Hybrid Model of gene regulatory networks, the case of the *lac-operon*

Laurent Tournier and Etienne Farcot

LMC-IMAG, 51 rue des Mathématiques,
38041 Grenoble Cedex 9, France
Laurent.Tournier@imag.fr, Etienne.Farcot@imag.fr

Abstract. The study of genetic regulatory networks has become an important field of research in biomathematics and biocomputing. With the early progresses made in biotechnology, there will be more and more data available concerning the expression of genes ; the understanding of such complex systems will then become crucial. However, even though several serious models have been proposed ([3], [1]), these systems still raise some challenging questions, from both qualitative and quantitative points of view. What we propose in this paper, after a brief description of the well-known *lac-operon*, is a hybrid model of the regulatory mechanism responsible for the treatment of lactose by the bacterium *e-coli*.

1 Introduction of the problem

The notion of *gene regulatory networks* comes from cellular biology. The cell, which is frequently used as experimental unit, can be viewed as a network of biochemical activities, involving gene products (mRNA and proteins in active and inactive states) and other biochemical entities (such as ions or other small molecules). We know today that proteins, which are bio-molecules essential to the life of the cell, are synthetized by the cell itself using the information contained in the DNA. This major concept is known as the **central dogma** of cell biology [4]. But even though it was a fundamental discovery, this dogma was not sufficient to explain how the cell regulates this synthesis during time and space in order to organize the different functions it must achieve to grow and divide.

Within the past decades, notably thanks to the work of F. Jacob and J. Monod, a lot of concepts have been developed by cell biologists to explain this regulation process. The first step was done thanks to the study of the mechanism of lactose treatment in the bacterium *e-coli* [2]. Jacob and Monod isolated a little group of genes that had a coordinated expression and they studied the mechanisms that were responsible for the regulation of this expression. This regulation unit was named **lac-operon**. Thanks to their work, it was actually confirmed that some genes were regulating the state of other genes, but they do so after a series of regulatory events, involving proteins and other molecules. So gene regulatory networks are complex systems, including different types of regulatory interactions. Since this discovery, biologists have isolated other operons and the

notion of gene regulation has become an active field of research (we can cite for instance the tryptophane operon [7] or the λ-phage [6]).

What we propose in this paper is a model of regulatory networks - through the typical case of the lac-operon - with the intention to approach as precisely as possible their dynamics. A lot of models have already been proposed, and we can classify them in two main categories : discrete and continuous models. The discrete approach was initiated by S. Kauffman who proposed a very interesting boolean idealization of gene regulation [3]. This was followed by R. Thomas who presented asynchronous logical networks [9], [8]. The continuous models are numerous ; we will cite notably the work of L. Glass who suggested an original piecewise linear approximation of gene interactions [1]. What we want here is to use the hybrid systems methodology, because it seems to us that the dynamics of operons (considering the example of the lac-operon) can be effectively modeled by a hybrid system, mixing discrete and continuous approaches.

2 Biological description of the *lac-operon*

Escherichia coli is a procaryote organism that has been very well studied by biologists. Its genome is entirely known, and we member more than 4000 genes along a single DNA molecule. It is clear that the bacterium cannot express all these genes continuously over the time. It must so detect with precision when a specific protein becomes necessary for its metabolism and then produce it.

Let us consider the case of two specific enzymes : the *lactose permease* and the *β-galactosidase*. These two enzymes are responsible for the treatment of lactose : The first one allows the bacterium to make external lactose enter the cytoplasm of the cell, while the second is used for degrading internal lactose in glucose, which is its basic source of energy. When there is no glucose left, and if lactose is present outside, these two proteins are vital for e-coli, whereas in different conditions, they are useless. The bacterium has so to detect whether these enzymes are necessary or not, and if it is the case, it must produce them quickly. This behaviour is ensured by the lac-operon. Let us now describe its operating mode.

