**MODELING SIGNAL PATHWAYS.**

BENTOLILA SIMONE

IGM, university Marne la Vallee - 2 rue de la Butte Verte - 93166 Noisy le Grand Cedex - FRANCE

e-mail:tolila@genethon.fr

Keywords: formal language, grammar, automata, cell signaling, transduction pathway

The integration of the various types of cellular activities in a multicellular organism is mainly performed by the nervous system, the endocrine system and the immune system.

In fact although DNA is indispensable for the life of the cell, outside of the context of the living cell and the intercellular communications network, DNA is basically inert. Memory exists at 2 levels: the memory of the species which consists of the unchangeable DNA on one hand, and the active memory or short-term memory of each cell, which is its metabolic state at each instant. Enzymes constitute the short-term memory of the cell, its identity, and the network of indicators of what is going to happen. It is the same for the intercellular communications network.

In fact DNA is indispensable to the life of the cell: during apoptosis, or programmed cell death, in which endonuclease enzymes which digest the cell's own DNA are activated, thus rendering the DNA unusable. DNA is necessary for life, it is a matrix which is read and interpreted by the cell according to its own identity, its own biochemical context and its environment. The identity of a differentiated cell is maintained by its metabolism; a cell which loses control of its regulation dedifferentiates and loses its identity. Enzymes activate or inhibit the metabolic and differentiation pathways in which the cell may engage (Figure 1): some enzymes regulate the expression of genes and the synthesis of proteins from the DNA template (both structural proteins and the enzymes themselves); while others enzymes catalyse metabolic reactions, anabolic, in which products required by the cell are synthesized, and catabolic in which substrates are degraded to elementary subunits with production of energy.

Intercellular communications are mainly conducted by secreted proteins (ex: hormone, growth factors, cytokines, antibodies) and exogenous ligands (ex: antigen). These circulating substances transmit a signal to competent cells using trans-receptor proteins as intermediaries; the signal is then relayed to the interior of the cell by transduction by an enzymes cascade. The signal may be transmitted to the nuclear by transcription factors which provoke the expression or repression of a gene, or to the cytoplasm where a metabolic pathway may be activated or inactivated (Figure 1). Without the enzymes to promote a given pathway, reactions would occur, but would proced so slowly that the products of of a given reaction might be degraded before they could serve as substrates for the next reaction of the pathway. The progression of enzymes that serve as catalysts for a metabolic pathway form a code which switch on or off, these enzymes form the code for the metabolic pathway or word of the language.

Molecules interact by contact and chemical interactions, a binding between two molecules may produce activation or inhibition of the catalytic site of one of the two molecules. This usually involves allosteric alteration or covalent modification by phosphorilation. An enzyme can be described by these 2 sites: the catalytic site and the allosteric site. An inhibitor may bind to the catalytic site and block it (isosteric modification, inhibition of the catalytic site by an analog of the substrate) or to the allosteric site and cause a change in conformation which will activate the catalytic site. An enzyme which catalyse covalent binding (interconversion), usually by phosphorilation or dephosphorilation, of another enzyme may cause activation or inactivation of the latter.

In a previous paper (Bentolila, 96) we described a context sensitive grammar which models the 4 main types of genes regulation. The proposed model considers two types of objects: transcriptional units on DNA and regulatory or structural proteins which are synthesized, and which are, in the case of regulatory proteins, themselves destined to activate or repress other transcriptional units in a later phase. A transcriptional unit is described by the list of its active sites (operator, promoter, binding sites for transcription factors). A regulatory protein is described by the list of its active sites (binding domain, activation domain, binding domain for ligand). The DNA sites and the protein domains are the terminal symbols of the proposed grammar. The interaction of these proteins with the DNA, and in certain cases preliminary interactions between proteins, leads to one of two antagonistic actions: expression or repression of the transcriptional unit. These protein-protein and protein-DNA interactions are grouped into syntactic categories (induction, inhibition, initiation complex, repressor complex, activation complex) which are called biological binding operators. The expression/repression actions are described by grammar rules which provide the chain of execution by biological binding operators for the four activable/repressible regulatory systems modulated by positive/negative co-factors.

If we suppose that the semantics of biological binding operators is already implemented (using a database), it is sufficient to write a context-free grammar which describes the order of application of biological binding operators, similar to the context-free grammar of arithmetic operators, for example. In the case of arithmetic operators, the semantics of the operations is recognized and implemented in the compiler, whereas in the case of biological binding operators, the semantics, i.e. the result of the operation cannot be given by computation; a table of meaningful bindings and their result is necessary as an extension.

We have extended this model to intercellular communication pathways or signal pathway. The grammar that we have developped describes the series of operations that leads to either the activation or inactivation of an anabolic / catabolic pathway (ex: glycogensis / glycogenolysis) or the expression / repression of a protein ( ex: antibody, ccytokines). We have applied this model to 2 examples: the key enzymes involved in sugar metabolism regulation in the liver (Table 1) which is under hormonal control; and a simplified model of the immune response (Table 2).

**An sql simulator based on a relational database**

The simulation models the binding of 2 molecules either in the same cell, or a receptor and a circulating molecule or 2 circulating molecules (or protein complexe).

Some operators are not detailed, and reference is made to a sub-automaton such as the sub-automaton "transduction" which cannot be detailed in some cases because all of the binding elements of the pathway are not yet known; or the sub-automata Expression / Represssion which was previously described (Bentolila, 96).

One cell refers to a population of cells.

The object of this modelization is the observation of a cell in a given state for a given process.

**The knolewdge database** which describes the process studied, contains the list of binding operators (1st operand, 2nd operand, operator, result), each operand described by (cell, molecule, type, state) (Table 1 and 2).

