses that biologists suggest. These hypotheses are more and more complex and it would be useful to express them in a form which can be handled by computer. All models which are planed to be proposed possess some properties corresponding to hypotheses previously mentioned [B02]. But a model which validates the hypotheses is not always the true model, it may fail with biological experiments. We have to call into question and to improve it. Thus, a model which agrees to biologists' experiments, ties up with the true model.

We present a formal language allowing to express in a modal logic [R00] some properties of a genetic regulatory network. Firstly, we present a formal description of the regulatory networks which are the semantic of this formal language. Secondly, we explain how model-checking is used to eliminate a majority of models. The model-checking is an automatic method to check if a model satisfies a logical formula which correspond for example to a temporal behaviour. To validate the retained models, a plan of experiment is proposed. Finally, we

illustrate this formal language using the system of mucus production in *Pseudomonas aeruginosa*, which is a mucoid bacteria that plays an important role in the cystic fibrosis [G01].

2 Formal description

Biological regulatory networks describe the interactions between genes and adjust production rate of system elements. Simulating them by a computer in order to predict their behaviour requires a formal description. The logical analysis developped by Thomas [TTK95-1,TT95-2] allows to study this type of models. It constitutes the basis of the purposed frame to study the regulatory networks, represented by an oriented graph (graph of interactions) of which the vertices represent the variables of the system (genes, proteins ...) and the edges represent the interactions between the variables. When the interactions form an oriented circuit, this constitutes a feedback loop. These loops regulate the production of the system variables and the behaviour of a specific variable depends on the possible paths.

Generally the interactions are described by differential equations and their solutions are sigmoid functions in shape. These functions determine values called thresholds, which correspond to the minimal concentration rate necessary for interaction between two variables. One approximates the sigmoids with step functions. In such a simplification, the number and location of steady states are preserved [ST93]. The expression level of each variable is discretised according to threshold values. But, there is no reason for all the thresholds of one variable to be equal because a variable does not act on the others with the same concentration. The interactions are ordered using the associated thresholds and the n^{th} threshold of a variable x is named s_x^n.

For example, let us suppose that a variable x acts on two variables x and y. The expression level of the different target variables in function of the expression level of x are sigmoids (figure 1). This figure shows three behaviours of x. For each of them, a specific discretised level of x is associated.

- x=0 : x acts on no variable
- x=1 : x acts only on y
- x=2 : x acts on y and x.

Generally, each edge $m \rightarrow n$ of the interaction graph is labelled with one of these thresholds and with the sign + if m has a positive influence on n and with the sign − if m represses n (graph of thresholds). For the same example, if x has a positive influence on y and on itself, the graph of interactions is presented in figure 2.

2.1 Graphs of states

To describe the evolution of the system with r variables, one has to make explicit the level reached by each variable depending on the presence (resp. absence) of

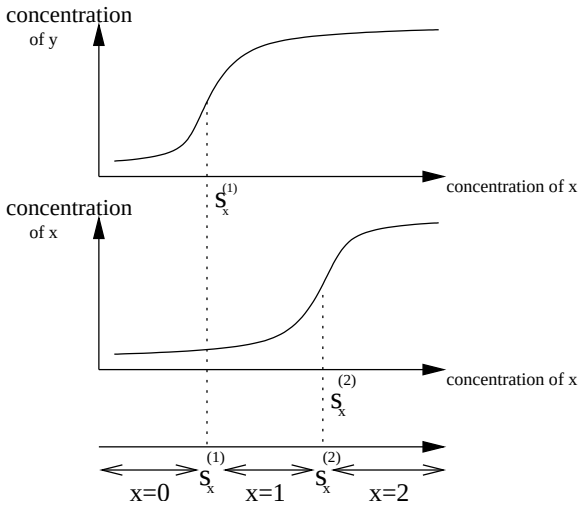

Fig. 1. Discretisation of the concentration of x

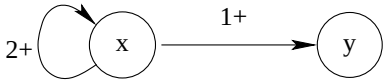

Fig. 2. Graph of interactions

variables which activate it (resp. which repress it). These values are represented by a function $\mathcal{K}$. This function takes two arguments as input (a target variable and the set of its activators) and returns the value towards which the target variable tends in presence of activators. This function can be determined using the same approach than Thomas [TT95-2,ST93], who used the different feedback loops and their functionality to find the parameters of the system. A positive feedback loop is functional if it ensures multi-stationarity and a negative one is functional if it ensures homeostasis. A loop is functional if the state located at the thresholds of the loop (*characteristic state*) is steady. One looks for the values of function $\mathcal{K}$ that make loops functional, obtains many contraints on the function $\mathcal{K}$ and can deduce its values (several solutions are possible).