As we can see on fig 1, the genes coding for permease and galactosidase are adjacent along the DNA molecule (with another gene coding for a third protein whose role will be neglicted in this paper). There exists a specific protein, noted i and named *repressor*, coded by another gene, I, that have an affinity with a special site of the DNA molecule, named **operator**, located just before the lactose genes (see fig 1). The binding between the repressor and the operator prevents the genes from being transcripted. When the repressor is present in suitable concentration in the cell, the expression of the lac-operon is therefore blocked. The repressor has another important property : when it is in presence of a lactose molecule, it binds to it on a specific sub-sequence of amino-acids. The effect of this binding is a change in the conformation of the protein, which results in the loss of its affinity with the operator (the repressor is then said to be *inactive*). As a consequence, when the bacterium is in presence of lactose, the

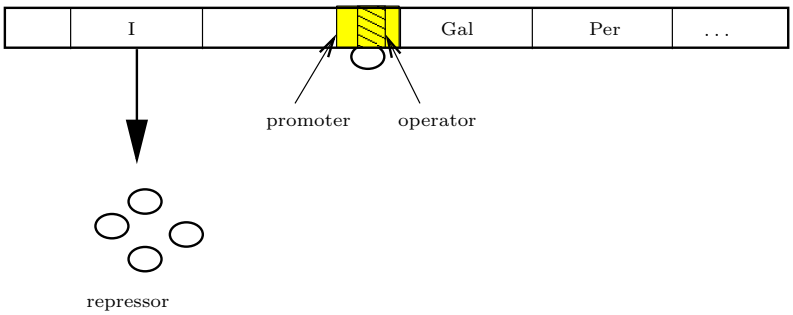

Fig. 1. *schematic representation of the lac-operon.*

latter penetrates the cell in very small quantity (thanks to the residual amount of permease) and this is enough to drop the concentration of active repressor and to start the expression of the lac-operon genes. This phenomenon is called **induction** and the role of inducer is therefore played by lactose itself.

This set of interactions is often called negative loop, but it is not the unique regulation process of the lac-operon. Indeed, cellular biologists pointed out another mechanism which is not dependent on lactose but on glucose. As a matter of fact, even if lactose is available, the bacterium will choose to treat it only if it is strictly necessary, that is to say, only if there is no glucose left. To that end, there exists a second loop, said positive, based on the concentration of glucose within the cell. As soon as glucose should happen to lack, the consequence is an increase of the concentration of the molecule *c-AMP*. This *c-AMP* binds itself with the protein *CAP* to form a complex that have a particular affinity with a site of the DNA molecule situated before the lactose genes. This site, named **promoter** (see figure 1), is the place where the RNA-polymerase binds itself to the DNA in order to initiate the transcription. The presence of the complex (*CAP-cAMP*) strongly increases the affinity between polymerase and the promoter, and thus allows to speed up the transcription (the acceleration is often estimated at a factor 50). That is why the drop of glucose concentration is another induction mechanism of the lac-operon. This phenomenon is known as **catabolite repression**. With its two regulation loops, the lac-operon allows e-coli to adapt its production in lactose-metabolism enzymes according to the external and internal conditions of the bacterium.

It is interesting to note that this description is representative of the functionning of any operon. Indeed, biologists discovered that the regulated genes often had one or several operators and that the sequences of these operators were similar. This is mainly thanks to these operators that regulating proteins like i may inhibit or, on the contrary activate the transcription. The understanding of the lac-operon regulation process is therefore important to have a deeper knowledge of gene regulation mechanisms.

3 Formalization of the gene regulation network

As we said previously, there already exist several models of the lac-operon in the litterature ([3], [10], ...), either discrete or continuous. The hybrid system that we build here is inspired from both approaches, trying to use the simplicity of the boolean view to represent biological switches, while studying the continuous evolution in time of biological variables.

In that case, we have to consider a priori five continuous variables representing the concentration of five important species. First, the concentrations of the two enzymes will be noted $x(t)$ and $y(t)$, x standing for the β-galactosidase and y for the permease. Then the concentration of RNA molecules issued from the transcription of the genes will be noted $z(t)$. We will also have to represent the concentrations in internal glucose and lactose, respectively by $g(t)$ and $l(t)$. Finally, we will have two *control* variables, $g_{ext}(t)$ and $l_{ext}(t)$ modeling the external sources of the two glucides.

If we consider the biological description we made in the first section, we notice two different parts in our system. The first part concerns the regulation at the genetic level and the second part takes into account the metabolism of the cell, involving the kinetics of enzymatic catalyse reactions.

3.1 Modeling genetic regulation

As we said, the regulation of the transcription of the lac-operon genes depends indirectly on the concentrations of glucose and lactose inside the cell. We decided in our model not to take into account the repressor and the activator molecules, but to make the production of enzymes result directly from the inner concentrations of the two glucides. To represent this dependency, we have been inspired by the discrete approach initiated by L. Glass [1], who models the interactions between genes with piecewise linear functions on rectangular boxes.

