

# Modeling the Molecular Network Controlling Adhesion Between Human Endothelial Cells: Inference and Simulation Using Constraint Logic Programming

Eric Fanchon<sup>1</sup>, Fabien Corblin<sup>1,2</sup>, Laurent Trilling<sup>2</sup>, Bastien Hermant<sup>1</sup>,  
and Danielle Gulino<sup>1</sup>

<sup>1</sup> LCM and LIM, Institut de Biologie Structurale Jean Pierre Ebel,  
CEA-CNRS-Université Joseph Fourier,  
41, rue Jules Horowitz, 38027 Grenoble Cedex 1, France  
`eric.fanchon@ibs.fr`

<sup>2</sup> IMAG-LSR, Université Joseph Fourier BP 53, 38041 Grenoble Cedex 9, France  
`corblinf@ufrima.imag.fr`  
`laurent.trilling@imag.fr`

**Abstract.** Cell-cell adhesion plays a critical role in the formation of tissues and organs. Adhesion between endothelial cells is also involved in the control of leukocyte migration across the endothelium of blood vessels. The most important players in this process are probably identified and the overall organization of the biochemical network can be drawn, but knowledge about connectivity is still incomplete, and the numerical values of kinetic parameters are unknown. This calls for qualitative modeling methods. Our aim in this paper is twofold: (i) to integrate in a unified model the biochemical network and the genetic circuitry. For this purpose we transform our system into a system of piecewise linear differential equations and then use Thomas theory of discrete networks. (ii) to show how constraints can be used to infer ranges of parameter values from observations and, with the same model, perform qualitative simulations.

## 1 Introduction - Modeling Objectives

With the development of high-throughput projects the quantity of molecular level data is exploding. It is now clear that biology is entering a new era in which all these molecular components have to be assembled into a *system* in order to reach new levels of understanding.

In general terms, our goal is to formalize 'verbal models' or, stated differently, build a formal model from a word description of a biological phenomenon. This means in practice that the knowledge is incomplete, that most information is not precise but qualitative, and that we may have to deal with several hypotheses. In any case, we want to be able to exploit what we do have, even if it is qualitative information. In such a state of partial knowledge we view modeling as a tool

to formalize different competing hypotheses and explore their consequences; to help in interpreting new data and use data to discriminate among competing models; to infer parameters and to devise maximally informative experiments. In short we look for rigorous methods to reason about models and data in the context of incomplete knowledge on complex systems.

Our first goal is the integration of biochemical reactions and genetic regulatory interactions in a single unified framework. It is possible in the case of genetic networks to describe the regulatory interactions by logical (or discrete) equations [22] without explicit reference to Ordinary Differential Equations (ODEs). The situation is different in the case of biochemical or signal transduction networks because the types of reaction are more diverse (phosphorylation, complexation, transport, *etc.*). So, although differential equations are not well suited to this knowledge level, it is nevertheless useful to describe the cellular process in term of differential equations (even with unknown parameters) and to transform them into a discrete model with the same logical structure. The modeling process can be summarized as follows: graph of biochemical reactions  $\xrightarrow{(1)}$  detailed ODEs  $\xrightarrow{(2)}$  simplified ODEs  $\xrightarrow{(3)}$  PLDEs  $\xrightarrow{(4)}$  discrete equations/interaction graph. Step (1) is straightforward. It relates two equivalent descriptions and can be automated [19]. Step (2) involves biochemical approximations to obtain a system of lower dimensionality based on sigmoids. In transformation (3) sigmoid functions are replaced by step functions. At this stage the system often reduces to piecewise linear differential equations (PLDEs). For step (4) R. Thomas and others [20, 21, 22, 13] have developed a method to transform a special class of PLDE into discrete equations to perform a qualitative analysis of the dynamics.

The second goal is to show that such a formal description of the biological system at hand can be easily exploited via a Constraint Logic Programming (CLP) implementation. The advantages of the CLP approach are (i) that the implementation is expressed in a very similar way to the formal specification, thus guarantying the correctness of the implementation, (ii) that many different queries can be easily asked to this formal specification due to its logical form. For example, queries equivalent to simulation as well as inference of model parameters in a context of incomplete knowledge.

These principles are illustrated by the study of endothelial cell-cell adhesion. This case makes clear that the chemical reaction graph should not be confused with the interaction graph (as one might think from the study of genetic networks alone). Once a discrete model is at hand, we focus on the inference of parameter values satisfying some constraints about the existence of steady states and the existence of paths corresponding to the junction repairing process.

The organization of the paper is the following: after a short introduction to the biological phenomenon, we describe the pathway and its components. In the section 3 the modeling choices and approximations are explained. This allows us to transform the initial ODEs into a simpler PLDE system, with reduced dimensionality. The form of this system is slightly different from the one usually encountered in the field of genetic networks and we show how the theory of

Snoussi and Thomas [20] can be extended to account for this kind of piecewise linear system. We then perform an analysis following the lines of this theory. We explain briefly how the model is implemented in the CLP language Prolog IV [4] and give examples of queries.

## 2 Description of the Biological System

The phenomenon of cell-cell adhesion and its control by the cell is rather complex and our knowledge about it is far from complete.

