



HAL
open science

Default Logic for Diagnostic of Discrete Time System

Tan Le, Andrei Doncescu, Pierre Siegel

► **To cite this version:**

Tan Le, Andrei Doncescu, Pierre Siegel. Default Logic for Diagnostic of Discrete Time System. BWCCA 2013: 8th International Conference on Broadband and Wireless Computing, Communication and Applications, Oct 2013, Compiègne, France. 6p. hal-00992206

HAL Id: hal-00992206

<https://hal.science/hal-00992206>

Submitted on 19 May 2014

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

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

Default Logic for Diagnostic of Discrete Time System

Tan Le, Andrei Doncescu
Laboratory for Analysis and Architecture of System
CNRS, University of Toulouse
7 Avenue du Colonel Roche, Toulouse, France
e-mail: tle@laas.fr, andrei.doncescu@laas.fr

Pierre Siegel
LIF - UMR CNRS 7279, Aix-Marseille
39 rue Frédéric Joliot Curie 13453,
Marseille cedex 13, France.
e-mail: p.siegel@free.fr

Abstract—Signaling pathways are seen as high criticalities in our understanding of mechanisms of biological functions. In this paper, we propose default logic for diagnostic of Discrete Time System (DTS) by focusing on automatic synthesis of the signaling pathways from factors within the cell. An essential component of this approach is using default logic to acquire facts about biological knowledge of intracellular communication. By choosing an adequate representation of biological knowledge, the "reasoning" is able to assign in acquisition of the facts and extract interactions necessary for the synthesis of the signaling pathways.

I. INTRODUCTION

Today, increasingly specialized experts need an appropriate evaluation of their know-how corroborated with the available experimental data in order to discover new knowledge. This scientific approach apprehends the biology systems by hypotension and validation of it. In this context, in recent decades, biology has grown prolifically in all its facets. New fields of applications and studies such as biotechnology, nanotechnology and bioinformatics are immersed and take an important place in the context of current research with the goal of identifying, understanding and quantification of biological phenomena seen in within the biological system.

The System Biology is the research field, which handles biological complex system representation and analysis. The dynamic nature and non-linear inherence of biological course make the system identification difficult. By system identification we understand the model associated to biological system. The majority of kinetic models in biology could be described by a couple of differential equations [1]. It is almost impossible to find out an analytical solution to these non-linear systems [3]. The only feasible approach is a bottom-up analysis allowing the solving and simulating of the biological systems.

From the standpoint of Artificial Intelligence, cells are sources of information that include a myriad of intra and extra cellular signals that as the ultimate goal of optimal output describing proteins. Disease and cancer in particular can be seen as a pathological alteration in the signaling networks of the cell. The study of signaling events appears to be the key of biological, pharmacological and medical research. The spread of these types of signals are not

changing the behavior of proteins on three levels: regulation of the activity, interaction and expression. The three levels are synchronized in a strong momentum that leads to changes in protein activity. Since a decade signaling networks have been studied using analytical methods based on the recognition of proteins by specific antibodies. Parallel DNA chips (microarrays) are widely used to study the co-expression of candidate genes to explain the etymology of certain diseases, including cancer.

This huge amount of data allows the modeling of gene interactions. The biological experts look for evidence of interactions between proteins or genes. Therefore, the representation by graphs is the best way of understanding of biological systems. This representation includes mathematical properties as connectivity; presence of positive and negative loops which is related to a main property of genetic regulatory networks. Biochemical reactions are very often a series of time steps instead of one elementary action. Therefore, one of the research directions in system biology is to capture or to describe the series of steps called pathways by signaling engineering.

The study of gene networks poses problems well identified and studied in Artificial Intelligence over the last thirty years. In this article we present how the possibility to reason from incomplete, uncertain, revisable, contradictory and multiple sources. Indeed, the logical or mathematical description of signaling pathways is not complete: biological experiments provide a number of protein interactions but certainly not all of them. On the other hand, the conditions and sometimes the difficulties of the experiments involve these data are not always accurate. Some data may be very wrong and must be corrected or revised in the future. Finally, the information coming from different sources and experiences can be contradictory. It is the goal of different logics and particularly the non-monotonic logics to handle this kind of situations. Afterwards this interaction maps should be validates by biological experiments. Of course, these experiments are time consuming and expensive, but less than an exhaustive experiment.