Let us consider two thresholds θ_g and θ_l such that :

$$\begin{cases} 0 < \theta_g < [g]_{max} \\ 0 < \theta_l < [l]_{max} \end{cases}$$

The phase space is thus cut into four boxes (see fig 2). Each one has a biological interpretation. For instance the box $\bar{G}L$ ($g < \theta_g$ and $l > \theta_l$) corresponds to the activated induction state.

Formally, the evolution of the RNA variable will follow a piecewise linear differential equation, with a piecewise constant production term and a linear decay term :

$$\dot{z} = \gamma(g, l) - \nu_1 z \tag{1}$$

with $\gamma(g, l)$ piecewise constant :

$$\gamma(g,l) = \begin{cases} \epsilon & \text{if : } l < \theta_l & \textit{(repression)} \\ \epsilon + \gamma & \text{if : } l > \theta_l \text{ and } g > \theta_g & \textit{(moderate induction)} \\ \epsilon + 50\gamma & \text{if : } l > \theta_l \text{ and } g < \theta_g & \textit{(accelerated induction)} \end{cases}$$

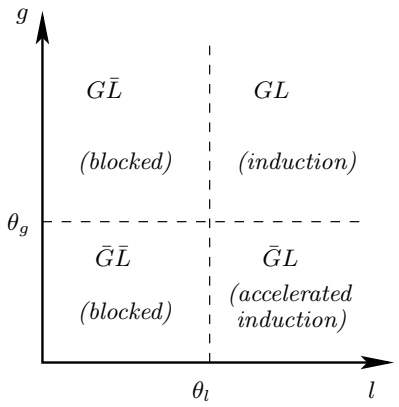

Fig. 2. *the four operating modes of the lac-operon.*

The term ϵ is a small perturbation term that represents a very weak production of the enzymes even if the transcription is blocked. Thanks to this assumption, we can assure that there is constantly a residual amount of enzymes in the cell in order to start the induction phenomenon.

The equation (1) represents the transcriptional process. Concerning the concentration in enzyme, we model the translational process by a linear differential system :

$$\begin{cases} \dot{x} = \alpha z - \nu_2 x \\ \dot{y} = \beta z - \nu_3 y \end{cases} \tag{2}$$

and we considered like in [10] that $\beta = 2\alpha$. (We will refer to table 1 for the values of the different constants)

The equations (1) and (2) establish the dynamics of the RNA and of the enzymes. They actually give the different biological modes we described previously, and as in any hybrid system, the biological switches are modeled by instantaneous switches. Though this may appear as a rough approximation, this piecewise linear model offers several advantages. First the equations can be solved formally, and secondly it allows an analytical analysis that makes possible a qualitative interpretation of the curves obtained (notably of the equilibria observed).

3.2 Modeling metabolism kinetics

At this point, we have to write the equations for glucose and lactose concentrations. These variables are critical because their values control the switches of the operon from one mode to another. Unfortunately, the usual differential equations used in biophysics to describe protein kinetics are often strongly non linear.

Let us establish the different terms of lactose and glucose appearance and disappearance :

- external lactose enters the cell thanks to the permease transporter,
- inside the cell, it disappears thanks to β-galactosidase.
- external glucose enters the cell thanks to a special mechanism named PTS (for sugar PhosphoTransferase System),
- glucose is produced by the degradation of lactose by β-galactosidase,
- glucose is degraded by the bacterium in order to produce energy.

Taking into account these factors, we can write the equations :

$$\begin{aligned} \dot{l} &= Cat(l_{ext}; y) - Cat(l; x) \\ \dot{g} &= PTS(glu_{ext}) + Cat(l; x) - Degr(g) \end{aligned}$$

where $Cat(substrate; enzyme)$ represents the kinetics of the enzymatic catalysis of the chemical reaction : $Substrate \xrightarrow{Enzyme} Product$; and $PTS(glu_{ext})$ and $Degr(g)$ represents respectively the supply in glucose and the degradation of glucose to produce energy.

We made here several **assumptions** :

First, the terms $PTS(glu_{ext})$ and $Degr(g)$ were approximated by linear terms as in [10], with a common coefficient δ.