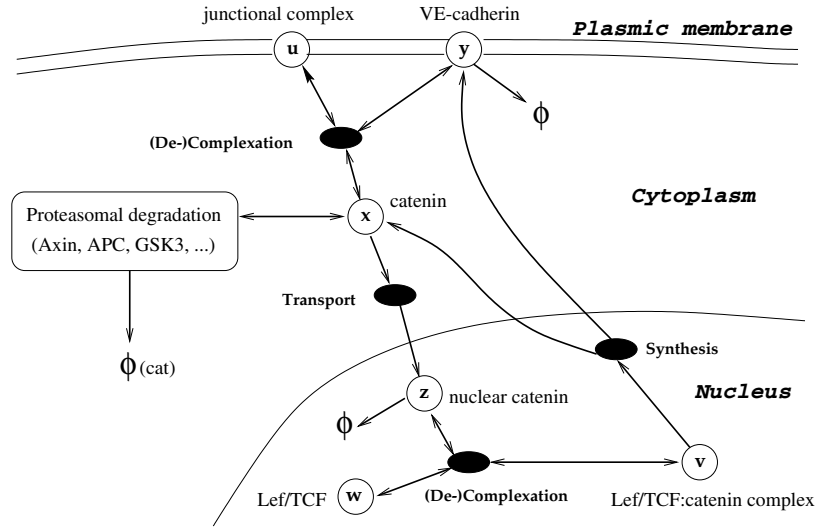
Blood vessels are lined with a monolayer of endothelial cells that form a barrier between blood and underlying tissues. This monolayer, called the endothelium, plays a central role in regulating the recruitment of leukocytes at sites of injury or inflammation. It does this by detecting changes in both the flow and chemical composition of blood, which triggers the expression and/or release on the cell surface of a variety of mediators.

Junctions between endothelial cells (adherens junctions) rely on the interaction of Vascular Endothelium (VE) cadherin molecules that are specifically expressed at these junctions [14, 15]. VE-cadherin is directly involved in the maintenance of endothelium permeability [9, 12] and in the control of the traffic of leukocytes from blood toward inflamed tissues [10]. Thus, following close contact between leukocytes and endothelial cells, proteases previously stored within leukocytes are transported at the cell surface and locally cleave VE-cadherin. This results in the subsequent disruption of adherens junctions that open the way for leukocyte migration. Once leukocytes have gone through, the integrity of endothelium must be restored to avoid an excessive accumulation of leukocytes within inflamed tissues.

This restoration process is the focus of our study. We are able to grow in culture endothelial cells extracted from human umbilical cords. The cells grow to confluence and reproduce an endothelium on a 2-dimensional plate. The migration of leukocytes through the tissue and subsequent destruction of adherens junctions is simulated in the culture by anti-cadherin antibodies which destabilize the junctional complexes. The biochemical structure of the system is informally illustrated on Fig. 1. We give now a brief description of the different components.

Adherens junctions in endothelial cells are constituted of VE-cadherin hexamers [16]. These VE-cadherins are membrane proteins with a cytoplasmic tail and a multidomain extracellular part. An hypothesis is that three VE-cadherin molecules from cell 1 and three from cell 2 self-assemble to form antiparallel hexamers. These hexameric units are held together via intracellular partners of VE-cadherin.

$\beta$ -catenin is one of these intracellular partners. It binds to the cytoplasmic part of VE-cadherin, and Fig. 1 displays its central role in adhesion. One hypothesis is that the VE-cadherin oligomer assemble first and then  $\beta$ -catenin binds to it. This is the hypothesis we incorporate in our model although this point is still controversial.  $\beta$ -catenin links cadherins to the actin cytoskeleton and has thus



**Fig. 1.** Schematic and informal representation of the endothelial cell-cell adhesion system. This graph stresses the overall architecture of the system. Circles represent chemical species and black ovals chemical reactions.  $\phi$  represents a degradation process. The molecules are distributed over three cellular locations: cytoplasm, nucleus and plasmic membrane. The concentration variables ( $x, y, z, u, v$  and  $w$ ) are defined by the triples (concentration variables, chemical species label, verbal definition) as follows: ( $x, cat$ , cytoplasmic unphosphorylated  $\beta$ -catenin), ( $y, cad$ , monomeric cadherin in the membrane), ( $z, catnuc$ ,  $\beta$ -catenin in the cell nucleus), ( $u, catcad3$ , complex of  $\beta$ -catenin with the cadherin trimer), ( $v, lefcat$ , complex of Lef/Tcf and  $\beta$ -catenin), ( $w, lef$ , Lef/Tcf transcription factor). Single-headed and double-headed arrows denote irreversible and reversible reactions, respectively

a structural role in the cell, but it is also able to act as a signalling molecule. Under certain conditions it can enter the nucleus where it complexes with members of the Tcf family of transcription factors, controlling gene expression. It is probable that this move to the nucleus is an active transport process (by opposition to passive diffusion) : a protein analogous to  $\beta$ -catenin, p120, is known to bind to motor proteins and to move along microtubules toward the cell nucleus (microtubules are molecular tubes made from a protein called tubulin). The real mechanism is poorly characterized and we will consider below two different ways of modeling it.

Cytoplasmic  $\beta$ -catenin is degraded by a complex called the proteasome.  $\beta$ -catenin is first complexed to APC and Axin (a scaffold protein), phosphorylated by two kinases acting in sequence (Ck1, Gsk3) and then released in the cytoplasm. The phosphorylated  $\beta$ -catenin is then tagged for destruction by the proteasome. The levels of proteins like Ck1 and Gsk3 are probably regulated but there is currently no information available. We thus considered that these proteins are maintained at fixed levels and consequently that proteasomal degra-

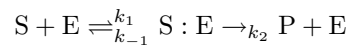
dation can be considered as a separate module with a single input. If regulatory interactions coupling proteins of this module to the system studied here are later identified, our model will have to be embedded in a larger model.