So graphs of states are constructed :

- vertices are the r-uplets corresponding to each possible combination of values of the r variables
- the oriented edges ($n_1 \longrightarrow m_1$, $n_1 \longrightarrow m_2$, ...) give the possible futures of a state n_1.

The interactions are considered to be asynchronous : a variable does not act simultaneously on the others. Thus, the future states m_1, m_2,... of n_1 changes the value of one and only one variable in one step. The obtained level traces are the possible paths of these graphs which shows the possible behaviours of the system when time passes.

2.2 Model-checking

Biologists could frequently propose hypotheses when studying a real phenomenon. These hypotheses often express a temporal property of the entire system. For example, we could wonder in figure 2 : if at a given time x has a positive influence on y ($x = 1$) then at the next time x has always a influence on y ($x = 1$). When a set of models is proposed, it is necessary to keep those which validate the hypotheses. We want to check if a model fulfills the temporal properties associated to the hypothesis.

In this aim a formal language is introduced, based on the temporal logic [E90] which is adapted to this kind of models. It allows to express temporal properties of the system. The temporal logic chosen here is CTL[3][CE81,EH82]. It is a tree structure which can define several futures. It can be applied to our model because all the possible outcoming edges from a state of the graph define a different future. The precedent hypothesis on figure 2 can be translated in CTL :

$$(x = 1) \Rightarrow AX(x = 1)$$

where

- $\Rightarrow$ is the logical connective of implication
- AX is a temporal connective which means the next time in all possible futures.

The model-checking [R00] proves that a given system of finite states, satisfies or does not satisfy a temporal formula. It allows to find out properties of the system, for example the fact that the system tends to a given state. The algorithm takes a model, here a specific graph of states, and a formula ϕ as input and returns all the states which satisfy ϕ. The algorithm describes in [R00] can be summarized as follow :

1. Translate the fomula ϕ with the six chosen connectives (see Appendix).
2. For all sub-formula ψ of ϕ, determine the states which are labelled with ψ.
3. Return all the states which are labelled with ϕ.

There are several implementations of the model-checking algorithm. At the moment, the regulatory subnetworks only contain few variables [MTA99,ST2001], so a simple ad hoc checker is sufficient.

[3] for Computation Tree Logic, see the connective definition in Appendix

A model satisfies a formula if all the states which belong to the model satisfy the formula. So, if the algorithm returns all the states of the model, the model satisfies the hypothesis. The next section shows how we have used this method to select models of the bacteria *pseudomonas aeruginosa* which satisfy a temporal hypothesis emitted by biologists.

3 Results

Pseudomonas aeruginosa is a bacteria which secretes mucus in lung affected by cystic fibrosis. The mucus production increases the respiratory deficiency of the patients. But in a healthy lung, there is no production of mucus. Moreover, if one isolates a population of cells from a sick lung and if one puts it in a healthy environment, the mucus production can persist or resume progressively its non-mucoid phenotype after numerous generations. Biologists have shown that when the production persists, mucoid strain is generated by mutation. But, this model does not explain why it may happen that the mucus production stops.

The main regulator for the mucus production, AlgU, supervises an operon which is made up of 4 genes among which one codes for a protein that is a repressor of AlgU[4]. Moreover AlgU favors its own synthesis, which constitutes a positive feedback loop. This simplified model takes into account only interactions which are involved in feedback circuits (figure 3).

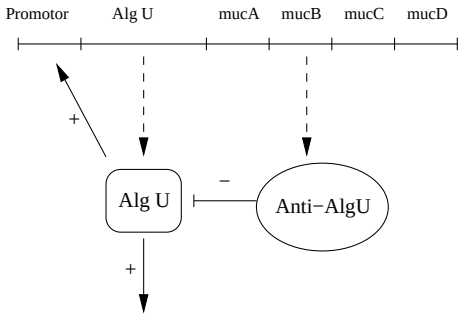

Fig. 3. Interactions of genes involved in the mucus production in P. aeruginosa

3.1 Hypothesis and modelisation of the bacteria

Thomas' theory [TTK95-1,TT95-2] allows us to assert that a positive feedback loop is a necessary condition for multi-stationarity. Moreover, epigenetic modifications could be a consequence of multi-stationarity [TK01]. Epigenetic modifications are phenotypic changes (transmitted from a cell to its progeny) without genetic or environmental modifications [G01].