Then, to model the terms $Cat(substrate; enzyme)$, we used the well-known Michaelis and Menten formalism [5]. The enzymatic catalyse reaction can be represented by the chemical equation :

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} (ES) \overset{k_2}{\to} E + P$$

where E is the enzyme, S the substrate, (ES) a short-life intermediate complex, and P the product of catalysis. The velocity of the reaction follows the Michaelis-Menten law :

$$\frac{d[S]}{dt} = -\frac{k_2[E][S]}{K_M + [S]}$$

here, $[A]$ is the concentration of metabolite A, and K_M is the Michaelis constant : $K_M = \frac{k_{-1}+k_2}{k_1}$. (see fig 3)
Following our will to obtain a piecewise linear differential system, we decided to propose a piecewise linear version of this velocity :

$$-\frac{d[S]}{dt} \simeq \begin{cases} v_m & \text{if : } [S] < \frac{K_M}{2} \\ v_m + \tilde{k}[E] & \text{if : } [S] > \frac{K_M}{2} \text{ and } [E] < [E]_{sat} \\ v_m + \tilde{k}[E]_{sat} & \text{if : } [S] > \frac{K_M}{2} \text{ and } [E] > [E]_{sat} \end{cases} \tag{3}$$

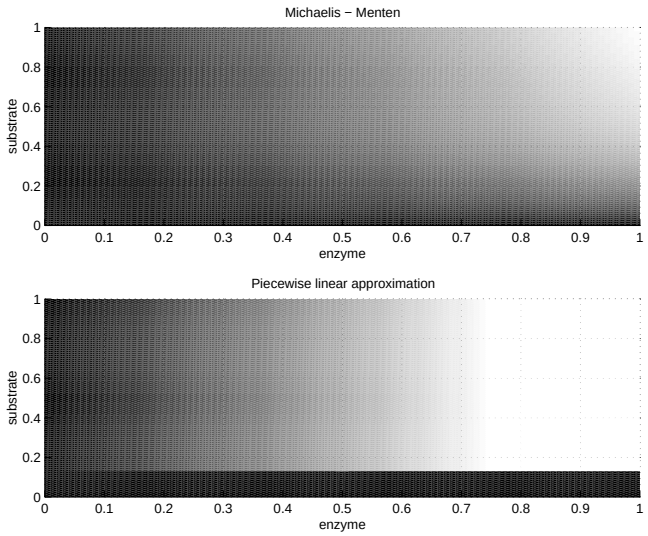

Fig. 3. *Michaelis-Menten function and piecewise linear approximation (3) with parameters :* $K_M = 0.25$*,* $k = 1$*,* $[E]_{sat} = 0.75$*,* $\tilde{k} = 0.7$ *and* $v_m = 0.1$

where v_m, $[E]_{sat}$ and $\tilde{k}$ are constant.

We can see on fig 3 the shape of the Michaelis-Menten surface and of its piecewise linear approximation.

We used this approximation to represent the two enzymatic reactions of our system :

$$Lactose_{ext} \xrightarrow{perm} Lactose_{int}$$
$$Lactose_{int} \xrightarrow{\beta-gal} Glucose + Galactose$$

4 Results

The model described in previous section has been implemented, with the set of parameters shown in table 1.
Although we do not claim that this model yields precise quantitative values, we tried to deal as far as possible with realistic orders of amplitudes. The concentrations are expressed in mol/g DCW (dry cell weight), and the time in minutes. With these units, we obtained plausible simulations, according to [10] results.