### 3 Modeling Choices - Derivation of a Simplified Differential System

We give here an outline of the derivation of simplified PLDEs from detailed 'chemical' ODEs. Our purpose is to obtain a system which is plausible from the biological point of view, knowing that other choices are possible in the current state of knowledge on cellular adhesion.

**Pseudo-steady State (PSS) Approximation.** The PSS approximation is the main approximation we are going to use. It is classically used in the derivation of the Michaelis-Menten equation for enzymatic reactions (see for example [17] for a detailed presentation) and of the equation for regulatory interactions. It is in fact potentially usable in cases where the first step is a reversible complexation-decomplexation reaction.

The classical mechanism of the Michaelis-Menten kinetics is:



where S is the substrate (concentration  $s$ ), E the enzyme (concentration  $e$ ), S:E the transient complex (concentration  $c$ ) and P the product (concentration  $p$ ). The parameters  $k_1$ ,  $k_{-1}$  and  $k_2$  are the kinetic constants. The substrate binds reversibly (double arrow) to the enzyme E. When the substrate is bound, reaction occurs, product P is released and enzyme E is ready for a new event. This mechanism is represented by a system of four differential equations (one for each chemical species). The rate of variation of the concentration  $c$  is given by:  $\dot{c} = k_1 \cdot s \cdot e - k_{-1} \cdot c - k_2 \cdot c$ .

In most biological situations the concentration of substrate is much larger than that of enzyme. There is a short transient time during which  $c$  increases very fast and then reaches a steady level ( $\dot{c} = 0$ ). This is the pseudo-steady state (PSS) hypothesis. In a closed system, substrate concentration decays slowly as reaction proceeds, and product concentration rises accordingly.

From  $\dot{c} = 0$  one gets the algebraic equation  $(k_{-1} + k_2) \cdot c = k_1 \cdot s \cdot e$  in which time does not appear explicitly, only implicitly through variables  $s$  and  $e$ . When using the PSS hypothesis we neglect the transient phase and assume that the concentrations adjust *instantly* to the steady-state values after a perturbation. In other words the characteristic time of the transient phase is supposed to be short with respect to the kinetics of the perturbation.

The total quantity of enzyme is conserved:  $e_0 = e + c$ . From this and the above relation we get:  $c = f(s) = \frac{e_0 \cdot s}{K + s}$  with  $K = \frac{k_{-1} + k_2}{k_1}$ .

The function  $f(s)$  for  $s \geq 0$  is a hyperbola branch. When cooperative binding occurs as in the case of allosteric enzymes (enzymes made of several sub-units

and thus several binding sites), the function  $f$  has sigmoid shape. The rate of production of P is:  $\dot{p} = k_2.c = k_2.f(s)$ .

The parallel with a regulatory interaction is readily seen: regulatory protein R (the analog of substrate S) binds reversibly to promoter Pr (DNA sequence); When R is bound to Pr, gene transcription occurs. The kinetic constant  $k_p$  associated to this process represents the average number of 'polymerase start' events per time unit. The previous reasoning applies, with the difference that nothing is consumed here. This is in fact a particular case in which  $k_2 = 0$  and thus the system is in quasi-*equilibrium* at all times. The rate of production by the regulated gene is then:  $\dot{p} = k_p.occ_{Pr} = k_p.f_p(r)$  where  $r$  is the concentration of R and  $occ_{Pr}$  the promoter occupancy. Since regulatory proteins are often dimers, binding to the promoter is cooperative and function  $f_p$  has sigmoid shape in most cases. In this treatment concentration fluctuations are assumed to be negligible (no stochastic effects). We apply the PSS approximation in two instances below.

**Reactions Taking Place in the Cell Nucleus.** The rates of variation of the nuclear species (see Fig. 1) are given informally by:

$$\begin{aligned}\dot{z} &= [\text{transport}] + [\text{decomplexation lefcat}] - [\text{complexation lef \& catnuc}] \\ &\quad - [\text{z degradation}] \\ \dot{v} &= [\text{complexation lef \& catnuc}] - [\text{decomplexation lefcat}] \\ \dot{w} &= [\text{decomplexation lefcat}] - [\text{complexation lef \& catnuc}]\end{aligned}$$

The lefcat complex activates several genes among which the  $\beta$ -catenin and cadherin genes [9]. This is a typical situation where the PSS approximation can be used. First at the usual step of binding to DNA (binding of the lefcat complex to the regulatory site) and also at the step of complexation of lef and catnuc:  $\dot{v} = -\dot{w} \approx 0$ . The analysis made above can be reproduced with protein Lef/Tcf in the role of E and nuclear  $\beta$ -catenin in the role of S:

$$v = f_v(z) = \frac{w_0.z}{K_v + z} \quad (1)$$

where  $w_0$  is the total quantity of Lef/Tcf in the nucleus, and  $K_v$  is the Michaelis-Menten constant of the enzyme.

The transport and degradation terms are given by:

$$[\text{transport}] = k_t.f_t(x), \quad [\text{z degradation}] = m_z.z$$

**Reactions Taking Place in the Cytoplasm and Within or Near the Plasmic Membrane.** Trimerization of cadherin and complexation of the trimer with  $\beta$ -catenin: we assume that these reactions are at quasi-equilibrium (they are fast with respect to synthesis, degradation, transport, etc...). Consequently the global reaction  $\text{cat} + 3.\text{cad} \rightleftharpoons \text{catcad}_3$  gives the relation:  $u = K_c.x.y^3$ , where  $K_c$  is the equilibrium constant.

Binding of  $\beta$ -catenin to cadherin may be cooperative, in which case the relation is different. This does not matter in our context, the important thing being that we have an algebraic relation  $g(x, y, u) = 0$ .