[4] see for details [G01]

This is a reason why biologists researchers [G01] wonder whether the mucoid state is a steady state alternative to the non-mucoid state, which is activated by an external signal and self-maintained by a high concentration rate of AlgU. In other words, they wonder if the mucus production is an epigenetic phenomenon.

The model is represented by an oriented graph with two vertices x and y (Figure 4). x represents gene AlgU, y represents the inhibitor genes of gene AlgU. The edges depict :

- $x \to x$: self-maintenance of variable AlgU,
- $y \to x$: inhibitor effect on gene AlgU,
- $x \to y$: influence of AlgU on its own inhibitors.

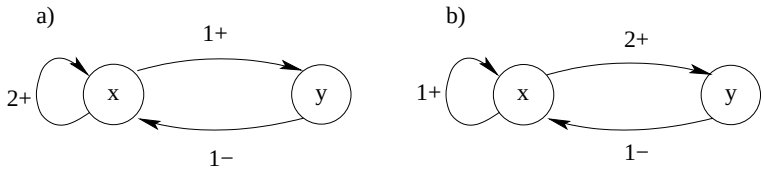

Fig. 4. Two possible graphs of thresholds

y acts on one variable, so the edge $y \longrightarrow x$ is labelled with 1. For biological reasons, one knows that the mucus production occurs when x is over its second threshold value (x=2). But we do not know if x acts on y at a lower threshold than it acts on x or at a upper one. So there are two possible graphs of thresholds (Figures 4a & 4b).

The state table of figure 4a is deduced. (X,Y) is the state towards which (x,y) tends. $\mathcal{K}_x\{y\}$ is the future value of x when y does not repress x in other word y is considered as an activator.

State table of figure 4a :

x y	X	Y
0 0	$\mathcal{K}_x\{y\}$	$\mathcal{K}_y\{\}$
0 1	$\mathcal{K}_x\{\}$	$\mathcal{K}_y\{\}$
1 0	$\mathcal{K}_x\{y\}$	$\mathcal{K}_y\{x\}$
1 1	$\mathcal{K}_x\{\}$	$\mathcal{K}_y\{x\}$
2 0	$\mathcal{K}_x\{x,y\}$	$\mathcal{K}_y\{x\}$
2 1	$\mathcal{K}_x\{x\}$	$\mathcal{K}_y\{x\}$

State table of figure 4b :

x y	X	Y
0 0	$\mathcal{K}_x\{y\}$	$\mathcal{K}_y\{\}$
0 1	$\mathcal{K}_x\{\}$	$\mathcal{K}_y\{\}$
1 0	$\mathcal{K}_x\{x,y\}$	$\mathcal{K}_y\{\}$
1 1	$\mathcal{K}_x\{x\}$	$\mathcal{K}_y\{\}$
2 0	$\mathcal{K}_x\{x,y\}$	$\mathcal{K}_y\{x\}$
2 1	$\mathcal{K}_x\{x\}$	$\mathcal{K}_y\{x\}$

3.2 Models fulfilling the formula

The values of $\mathcal{K}$ are calculated with the Thomas' theory. Several solutions are possible. The figure 5 is a possible state graph of figure 4a [5]. For $x = 1$ and $y = 0$, $X = \mathcal{K}_x\{y\} = 0$ and $Y = \mathcal{K}_y\{x\} = 1$. Because the state 10 tends to 01, we write in the state 10 the values 01. As the interactions are asynchronous, only one variable can change at a given moment. So the possible future states of 10 are 00 and 11.

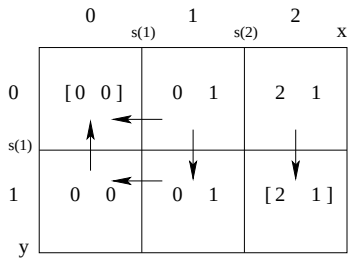

Fig. 5. Possible state graph of figure 4a

27 possible functions $\mathcal{K}$ are found, with the hypothesis that there is a functional positive feedbak loop in the system which is a necessary condition for multi-stationarity. So 27 state graphs can be constructed.