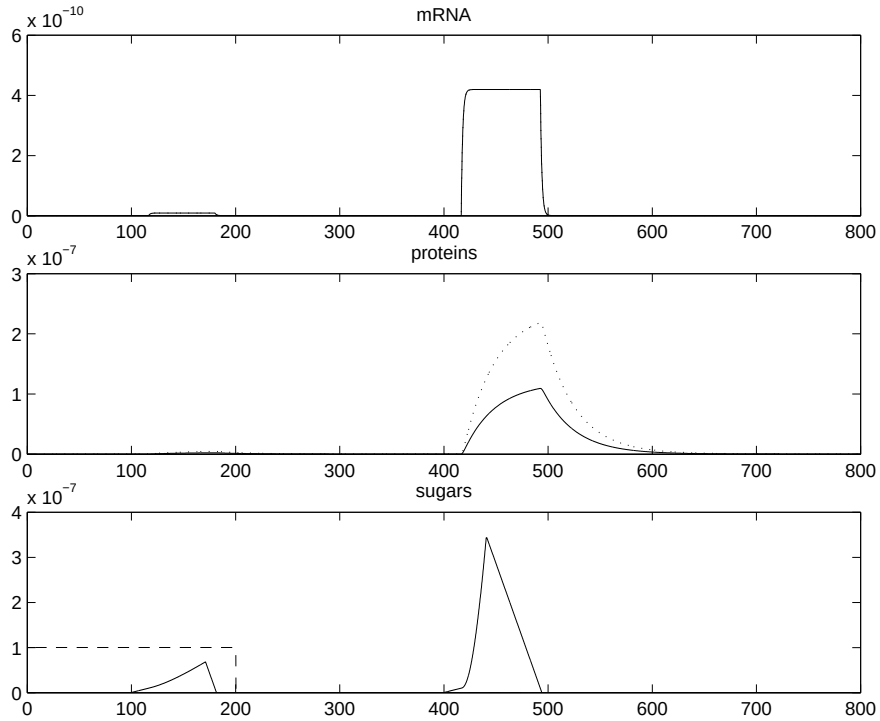

Fig. 4. *Evolution of different concentrations versus time. We can observe a period of moderate induction between times 100 and 200, and a period of active induction between times 400 and 500. The first of these is due to the simultaneous presence of lactose and glucose. The second to the presence of lactose only. In second plot, permease is the dotted line and β-galactosidase the full line. In the last plot, lactose is the full line and glucose the dashed line. Between moderate and active induction, there is no difference in the shape of curves, but the amplitude varies with a multiplicative coefficient of 50.*

name	**value**
ν_1	0.716
ϵ	$5e-13$
γ	$6e-12$
α	9.4
$\nu_2 = \nu_3$	0.033
δ	20
K_{M_1}	$4e-7$
K_{M_2}	$1e-9$

name	**value**
k_1	0.09
$\tilde{k}_2$	0.2
v_{m1}	$6.5e-9$
v_{m2}	$7e-9$
x_{sat}	$1.2e-7$
y_{sat}	$2.4e-7$
θ_g	$1e-7$
θ_l	$1e-8$

Table 1. Table of parameters used for the simulation

But the most important feature, beyond this quantitative considerations, is the ability of our piecewise linear model to reproduce the qualitative behaviours of the *lac*-operon, ie. the induction of this operon by lactose, moderated by the presence of glucose. This is shown in figure (4).

5 Future work

The model we present in this paper is built upon several assumptions. In comparison with the article [10], we neglicted some biological phenomenons that it would be interesting to integrate :

- the transport of glucides through membranes is a very complicated process including for instance the **inducer exclusion** phenomenon. Actually, the presence of glucose outside the cell would have a negative effect on the penetration of lactose inside the cell. Although the exact mechanism is not known, it is assumed (see [10]) that lactose transport follows saturation kinetics. It would be interesting to consider this assumption in the hybrid model.

- another fact is that lactose is also present in the cell in a different form : the *allo-lactose*. They follow similar kinetics but with different constants. It is the allo-lactose which is responsible for the induction of the operon.

- we can also imagine different piecewise linear approximations of Michaelis-Menten kinetics function, in order to better approach the enzymatic catalysis kinetics.

Finally, a subsequent step will be to apply the hybrid methodology to model other gene networks. We can cite especially the case of the tryptophan operon, which is quite similar to the lac-operon, and then the regulation network responsible for the λ-phage infection of e-coli, which is certainly more complex.

References

1. L. Glass, J. S. Pasternack *Prediction of limit cycles in mathematical models of biological oscillations.* Bull. Math. Biol. 40 : 27-44 (1978)
2. F. Jacob, J. Monod *Genetic regulatory mechanisms in the synthesis of proteins.* J. Mol. Biol. : 3-318 (1961)
3. S. A. Kauffman *The origins of order.* Oxford university press (1993)
4. B. Lewin *Genes VII.* Oxford university press (1999)
5. J. D. Murray *Mathematical Biology.* Springer - Verlag, New York (1989)
6. M. Ptashne *A Genetic Switch, second edition phage λ and higher organisms* Cell Press & Blackwell scientific publications (1992)
7. M Santillan, M. C. Mackey *Dynamic regulation of the tryptophane operon : a modeling study and comparison with experimental data.* P.N.A.S. 98 : 1364-1369 (2001)
8. E. H. Snoussi *Qualitative dynamics of piecewise-linear differential equations: a discrete mapping approach.* Dyn. Stab. Syst. 4(3-4) : 189-207 (1989)
9. R. Thomas, R. D'Ari *Biological feedback.* CRC Press (1990)
10. P. Wong, S. Gladney, J. D. Keasling *Mathematical model of the lac-operon : inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose.* Biotechnol. Prog. 13 : 132-143 (1997)