The remaining equations are:

$$\begin{aligned}\dot{x} &= [\text{x synthesis}] - [\text{transport}] - [\text{x proteasomal degradation}] \\ \dot{y} &= [\text{y synthesis}] - [\text{y degradation}]\end{aligned}$$

The bracketed terms are given by:  $[\text{x synthesis}] = k_{x0} + k_{xs} \cdot \sigma^+(v, s_v)$

$$\begin{aligned}[\text{x proteasomal degradation}] &= (m_{x1} + m_{x2} \cdot \sigma^+(x, s_{x2})) \cdot x \\ [\text{y synthesis}] &= k_{y0} + k_{ys} \cdot \sigma^+(v, s_v) \\ [\text{y degradation}] &= m_y \cdot y\end{aligned}$$

where  $\sigma^+(a, s_a)$  is a positive sigmoid with an inflection point at  $s_a$ .

Proteasomal degradation differs from spontaneous degradation by the existence of two levels ( $m_{x1}$  and  $m_{x1} + m_{x2}$ ). Note also that we include a basal level of expression of the  $\beta$ -catenin and cadherin genes ( $k_{x0}$  and  $k_{y0}$ ).

All occurrences of  $v$  are replaced by  $f_v(z)$  (Eq. 1). The [x synthesis] term can be rewritten as follows:

$[\text{x synthesis}] = k_{x0} + k_{xs} \cdot \sigma^+(f_v(z), s_v)$ . Under the condition that the threshold on lefcats is less than the lefcats saturation value, the composition of the hyperbola branch with a sigmoid gives a slightly deformed sigmoid (with parameters different from those of the original sigmoid):  $[\text{x synthesis}] = k_{x0} + k_{x1} \cdot \sigma^+(z, s_{z1})$  where  $k_{x1}$  is a new constant defined from  $k_{xs}$  and the hyperbola parameters.

We are left with the following system :

$$\begin{cases} \dot{x} = k_{x0} + k_{x1} \cdot \sigma^+(z, s_{z1}) - k_t \cdot f_t(x) - (m_{x1} + m_{x2} \cdot \sigma^+(x, s_{x2})) \cdot x \\ \dot{y} = k_{y0} + k_{y1} \cdot \sigma^+(z, s_{z2}) - m_y \cdot y \\ \dot{z} = k_t \cdot f_t(x) - m_z \cdot z \\ g(x, y, u) = 0 \end{cases}$$

Note that  $y$  does not influence other variables, and can be integrated once  $z(t)$  is known. Thus we can work on the sub-system constituted by the two differential equations on  $x$  and  $z$ . From a given solution  $(x(t), z(t))$ ,  $y$  and  $u$  can be calculated.

We consider two ways of *qualitatively* representing transport, called 'sigmoidal' and 'linear' for short:

- Linear transport:  $f_t(x) = x$
- Sigmoidal transport :  $f_t(x) = \sigma^+(x, s_{x1})$

In this last case the system we obtain is close to the classical form used in the field of genetic regulatory networks:

$$\dot{x}_j = h(x_1, \dots, x_i, \dots) - m_j \cdot x_j \quad (2)$$

where  $h$  is a *sum* of products of (positive or negative) step functions depending on several variables  $x_i$  (possibly including  $x_j$  itself).

The noticeable differences between these equations and ours are twofold: (i) the coefficient in the x degradation term is not a constant but contains a step

function, (ii) the complete system includes an algebraic equation. The three differential equations can be solved independently of the 4<sup>th</sup> equation since they do not involve  $u$ .

The version with a linear transport term introduces a third difference: the differential system is not diagonal any more since  $\dot{z}$  involves  $z$  and  $x$ .

## 4 Discrete Models and Thomas-Snoussi Theory

L. Glass [7, 8], R. Thomas [22] and others have developed a logical description for genetic regulatory networks. It allows to analyse *qualitatively* the dynamics of such networks. As said above the logical equations are derived from PLDEs of the form (Eq. 2). The variables  $x_i$  can be viewed as the concentrations of the proteins produced by the genes. A state  $\mathbf{x}$  of the system is defined by a vector  $(x_1, \dots, x_j, \dots)$  and a path is a sequence of states.

In the asynchronous description a focal point  $\mathbf{X}$  is associated to each state  $\mathbf{x}$ :  $\mathbf{X} = f(\mathbf{x})$ . The state  $\mathbf{X}$  is the state toward which the system *tends*. If only one variable differs between  $\mathbf{X}$  and  $\mathbf{x}$ , this variable changes value in the transition  $t \rightarrow t + 1$  and state  $\mathbf{X}$  is reached at the next step. If  $n$  variables differ between  $\mathbf{X}$  and  $\mathbf{x}$ ,  $n$  transitions are considered because by definition *only one* variable can change value in a transition. In such a description the system evolution is non-deterministic.

When logical equations are viewed as abstractions of PLDEs, it is easily seen that the probability for two or more variables to switch synchronously is marginal. The state space (concentration space) is divided into rectangular domains by the thresholds existing along each axis  $x_i$ : in each domain  $D_k$  the DE system is linear, and a focal point  $F_k$  can be associated to  $D_k$ :  $F_k = \phi(D_k)$ . Another important aspect is that the kinetic parameters are poorly known or totally unknown. Snoussi, Thomas and colleagues [20, 21, 22] have developed a method in which these parameters are discretized on the scale defined by the thresholds. This means that the location of the focal points is not precisely known. What is known is, for each domain  $D_k$ , which domain  $D_f$  the focal point  $F_k$  belongs to. From this mapping, the transitions between domains can be established and the *transition graph* built.