The article goal is to analyze, understand and associate a logical model to biological systems. However, we want to

build-up a knowledge based-system, able to discover biological mechanisms.

The rest of this paper is organized as follows. Section II gives the problem from a biological point of view. Section III presents the causal relations between enzymatic activities. Section IV introduces the logical representation of these relations. Section V shows experimental results obtained by application of the default logic for estimation of possible reaction states from given observations. Section VI concludes this paper.

II. SIGNALING PATHWAY

In attempts to describe the behavior of living systems, where the deductive model is not successful, the process by qualitative reasoning based on the function of molecules has shown its limits [4]. Similarly, if the properties are known, we cannot clearly deduce their function in the living cell, and from the characteristics of living cells, calculate their behavior in a given environment. In general, the deductive approach fails because the functions of the living system components depend simultaneously of the interactions with other elements. The recurring problem of this reasoning is how the functional properties of the cell can be derived from properties of its components alone. In this context, with their ability to describe the complexity, the logic tools offer a perspective to analyze these structural elements organized in a complex network.

If the analytical models based on differential equations are impossible to solve and the multi-scale analysis seems utopic, we propose in this paper an elegant solution to find out the main signaling reaction [2].

While the proteins have different morphologies and structures and their roles in different organisms are different, their basic functionality is the same. One of these cell-based activities is to ensure their own division (production). Its activity can be wholly summarized in two points. First, proteins can promote an activity of others, which is called cause. Second, proteins can inhibit an activity of others, which is called block. Most of these reactions, which take place in a cell, are catalyzed by special molecules called enzymes. Such a large amount of data on the signaling is represented as a network, called a signaling pathway, and has been stored and maintained in a database on a large scale [4].

Fig. 1 gives a very simplified example of interactions in a cell. Through different mechanisms not shown here, ultraviolet (UV) is the cell apoptosis (it actually becomes immortal) from cancer. This is shown by an arrow. On the other hand, the UV activates the production of the protein p53. This protein will activate protein A which will block the cancer. But p53, binds to the protein Mdm2, will produce B, which will block A. For a biologist, the question is how to block the cancer by blocking B. Biological experiments have shown that X could be a candidate for this

block. Fig. 2 and Fig. 3 provide two types of interactions with X to explain blocking B [3].

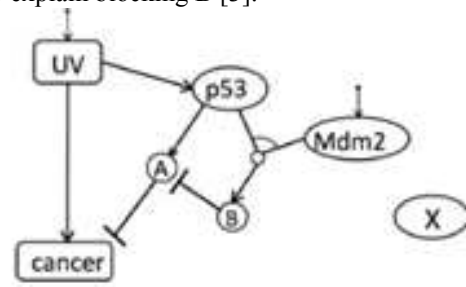


Figure 1. One example of interactions in a cell

The problem is, in the domain of discrete time, how the interactions like. For example, at a moment, t_0 , UV activates cancer and p53. At the next moment, t_1 , p53 will activate A. On the other hand, p53 will bind to the Mdm2 for active B, who will block A at the moment t_2 , etc. In this paper, we have completed the graph by hand, and by using the computer to do this, too.

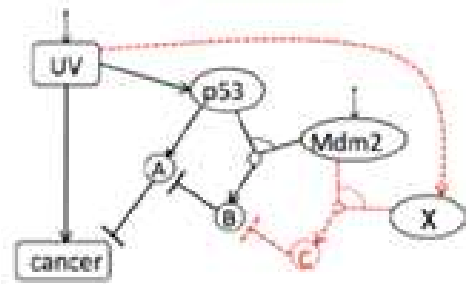


Figure 2. One possible solution

This example is a very basic case, of course. In practice this graphical representation of the network signaling pathways may involve several thousands of genes. But the problem is still not trivial because the discovery (abduction) will be firstly on the presence of protein C and secondly on the addition of links between proteins.