Two different cycles cohabit in the system : one positive feedback loop which imposes two steady states and a negative feedback loop which makes the system switch to one of the steady states according to the values of $\mathcal{K}$.

The aim is to show that the mucoid state can be reached without mutation, in other words that the loop $x \rightarrow x$ can be functional in the presence of y and so that it is possible, with the previous model, to obtain a recurrent state where $x = 2$. The following formula expresses the previous property :

$$(x = 2) \Longrightarrow AX\ AF(x = 2)$$

where $AX\ AF(\phi)$ means that for all possible futures (excluding the present) ϕ will be satisfied at a given time. The formula means that if the bacteria products mucus at a given time, then in a future time it will again product mucus. We want to show that for certain values of $\mathcal{K}$ which can be computed, $x = 2$ is a recurrent property.

We implemented a program which takes a function $\mathcal{K}$, a graph of interactions and a temporal formula (written in CTL) and returns true if the model satisfies the formula and false if not. The formula tested on the 27 models highlights 14 models which satisfy the hypothesis.

[5] One could construct the state graph of the figure 4b using the same method.

For example, the model which is represented in the figure 5 fulfills the formula which expresses the mucus production as an epigenetic modification.

The fact that these 14 models don't reject the epigenetic hypothesis can open new therapeutics in prospects. We have two classes of models : 14 models fulfill the hypothesis and 13 don't fulfill it. From this language, we can propose scenarii for confronting the models to experimentation. After experiment, we hope to be able to reject a class of models.

3.3 Experimentation

The experiment plan is deduced from the formula written in temporal logic. The following scenario allows one to test the property of the recurrent mucoid state : pulsing x to the value 2 by using an external signal, waiting a period of time in order to pass the transitory phase due to the pulse, and measuring whether x is above its second threshold. If, for a period of time, the state of the bacteria is mucoid, then the 14 models are considered as right and the others are rejected. But, if the experiment fails we cannot reject any class of model. If the experiment fails, it means that the mucus production has not been observed but it does not mean that the mucus production will never occur.

3.4 Limitations

The formal language does not take into account the external elements of the graph of figure 3 which is a subgraph of the graph with all the variables of the organism. Having neglected the edges outgoing from the subgraph has no serious consequences on our study since one is only interested on the subsystem supervising the mucus production. On the other hand, neglecting all the incoming edges could lead to an excessive simplification. If there are some edges regulating x and y whose influence does not change in the future, the only consequence of having extracted a subgraph is to shift the different thresholds associated to variables x and y. The system will have some other values for these thresholds, possibly some other values for $\mathcal{K}$, the steady states will not necessarily be the same, but the formula to be proved will remain identical. Only one case is a real problem : if there are some external regulators of x and y that have an influence which depends on time, the language will not allow one to translate the real behaviour of the system. The current study makes the hypothesis that these influences are negligible.

4 Conclusion

The presented method highlights the exixtence of models which are consistent with biologists' hypothesis. Thomas' theory is used to construct all the models and the temporal logic CTL is used to express a temporal property we want the system to have. Model-checking, a classical method in computer science, eliminates the models which do not fulfill the wished behaviours.

This method is applied to the production of mucus in *pseudomonas aeruginosa*. Thomas' theory on feedback loops is used to model some gene interactions of this bacteria. This theory says that a positive feedback loop is a necessary condition to have multi-stationarity. So to explain the phenomenon as an epigenetic shift, we only consider models that contain a functional positive feedback loop.

The hypothesis expressing the mucus production as an epigenetic modification would be in CTL : $(x = 2) \Longrightarrow AX\ AF(x = 2)$. 27 models are found by Thomas' theory, and 14 are kept by model-checking. But these models have to be validated by biological experiments. Thus, according to the CTL formula we propose a plan of experiment for confronting the model.

1. pulsing x to the value 2 by using an external signal
2. waiting a period of time in order to pass the transitory phase due to the pulse
3. measuring whether x is above its second threshold.

We do not have the result of the experiment yet, but if the experiment says the 14 models as right, we could have new points of view on the mucus production of this bacteria and on the therapeutic treatments. For example, instead of killing the bacteria, it would become possible to prevent a phenotypic shift to pathogeny [G01].

Acknowledgments

The research has led at the LaMI (LAboratoire des Méthodes Informatique) in Evry, France. We would like to thank Gilles Bernot for stimulating discussions and Janine Guespin for generous support.