As mentioned at the end of section 3, the differential equations we obtain do not have exactly the form of those used to describe genetic networks. We have shown (to be published) that PLDEs with piecewise linear degradation terms can be transformed to the form used by Thomas and colleagues. It is thus possible to represent the logical structure of the system by an equivalent interaction graph and perform an analysis in term of circuits [20, 21, 22]. It is also in principle possible to extend the theory to PLDEs with off-diagonal terms, but the transformation becomes complicated, and implies the introduction of many discrete  $K_{a,i}$  parameters (see next section for the definition). In the following, we present only the discrete equations obtained on our specific case.



## 5 Model Implementation in Constraint Logical Programming

CLP is a programming technique based on a declarative approach. A program is constituted of a set of predicates stating what is known to hold true on the system.

The construction of a program proceeds in two steps:

1. Statement of a logical specification. In our case this step corresponds to the statement of predicates defining the formalism used (asynchronous networks with discrete parameters) and those defining the model itself. The model is represented a set of constraints corresponding to the discrete equations.
2. Statement of queries. Some of the parameters involved in the queries can be constrained while others are left unconstrained. In our context the parameters are typically the discretized kinetic parameters and sequences of states of the biological system. Examples are shown below.

In the case of a query with known kinetic parameters and unknown sequence of states the program returns the paths compatible with the parameters. This corresponds to a qualitative simulation. In the case of a query with known sequence of states and unknown parameters, the program returns the parameters compatible with the paths, or in other words an inference on the kinetic parameters (reverse-engineering). Mixed requests can also be formulated when part of paths have been observed and parameters are partially characterized. The reversibility property is one of the most powerful and interesting in this context. The same model description can be used for simulation, inference and mixed requests. Finally, constraints allow an efficient representation of *sets of* models. It is the addition of new constraints which allows to reduce the set of possible models.

### 5.1 The Two Models

The real parameters used in the models are the following:

$$\begin{aligned}
 K_{x,0} &= \frac{k_{x0}}{m_{x1}} & K_{x,3} &= \frac{k_{x0}}{m_{x1}+m_{x2}} & K_{x,6} &= \frac{k_{x0}}{m_{x1}+k_t} & K_{x,9} &= \frac{k_{x0}}{m_{x1}+m_{x2}+k_t} \\
 K_{x,1} &= \frac{k_{x1}}{m_{x1}} & K_{x,4} &= \frac{k_{x1}}{m_{x1}+m_{x2}} & K_{x,7} &= \frac{k_{x1}}{m_{x1}+k_t} & K_{x,10} &= \frac{k_{x1}}{m_{x1}+m_{x2}+k_t} \\
 K_{x,2} &= \frac{k_t}{m_{x1}} & K_{x,5} &= \frac{k_t}{m_{x1}+m_{x2}} & K_{x,8} &= \frac{k_t}{m_{x1}+k_t} & K_{x,11} &= \frac{k_t}{m_{x1}+m_{x2}+k_t} \\
 K_z &= \frac{k_t}{m_z}
 \end{aligned}$$

At this point, we make an additional approximation by replacing sigmoids ( $\sigma$ ) by step functions ( $\mathfrak{s}$ ).

**Model A : Model with Linear Transport.** From the PLDEs it is possible to derive the following equations:

$$\begin{aligned} X &= \mathfrak{s}^-(x, s_{x2}) \cdot [K_{x,6} + K_{x,7} \cdot \mathfrak{s}^+(z, s_{z1})] + \mathfrak{s}^+(x, s_{x2}) \cdot [K_{x,9} + K_{x,10} \cdot \mathfrak{s}^+(z, s_{z1})] \\ Z &= K_z \cdot X \end{aligned}$$

giving the coordinates  $(X, Z)$  of the focal point associated to the point  $(x, z)$ .

Now, following Thomas [22], we define discretization operators. If for example, a real variable  $a$  has two thresholds ( $s_{\text{Sup}}, s_{\text{Inf}}$  with  $s_{\text{Sup}} > s_{\text{Inf}}$ ) it is abstracted in a discrete variable  $\mathbf{a} = d(a)$  as follows:

$$\begin{aligned} \mathbf{a} = 0 &\Leftrightarrow a < s_{\text{Inf}} \\ \mathbf{a} = 1 &\Leftrightarrow s_{\text{Inf}} < a < s_{\text{Sup}} \\ \mathbf{a} = 2 &\Leftrightarrow s_{\text{Sup}} < a \end{aligned}$$

We introduce the following discrete variables:

$$\mathbf{X} = d_x(X), \mathbf{Z} = d_z(Z), \mathbf{x} = d_x(x), \mathbf{z} = d_z(z).$$

There are three thresholds in this model:  $s_{x2}$ ,  $s_{z1}$ ,  $s_{z2}$  and consequently two threshold orders have to be examined:  $s_{z1} < s_{z2}$  (model A1) and  $s_{z2} < s_{z1}$  (model A2). The parameterized discrete models A1 and A2 are defined in Table 1. Note that, even though  $s_{z2}$  does not appear in these two equations, it must be taken into account because  $y$  depends on this threshold.