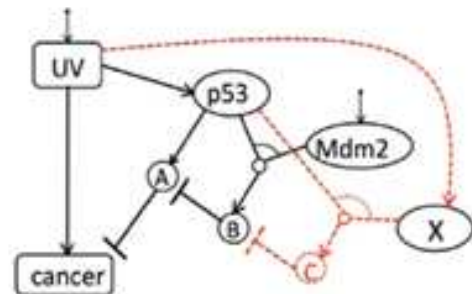


Figure 3. Another possible solution

III. LOGICAL MODEL OF SIGNALING

Genes and proteins are considered as a same object (genes produce proteins). We will often restrict here to a

propositional representation. In practice, the detailed study of interactions will require to represent increases or decreases in protein concentration. It therefore falls outside the scope propositional, but the basic problems are the same. To represent a change of concentration, for example, is possible to use predicates such as "increase" or "decrease" and to limit the use of these predicates.

To describe interactions between genes in the cell, we start from a classical logic language L (propositional or first order). In L, the proposition A (resp. $\neg A$) means that A is true (false). We can say, for example, *give(UV)* to say that the cell is subjected by ultraviolet, or even *glass-screen* $\rightarrow \neg$ *give(UV)* to say that a glass screen protects from ultraviolet rays. We are in a logical framework, so it is possible to represent almost everything you want in a natural way. The price to pay, this can be the combinatorial explosion of algorithms.

Interactions among genes are a very simple form of causality. To express these interactions, it is common to use two binary relations *cause(A, B)* and *block(A, B)*. The first relation means, for example, protein A initiates the production of protein B, while the second inhibits it. Conventionally, these relationships are represented in the network of genes by $A \rightarrow B$ and $A \dashv B$. Of course, this causality is basic and many works have been written for represent the causalities.

If the inference of classical logic $A \rightarrow B$ is formally described perfectly, the description of formal properties causation, is less straightforward. Causality cannot be seen as a logical classical relation. A basic example is that of the expression "If it rains, the grass is wet". This expression cannot be translated by the formula *rain* \rightarrow *grass-wet*, which will signifier that when it rains, the grass is wet automatically. Indeed, there may be exceptions to this rule (the grass is in a shed ...). We can also change the environment (we cover the grass). These revisable rules and exceptions are well known in Artificial Intelligence. They create, in particular, non-monotonic logics and theories of review. On the other hand, and more technical, we find here all the classical problems that arise when we want to try to formalize and use negation by failure in programming languages such as Prolog or Solar.

To give the links between our causal relations *cause* and *block*, in a classical language (propositional calculus or first order logic), it must therefore do two things:

- describe the internal properties in relations *cause* and *block*
- describe the links between these relations and the classical logic

All this take into account the problem of uncertain and revisable. For the first aspect we will explicitly give the links minimum and necessary between two causal relations. Links with classical logic will be described in a first time in default logic. Then, to take into account the aspect of discovery (abduction, field of production) we will use the default logic.

In our context, to give these links between the relations *cause* and *block*, we will more simple, use classical logic. The basic solution is then to explicitly give two schemes of axioms:

$$(C1) \text{ cause}(A,B) \wedge \text{ cause}(B, C) \rightarrow \text{ cause}(A, C)$$

$$(C2) \text{ cause}(A, B) \wedge \text{ block}(B, C) \rightarrow \text{ block}(A, C)$$

We believe that it is axiomatic minimum system necessary and probably sufficient for the application to the cell. At the moment, there is no formal link between two relations. It is of course possible to add other axioms to take into account these links.

IV. LOGICAL REPRESENTATION

In a first approach, the first properties that we want to give can be expressed naturally, by the rules of the type:

(1) If A *cause* B and if A is *true*, so B is *true*.

(2) If A *block* B and if A is *true*, so B is *false*.