Appendix : Note on Computation Tree Logic

Definition 1. *CTL formula is defined inductively with :*

- *a finite state of propositional variables :* $\{p_i\}$.
- *logical connectives :* $\bot, \top, \neg, \wedge, \vee, \Rightarrow$.
- *temporal connectives :* AX, EX, AG, EG, AU, EU, AF, EF.
- *rules of formation :*
  - p_i, $\bot$ *and* $\top$ *are formulas.*
  - *if* ϕ *and* ψ *are formulas, then* $(\neg\phi)$, $(\phi \wedge \psi)$, $(\phi \vee \psi)$, $(\phi \Rightarrow \psi)$, $AX\phi$, $EX\phi$, $A[\phi U \psi]$, $E[\phi U \psi]$, $AG\phi$, $EG\phi$, $AF\phi$, $EF\phi$ *are formulas.*

The logical connectives are the classical ones : *false*, *true*, *not*, *and*, *or*, *implication*. All the temporal connectives are pairs of symbols. The first element of the pair is A or E. The second one is X, F, G or U.

Meaning of the connectives :

- A : along All paths
- E : along at least one path (there Exist)
- X : neXt state
- F : some Future state
- G : all future states (Globally)
- U : Until

All the CTL connectives are equivalent to a combination of a set of six of them, for example $\{\bot, \neg, \wedge, AF, EU, EX\}$. So, all the formulas can be written with this set of connectives using the following equivalences.

Equivalences between CTL formulas [R00] :

1. $\neg AF\phi \equiv EG\neg\phi$
2. $\neg EF\phi \equiv AG\neg\phi$
3. $\neg AX\phi \equiv EX\neg\phi$
4. $AF\phi \equiv A[\top U\phi]$
5. $EF\phi \equiv E[\top U\phi]$
6. $A[pUq] \equiv \neg(E[\neg qU(\neg p \wedge \neg q)] \vee EG\neg q)$

References

[B02] G Bernot, J Guespin-Michel, A Zemirline, J-P Comet, F Delaplace, P Ballet, and P Amar. Modelling, observability and experiment: a case study, positive feedback loop in a genetic regulatory network. *Research report, LaMI, Univ. Evry, France*, 2002.

[CE81] E.M. Clarke and E.A. Emerson. Design and syntheses of synchronization skeletons using branching time temporal logic. In *Proc. Logics of Programs Workshop, Yorktown Heights, New York*, volume 131 of *LNCS*, pages 52–71. Springer, May 1981.

[E90] E.A. Emerson. *Handbook of theoretical computer science, Volume B : formal models and semantics*, chapter Temporal and modal logic, pages 995–1072. MIT Press, 1990.

[EH82] E.A. Emerson and J.Y. Halpern. Decision procedures and expressiveness in the temporal logic of branching time. In *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing, San Francisco, California*, pages 169–180, 5-7 May 1982.

[G01] J. Guespin-Michel and M. Kaufman. Positive feedback circuits and adaptive regulation in bacteria. *Acta Biotheoretica*, 49, 2001.

[R00] M. Huth and M. Ryan. *Logic in Computer Science: Modelling and reasoning about systems*. Cambridge University Press, 2000.

[ST93] E.H. Snoussi and R. Thomas. Logical identification of all steady states : the concept of feedback loop characteristic states. *Bulletin of Mathematical Biology*, 1993.

[TK01] R. Thomas and M. Kaufman. Multistationarity, the basis of cell differentiation and memory. i structural conditions of multistationarity and other nontrivial behavior. *Chaos*, 2001.

[TT95-2] R Thomas and D Thieffry. Dynamical behaviours of regulatory networks - II . immunity control in bactériophage lambda. *Bulletin of Mathematical Biology*, 57(2):277–297, 1995.

[TTK95-1] R Thomas, D Thieffry, and M Kaufman. Dynamical behaviours of regulatory networks - I . biological role of feedback loops and pratical use of the concept of feedback loop. *Bulletin of Mathematical Biology*, 57(2):247–276, 1995.

[MTA99] L. Mendoza, D. Thieffry, and E.R. Alvarez-Buylla. Genetic control of flower morphogenesis in arabidopsis thaliana : a logical analysis. *Bioinformatics*, 15(7/8):593–606, 1999.

[ST2001] L. Sanchez and D. Thieffry. A logical analysis of the drosophila gap-gene system. *J. Theor. Biol.*, 211:115–141, 2001.