**Table 1.** Tables for models A1 and A2

$\mathbf{x} \ \mathbf{z}$	$\mathbf{X}$	$\mathbf{Z}$	$\mathbf{X}$	$\mathbf{Z}$
0 0	$K_{x,6}$	$K_{z,6}$	$K_{x,6}$	$K_{z,6}$
0 1	$K_{x,6+7}$	$K_{z,6+7}$	$K_{x,6}$	$K_{z,6}$
0 2	$K_{x,6+7}$	$K_{z,6+7}$	$K_{x,6+7}$	$K_{z,6+7}$
1 0	$K_{x,9}$	$K_{z,9}$	$K_{x,9}$	$K_{z,9}$
1 1	$K_{x,9+10}$	$K_{z,9+10}$	$K_{x,9}$	$K_{z,9}$
1 2	$K_{x,9+10}$	$K_{z,9+10}$	$K_{x,9+10}$	$K_{z,9+10}$

There are eight discrete parameters  $K_{x,i}$  and  $K_{z,i}$  for these two models. These parameters  $K_{a,i}$  are defined as follows:  $K_{a,i} = d_a(K_{a,i})$ ,  $K_{a,i+j} = d_x(K_{a,i} + K_{a,j})$ . From this definition one can deduce the constraints:  $K_{x,6} \leq K_{x,6+7}$ ,  $K_{x,9} \leq K_{x,6}$ ,  $K_{x,9+10} \leq K_{x,6+7}$ ,  $K_{x,9} \leq K_{x,9+10}$ ,  $K_{z,6} \leq K_{z,6+7}$ ,  $K_{z,9} \leq K_{z,6}$ ,  $K_{z,9+10} \leq K_{z,6+7}$ ,  $K_{z,9} \leq K_{z,9+10}$ .

Due to the above inequalities, if  $K_{z,6+7} < 2$  then  $K_{z,6} < 2$ ,  $K_{z,9} < 2$  and  $K_{z,9+10} < 2$ . This means that  $\mathbf{Z}$  cannot be equal to 2 (see Table 1) and thus no state with  $\mathbf{z} = 2$  is reachable. We make the biological hypothesis that our system can reach states in which both  $\beta$ -catenin and cadherin are produced, which means that  $z$  is above both  $s_{z1}$  and  $s_{z2}$  ( $\mathbf{z} = 2$ ), so that we have necessarily  $K_{z,6+7} = 2$ . For each case (A1, A2), 84 parameter sets are compatible with the above constraints.

**Model B : Model with Sigmoidal Transport.** From the PLDEs it is possible to derive the following equations:

$$\begin{aligned}
X &= \mathfrak{s}^-(x, s_{x2}) \cdot [(K_{x,0} - K_{x,2}) + K_{x,1} \cdot \mathfrak{s}^+(z, s_{z1}) + K_{x,2} \cdot \mathfrak{s}^-(x, s_{x1})] \\
&\quad + \mathfrak{s}^+(x, s_{x2}) \cdot [(K_{x,3} - K_{x,5}) + K_{x,4} \cdot \mathfrak{s}^+(z, s_{z1}) + K_{x,5} \cdot \mathfrak{s}^-(x, s_{x1})] \\
Z &= K_z \cdot \mathfrak{s}^+(x, s_{x1})
\end{aligned}$$

There are four thresholds in this model:  $s_{x1}, s_{x2}, s_{z1}, s_{z2}$  and consequently four pairs of threshold orders have to be examined:  $s_{x1} < s_{x2}$  and  $s_{z1} < s_{z2}$  (model B1),  $s_{x1} < s_{x2}$  and  $s_{z2} < s_{z1}$  (model B2),  $s_{x2} < s_{x1}$  and  $s_{z1} < s_{z2}$  (model B3),  $s_{x2} < s_{x1}$  and  $s_{z2} < s_{z1}$  (model B4). The parameterized discrete models B1, B2, B3 and B4 are defined in Table 2.

**Table 2.** Tables for models B1, B2, B3 and B4

x z	X	Z	X	Z	X	Z	X	Z
0 0	$K_{x,0}$	0	$K_{x,0}$	0	$K_{x,0}$	0	$K_{x,0}$	0
0 1	$K_{x,0+1}$	0	$K_{x,0}$	0	$K_{x,0+1}$	0	$K_{x,0}$	0
0 2	$K_{x,0+1}$	0	$K_{x,0+1}$	0	$K_{x,0+1}$	0	$K_{x,0+1}$	0
1 0	$K_{x,0-2}$	$K_z$	$K_{x,0-2}$	$K_z$	$K_{x,3}$	0	$K_{x,3}$	0
1 1	$K_{x,0+1-2}$	$K_z$	$K_{x,0-2}$	$K_z$	$K_{x,3+4}$	0	$K_{x,3}$	0
1 2	$K_{x,0+1-2}$	$K_z$	$K_{x,0+1-2}$	$K_z$	$K_{x,3+4}$	0	$K_{x,3+4}$	0
2 0	$K_{x,3-5}$	$K_z$	$K_{x,3-5}$	$K_z$	$K_{x,3-5}$	$K_z$	$K_{x,3-5}$	$K_z$
2 1	$K_{x,3+4-5}$	$K_z$	$K_{x,3-5}$	$K_z$	$K_{x,3+4-5}$	$K_z$	$K_{x,3-5}$	$K_z$
2 2	$K_{x,3+4-5}$	$K_z$	$K_{x,3+4-5}$	$K_z$	$K_{x,3+4-5}$	$K_z$	$K_{x,3+4-5}$	$K_z$

There are nine discrete parameters  $K_{x,i}$  and  $K_z$  for these four models, which are constrained by their definitions:  $K_{x,0-2} \leq K_{x,0}$ ,  $K_{x,0} \leq K_{x,0+1}$ ,  $K_{x,0-2} \leq K_{x,0+1-2}$ ,  $K_{x,0+1-2} \leq K_{x,0+1}$ ,  $K_{x,3-5} \leq K_{x,3}$ ,  $K_{x,3} \leq K_{x,3+4}$ ,  $K_{x,3-5} \leq K_{x,3+4-5}$ ,  $K_{x,3+4-5} \leq K_{x,3+4}$ ,  $K_{x,3} \leq K_{x,0}$ ,  $K_{x,3+4} \leq K_{x,0+1}$ ,  $K_{x,3-5} \leq K_{x,0-2}$ ,  $K_{x,3+4-5} \leq K_{x,0+1-2}$ ,  $K_z = 2$ . The last constraint is due to the biological hypothesis, as stated above.