Depending on the context, the *true* can be called the *known*, *certain* ... or, more technically, in a demonstration automatically system *proved*. The first idea is to express these laws in classical logic by axioms:

$$\text{ cause}(A, B) \wedge A \rightarrow B$$

$$\text{ block}(A, B) \wedge A \rightarrow \neg B$$

We can also express more weakly the laws by inference rules of modus ponens nearby:

$$\text{ cause}(A, B) \wedge A / B$$

$$\text{ block}(A, B) \wedge A / \neg B$$

But these two formulations are problematic because there is a conflict. For example, if we have a set of four formulas $F = \{A, B, \text{ cause}(A, C), \text{ block}(B, C)\}$, it goes to both approaches above data inferred from F, C and $\neg C$, which is inconsistent. To resolve this type of conflict, we can try to use methods inspired by constraint programming, such as the use of negation by failure. It is also possible to use a revisable reasoning, especially a non-monotonic logic. The first approach poses many conceptual and technical problems if you leave the simple cases. These problems are often solved by adding properties to the formal system, that pose other problems ... and we arrive at a beautiful gas plant. We will study here a non-monotonic approach. At first, we use the default logic.

V. INTERACTION AND DEFAULT LOGIC

Default logic formalizes reasoning by default. It allows treating the rules by admitting exceptions without having to challenge the rules previously established whenever a new exception appears. A default theory consists of a set of facts W, which are formulas of propositional calculus from either of the first order logic, and a set of defaults D, which are rules of inference to specific contents. Defaults are used to manage incomplete information. In its most general form, a default is an expression of the form:

$$D = \frac{A_x(X):B_y(X) \wedge C(X)}{C(X)} \quad (1)$$

where $A_x(X)$, $B_y(X)$ and $C(X)$ ($x = 1, 2, \dots, m$, $y = 1, 2, \dots, l$) are well-formed formulas which contain first order as free variable X or $X = (x_1, x_2, x_3, \dots, x_n)$ as a vector of free variables. $A_x(X)$ are the prerequisites, $B_y(X)$ are the justifications and $C(X)$ is the consequent. The default equation (1) means informally: if $A_x(X)$ are verified, if it is possible that $B_y(X)$ are real ($B_y(X)$ are consistent), and if it is possible that $C(X)$ is true, then we infer $C(X)$.

The use of defaults increases the number of formulas derived from the knowledge base W : we get extensions that are sets of theorems derivable monotonically. An extension of the default theory $\Delta = (D, W)$ is a set E of formulas, closed for the deduction, containing W and satisfying the following property: if d is a default of D whose prerequisites $A_x(X)$ are in, the negation of justifications $B_y(X)$ and of consequent $C(X)$ are not in E , then the consequent of d is in E . Formally, the extensions are defined as follows:

E is an extension of Δ if and only if

$$(A1): E = \bigcup_{i=0, \infty} E_i,$$

with

$$(A2): E_0 = W,$$

and

$$(A3): E_i = Th(E_{i-1}) \cup \{C(X) / (\frac{A_x(X):B_y(X) \wedge C(X)}{C(X)}) \in D, A_x(X) \in E_{i-1}, \neg B_y \notin E_{i-1}, \neg C(X) \notin E_{i-1}, i > 0\},$$

where

$$(A4): Th(E_i) = \{w/E_i \vdash w\}.$$

The calculation of extensions allows to study the defaults one by one and to retain those who respond to the problem and are compatible with each other. Each extension corresponds to a possible solution of the problem. To calculate an extension, we must verify that the negation of justification does not belong to E_i . We can therefore use an incremental algorithm for computing extensions. For a default theory $\Delta = (D, W)$, with the set of defaults D and the knowledge base W , the calculation is extended according to the algorithm:

Input : $E = \emptyset$; (set of extensions E is empty).

Output : $E = \bigcup_{i=0, N} E_i$.

calcul_extension(E) :

{

(1) **while** there is a default $D = \frac{A_x(X):B_y(X) \wedge C(X)}{C(X)}$

that has not yet been inspected **do**

(2) Select the default D ,

(3) Verify that the prerequisites $A_x(X)$ are true,

(4) Verify that the justifications $B_y(X)$ are consistent with W ,

(5) Verify that the consequent $C(X)$ is consistent with W ,

(6) Add $B_y(X)$ and $C(X)$ to W .

(7) **end while**

(8) End of the calculation for an extension.

(9) Backtracking (deleting the last $C(X)$ and $B_y(X)$ added to W).

(10) calcul_extension(E).

In our example, to provide links between these relationships *cause* and *block*, the intuitive idea is to weaken the formulation of causation rules:

(1) If *cause*(A, B) is *true* (A causes B), and if A is *true*, and if it is *possible* that B , then B is *true*

(2) If *block*(A, B) is *true* (A blocks B), and if A is *true*, and if it is *possible* that B is *false*, then B is *false*

The question is then formally described by *possible*. We use here a non-monotonic logic of the best known, default logic. In this logic, the rules (1) and (2) will be expressed intuitively as:

(1') If *cause*(A, B) is *true*, and if A is *true*, and if B is not *contradictory*, then B is *true*

(2') If *block*(A, B) is *true*, and if A is *true*, and if $\neg B$ is not *contradictory*, then $\neg B$ is *true*

In the default logic, these rules will be represented by the set of defaults D and be written:

$$d1: \frac{cause(uv, cancer) \wedge uv: cancer}{cancer}$$

$$d2: \frac{cause(uv, p53) \wedge uv: p53}{p53}$$

$$d3: \frac{cause(p53, a) \wedge p53: a}{a}$$

$$d4: \frac{cause(joint(p53,mdm2),b) \wedge p53 \wedge mdm2: b}{b}$$

$$d5: \frac{block(b,a) \wedge b: \neg a}{\neg a}$$

$$d6: \frac{block(a,cancer) \wedge a: \neg cancer}{\neg cancer}$$

$$d7: \frac{block(c,b) \wedge c: \neg b}{\neg b}$$

We have also two general defaults, where $X \in \{uv, p53, mdm2, a, b\}$:

$$d8: \frac{cause(X,x) \wedge X: x}{x}$$

$$d9: \frac{cause(joint(X,x),c) \wedge X \wedge x: c}{c}$$

The conflict was resolved.

With the default theory $\Delta = (D, W)$, in which $D = \{d1, d2, d3, d4, d5, d6, d7, d8, d9\}$ and $W = \{uv, mdm2\}$, by applying the algorithm above, we have 18 extensions. The next are two could be easily interpreted that correspond to Fig. 2 and Fig. 3, respectively:

$$E1 = \begin{cases} uv \rightarrow p53 \\ p53 \rightarrow a \\ joint(p53,mdm2) \rightarrow b \\ uv \rightarrow x \\ joint(uv,x) \rightarrow c \\ a \rightarrow \neg cancer \end{cases}$$

and

$$E2 = \begin{cases} uv \rightarrow p53 \\ p53 \rightarrow a \\ joint(p53,mdm2) \rightarrow b \\ uv \rightarrow x \\ joint(mdm2,x) \rightarrow c \\ a \rightarrow \neg cancer \end{cases}$$

This algorithm is almost NP - problem. Of course the real problem of signaling pathways is more complicated than the model presented in this paper. The practical complexity can then be accessible by controlling the numbers of defaults.

If it is necessary to know, which molecule (a future drug), acts effectively we could represent this problem in a context of abductive reasoning. In simple terms, abduction find out the "minimum" information set added to a known facts F, able to deduce a result R which we would like to prove.

VI. CONCLUSION

We presented a first signaling pathways model using default logic to represent, complete and to find out the main reaction. All this was done using the reasoning with default assumption. We show-up all possible main reactions in the case of a simple model, the next work will be devoted to real signaling maps downloaded from KEGG.

ACKNOWLEDGMENT

We deeply thank to D.R. Jean-Charles Faye and C.R. Olivier Sordet of Claudius Regaud Cancer Institute (ICR). Moreover, we are particularly grateful to Vietnamese and French Government to finance this work.