## 5.2 Organization of the CLP Program

Now a few words about the overall organization of the CLP program. A general predicate defines the asynchronous multivalued framework. The predicate `multivalued_async_model(Model, Path)` is true if `Path` is a possible path of the model `Model`. A path is a list of states. A model is defined by the type of transport and the threshold order. It is represented by a transition table (see Tables 1 and 2) and a set of constraints between the model parameters which are deduced from the definition of the model. A general constraint defined in `multivalued_async_model` is for example that a variable value changes by unit steps in a transition (it cannot jump from 0 to 2 or from 2 to 0).

Due to the fact that we obtain discrete equations from the PLDEs, it is possible to use the constraint solver on intervals of Prolog IV. In our case, variables

have finite domains. As an example, consider a case where a model `Model` is known, defined by a table `T` and parameters `P`. Then a typical query concerning the existence of steady states is the following:

```
Path = [S,S] ,
Model = [T,P] ,
multivalued_async_model(Model,Path) .
```

### 5.3 Example of Queries and Results

**Questions about Steady States.** In the case of sigmoidal transport (four models: B1, B2, B3, B4), we ask the list of all possible steady states (whatever the threshold orders). We find four states: (0, 0), (1, 0), (1, 2), (2, 2). An additional query allows to see that there exists no set  $\{K_{a,i}\}$  having more than two steady states. We obtain the same result, i.e. no more than two stable states, in the case of linear transport (two models: A1 and A2).

The next query is: What are the sets of  $K_{a,i}$  values having just two steady states? For model B3 for instance, there are five solutions: one having steady states (0,0) and (2,2) and four having (1,0) and (2,2). In the case  $\{(0,0),(2,2)\}$ , we interpret state (0,0) as the normal state in which  $\beta$ -catenin is present in the cytoplasm and in the nucleus at low levels. In all cases, (2,2) is a pathological state of constant over-expression.

**Question About the Perturbation.** From this point, one would like to know in case B with the two steady states (0,0) and (2,2) whether it is possible to take into account the repairing of the junctions. More precisely, when junctions are destabilized by antibodies, cadherins are destroyed and  $\beta$ -catenins are released in the cytoplasm. In our modeling the perturbation is represented by setting the system in a state with larger values of  $x$ . It implies that (2,2) cannot be considered as a normal state of the cell (because  $x$  cannot be augmented). The resulting query checks that it exists a path beginning in state (1,0) and containing a state where  $z = 1$  (the nuclear  $\beta$ -catenin concentration is observed increasing [9]) which finishes in state (0,0) (the normal steady state is reached). It appears that only model B2 is acceptable.

Also, one would like to know the parameter sets in case B for a unique steady state (excluding (2,2)), which takes into account the repairing of junctions. The resulting query checks that it exists a path beginning with the perturbed state and containing a state where  $z$  has been increased which finishes in the steady state. Model B3 admits 29 parameter sets having exactly one steady state. The added constraint on the existence of a restoration path eliminates 14 sets.

## 6 Conclusion and Perspectives

Using the PSS approximation we have been able to represent our biochemical system by PLDEs very similar to the ones used for genetic networks. This allowed

us to describe in a unified way regulatory interactions and biochemical reactions. This kind of analysis is not general and must be performed on each specific case.

Then, using an exploratory strategy and refining progressively the queries, we were able to deduce interesting properties having biological interpretation. We focused on stable states and models exhibiting 'return to stable state' paths and it appeared that such constraints reduce notably the number of possible parameter sets. We were not interested in oscillating behaviors, but it is of course possible to express constraint-based queries about cyclic paths.

**Related Work.** Constraint Satisfaction Problem (CSP) technology is used by V. Devloo [6] to discover efficiently the steady-state of completely instantiated asynchronous models. It is mentioned also that this approach *could* be used, as we do in this paper, to induce parameters from system behaviors.

Chabrier and Fages [2] describe a model-checking approach in CLP [5] (with linear arithmetic constraints) to check properties of qualitative or quantitative systems expressed in Computation Tree Logic (CTL): this study considers that the biological system is known.

Peres and Comet [18] pursue a similar goal as ours (inference of a Thomas network) by using model-checking and CTL, but without using constraints. They generate 27 model instances and check the validity of a CTL formula on each of them. This allows them to reduce to 14 possibilities. An approach similar to ours was advocated by J. Cohen for the case of gene regulatory networks [3].

Bockmayr and Courtois [1] use HCC (Hybrid Concurrent Constraint) which can tackle more general differential equations. But this technology does not appear to possess the full capabilities of constraints, namely to infer pre-conditions from post-conditions. For the purpose of taking into account more general differential equations. We are thinking to rather use an approach similar to the one proposed by Hickey [11].

It is worth noting that the CLP technology as used in this paper, provides not only constraint solvers (in particular CSP solvers) but also a very flexible and powerful way to express queries via logical formulas. The queries mentioned in section 5 are typical of this aspect as each of them introduces new constraints to be added to the constraints defined by the predicate `multivalued_async_model`.