REFERENCES

- [1] Demongeot J, "Multi-stationarity and cell differentiation", J. Biol. Systems., 6, 1-2 (1998).
- [2] Doncescu A., Inoue K. and Yamamoto, "Knowledge-based discovery in systems biology using CF-induction". New Trends in Applied Artificial Intelligence: Proceedings of the 20th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA / AIE 2007), Lecture Notes in Artificial Intelligence, volume 4570, pages 395-404, Springer-Verlag.
- [3] Doncescu A, and Siegel P, "Utilisation de la logique des hypothèses pour la modélisation des voies de signalisation dans la cellule", JIAF 11, Lyon 8-10, June 2011.
- [4] Doncescu A., Waissman J., Richard G., Roux G. "Characterization of bio-chemical signals by inductive logic programming", Knowledge-Based Systems 15 (1), 129-137, 2002.
- [5] Christophe Chassagnole, Juan Carlos, A Rodriguez, Andrei Doncescu, Laurence T Yang "Differential evolutionary algorithms for in vivo dynamic analysis of glycolysis and pentose phosphate pathway in Escherichia Coli", Parallel Computing for Bioinformatics and Computational Biology: Models, Enabling Technologies, and Case Studies, 59-78, John Wiley & Sons, Inc., 2006.
- [6] Montseny E., Doncescu A., "Operatorial Parametrizing of Controlled Dynamic Systems-Application to the Fed-Batch Bioreactor Control Problem", 17th World Congress The International Federation of Automatic Control. Seoul, Korea, June 2008.
- [7] Forget L, Rish V., P Siegel. "Preferential Logics are X-logics" Journal of Computational Logic, 10, 2000, pp. 1-13.
- [8] Ginsberg, ML, Smith, DE (July 1988). "Reasoning about action II: the qualification problem". Artificial Intelligence Vol. 35 No. 3 pp.311-342.
- [9] Giunchiglia, E., Lee, J., Lifschitz, V., McCain, N., & Turner, H. (March 2004). "Nonmonotonic causal theories". Artificial Intelligence No. 1-2 vol.153 pp.49-104.
- [10] Inoue K, "Induction as Consequence Finding". Machine Learning, 55 (2) :109-135, 2004.
- [11] Inoue K, Saito H. "Circumscripta Policies for Induction" Proceedings of 14th Conf. on Inductive Logic Programming, LNAI 3194, pp.164-179, Springer, September 2004.
- [12] Kayser D., Levy F. "Modeling symbolic causal reasoning", Intellecta 2004 / 1, 38, pp 291-232
- [13] Nabeshima H., Iwanuma K., Inoue K. Ray O. "SOLAR: An automated deduction system for Finding consequence". AI Commun. 23 (2-3): 183-203 (2010)
- [14] Roux-Rouquié M., L. Hood, Imbeaud S., Auffray C. "Issues in Computational Methods for Functional Genomics and Systems Biology". CMSB 2003 : 182-186
- [15] Schwind P., Siegel P: "Modal Logic for Hypothesis Theory", Fundamentae Informaticae, cal 21, No. 1-2 89-101.
- [16] Synnaeve G, Inoue K, Doncescu A, Nabeshima N, Kameya Y, Ishihata M., Sato T, "Kinetic models and qualitative abstraction for relational learning in systems biology", BIOSTEC Bioinformatics 2011
- [17] Siegel P. : "A modal language for Nonmonotonic Reasoning", Proc. Workshop DRUMS / EEC Marseille 24-27 February 90.
- [18] P. Siegel, C. Schwind (93) "Modal logic based theory for nonmonotonic reasoning". Journal of Applied Non-classical Logic, Volume 3 - No. 1 / 1993, P 73-92.

- [19] Synnaeve G., Doncescu A., Inoue K., “Kinetic models for logic-based hypothesis finding in metabolic pathways”, Int’l Conf. on Inductive Logic Programming (ILP-09), 2009.
- [20] Tran N. , C. Baral (2007) “Hypothesizing and reasoning about signaling networks”. Journal of Applied Logic 7 (2009) 253-274