**Perspectives.** DNA chip experiments as well as proteomic experiments are being done in our lab to identify new players (genes and proteins). This will extend the molecular network but will also bring new data and thus new constraints. Other biological hypotheses should be included concerning for example the adherens junction assembly process. Taking into account combinations of such hypotheses, we will generate a 'model space'. It will then be even more important to have formal reasoning tools to discriminate these models.

Also we intend to study carefully the language of the interesting queries which can be answered efficiently. The present network is tractable, but it could be a different matter as complexity increases.

## References

1. A. Bockmayer and A. Courtois. Using Hybrid Concurrent Constraint Programming to Model Dynamic Biological Systems. In Proceedings of the 18th International Conference on Logic Programming, LNCS 2401, Springer, 85–99 (2002).
2. N. Chabrier and F. Fages. Symbolic Model Checking of Biochemical Networks. Computational Methods in Systems Biology. In Computational Methods in System Biology 2003, C. Priami (ed.), LNCS 2602, Springer, 149–162 (2003).
3. J. Cohen. Approaches for simulating and modeling cell regulation : search for a unified view using constraints. In Linköping Electronic Articles in Computer and Information Science, 3, no7 (2001).
4. A. Colmerauer. Prolog – Constraints Inside, Manuel de Prolog, PROLOGIA, Case 919, 13288 Marseille cedex 09, France (1996).
5. G. Delzanno and A. Podelski. Model Checking in CLP. In Proceeding of the 5th International Conference TACAS'99, Springer, LNCS 1579, 223–239 (1999).
6. V. Devloo, P. Hansen and M. Labbé. Identification of All Steady States in Large Biological Systems by Logical Analysis. Bulletin of Mathematical Biology 65, 1025–1051 (2003).
7. L. Glass, S. A. Kauffman. Co-operative components, spatial localization and oscillatory cellular dynamics. J. Theor. Biol. 34, 219–237 (1972).
8. L. Glass, S. A. Kauffman. The logical analysis of continuous, non-linear biochemical control networks. J. Theor. Biol. 39, 103–129 (1973).
9. D. Gulino, E. Delachanal, E. Concord, Y. Genoux, B. Morand, M. O. Valiron, E. Sulpice, R. Scaife, M. Alemany and T. Vernet. Alteration of Endothelial Cell Monolayer Integrity Triggers Resynthesis of Vascular Endothelium Cadherin. The Journal of Biological Chemistry 273, 29786–29793 (1998).
10. B. Hermant, S. Bibert, E. Concord, B. Dublet, M. Weidenhaupt, T. Vernet and D. Gulino-Debrac. Identification of Proteases Involved in the Proteolysis of Vascular Endothelium Cadherin during Neutrophil Transmigration. The Journal of Biological Chemistry 278, 14002–14012 (2003).
11. T. J. Hickey and D. K. Wittenberg. Rigorous Modeling of Hybrid Systems using Interval Arithmetic Constraints. In Technical Report CS-03-241, Computer Science Departement, Brandeis University (2003).
12. P. L. Hordijk, E. Anthony, F. P. Mul, R. Rientsma, L. C. Oomen and D. Roos. Vascular-Endothelial-Cadherin Modulates Endothelial Monolayer Permeability. J. Cell Sci. 112, 1915–1923 (1999).
13. H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Hybrid modeling and simulation of genetic regulatory networks: A qualitative approach. In Hybrid Systems : Computation and Control (HSCC 2003), A. Pnueli and O. Maler (ed.), LNCS 2623, Springer, 267–282 (2003).
14. M. G. Lampugnani, M. Corada, L. Caveda, F. Breviario, O. Ayalon, B. Geiger and E. Dejana. The molecular organization of endothelial cell to cell junctions: differential association of plakoglobin,  $\beta$ -catenin, and  $\alpha$ -catenin with vascular endothelial cadherin (VE-cadherin). J. Cell Biol. 129, 203–217 (1995).
15. M. G. Lampugnani, M. Resnati, M. Raiteri, R. Pigott, A. Pisacane, G. Houen, L. P. Ruco and E. Dejana. A novel endothelial-specific membrane protein is a marker of cell-cell contacts. J. Cell Biol. 118, 1511–1522 (1992).
16. P. Legrand, S. Bibert, M. Jaquinod, C. Ebel, E. Hewatt, F. Vincent, C. Vanbelle, E. Concord, T. Vernet and D. Gulino. Self-assembly of the vascular endothelial cadherin ectodomain in a  $\text{Ca}^{2+}$ -dependent hexameric structure. Journal of Biological Chemistry 276, 3581–3588 (2001).

17. J. D. Murray, *Mathematical Biology*, Springer-Verlag (1989).
18. S. Peres and J. P. Comet. Contribution of Computational Tree Logic to Biological Regulatory Networks: Example from *Pseudomonas Aeruginosa*. In CMSB 2003, C. Priami (ed.), LNCS 2602, 47–56 (2003).
19. B. E. Shapiro, A. Levchenko and E. Mjolsness. Automatic model generation for signal transduction with applications to MAPK pathway. In *Foundations of Systems Biology*, H. Kitano (ed.), MIT Press (2002).
20. E. H. Snoussi and R. Thomas. Logical Identification of All Steady States : The Concept of Feedback Loop Characteristic States. *Bulletin of Mathematical Biology* 55, 973–991 (1993).
21. D. Thieffry, M. Colet and R. Thomas. Formalisation of Regulatory Networks : a Logical Method and Its Automation. *Math. Modelling and Sci. Computing* 2, 144–151 (1993).
22. R. Thomas and M. Kaufman. Multistationarity, the Basis of Cell Differentiation and Memory. II. Logical Analysis of Regulatory Networks in Term of Feedback Circuits. *Chaos*, 11, 180–195 (2001).