**The current state of the simulation**contains the initial state of the simulation: the list of cells described by the list of their receptors and their cytoplasmic proteins, as well as the list of circulating molecules.

**In each cycle, the simulation considers** the licit bindings based on the current state and applies the resulting actions to the current state, such as: addition of a circulating / cytoplasmic / receptor protein (expression of a gene); change in state of a protein (activation / inactivation); addition of a cell (cell division); suppression of a cell (cell destruction).

The simulation uses sql to provides a better simulation of parallelism.

Figure 1 A modeling of signal pathway

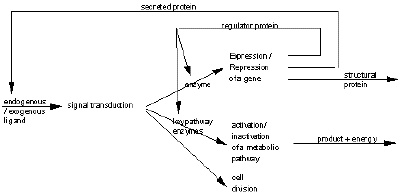


Figure 2 Key enzymes in regulation of sugar in the liver

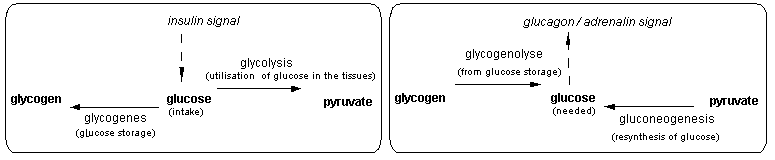


Table 1 Signal pathway in the regulation of sugar metabolism in the liver

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1st operand** | | | | **2nd operand** | | | | **Operator** | **Result** | |
| Cell | molecule | type | Cell | molecule | type |  | Cell | molecule | type | state |
|  |  |  |  | glucose |  | Expression |  | insulin | circulating |  |
| liver | insulin receptor | receptor |  | insulin | circulating | transduction | liver | glycogen-synthase | cytoplasmic | activated |
| liver | glycogen-synthase |  | liver |  |  | metabolic pathway | liver | glycogen | cytoplasmic |  |
| liver | insulin receptor | receptor |  | insulin | circulating | transduction | liver | glycolysis  key enzymes | cytoplasmic | activated |
| liver | glycolysis  key enzymes |  | liver |  |  | metabolic pathway | liver | pyruvate | cytoplasmic |  |
| liver | insulin receptor | receptor |  | insulin | circulating | transduction | liver | gluconeogenesis key enzymes | cytoplasmic | inactivated |
| liver | insulin receptor | receptor |  | insulin | circulating | transduction | liver | glycogen-phosphorylase | cytoplasmic | inactivated |
|  |  |  |  |  |  | Expression |  | glucagon | circulating |  |
| liver | glucagon receptor | receptor |  | glucagon | circulating | transduction | liver | gluconeogenesis key enzymes | cytoplasmic | activated |
| liver | gluconeogenesis key enzymes |  | liver |  |  | metabolic pathway | liver | glucose | circulating |  |
| liver | glucagon receptor | receptor |  | glucagon | circulating | transduction | liver | glycogen-phosphorylase | cytoplasmiqc | activated |
| liver | glycogen-phosphorylase |  | liver |  |  | metabolic pathway | liver | glucose | circulating |  |
| liver | glucagon receptor | receptor |  | glucagon | circulating | transduction | liver | glycolysis  key enzymes | cytoplasmic | inactivated |
| liver | glucagon receptor | receptor |  | glucagon | circulating | transduction | liver | glycogen-synthase | cytoplasmic | inactivated |

Detail

|  |  |
| --- | --- |
| glycogense key enzymes | glycogen synthase |
| glycogenolyse key enzymes | glycogen phosphorylase |
| gluconeogenesis key enzymes | pyruvate carboxylase |
|  | PEP carboxykinase |
|  | fructose1,6 bisphosphatase |
|  | glucose 6 phosphatase |
| glycolysis key enzymes | hexokinase |
|  | 6-phosphofructokinase |
|  | pyruvate kinase |

Detail of glucagon transduction pathway

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1st operand | | | | 2nd operand | | | | Operator | Result | |
| Cell | molecule | type | Cell | molecule | type |  | Cell | molecule | type | state |
| liver | glucagon | circulating | liver | glucagon receptor | receptor | binding | liver | glucagon receptor | receptor | activated |
| liver | glucagon receptor | receptor | liver | G protein | cytoplasmic | binding | liver | G protein | cytoplasmic | activated |
| liver | G protein | cytoplasmic | liver | adenylate cyclase | cytoplasmic | binding | liver | adenylate cyclase | cytoplasmic | activated |
| liver | adenylate cyclase | cytoplasmic | liver | ATP | cytoplasmic | binding | liver | AMPc | cytoplasmic | activated |
| liver | AMPc | cytoplasmic | liver | protein kinase A | cytoplasmic | binding | liver | protein kinase A | cytoplasmic | activated |
| liver | protein kinase A | cytoplasmic | liver | phosphorylase-kinase | cytoplasmic | binding | liver | phosphorylase-kinase | cytoplasmic | activated |
| liver | phosphorylase-kinase | cytoplasmic | liver | glycogen phosphorylase | cytoplasmic | binding | liver | glycogen phosphorylase | cytoplasmic | activated |
| liver | phosphorylase-kinase | cytoplasmic | liver | glycogen synthase | cytoplasmic | binding | liver | glycogen synthase | cytoplasmic | inactivated |

Table 2 Simplified model of the immune response

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1st operand | | | | 2nd operand | | | | Operator | Result |
| Cell | molecule | type | Cell | molecule | type | Act | Cell | molecule | type |
| Target cell | cell receptor | receptor |  | antigen | circulating | Expression | infected cell | MHCI + antigen | receptor |
| Target cell | cell receptor | receptor |  | antigen | circulating |  | infected cell | antigen | receptor |
| B | antibody Ig | receptor |  | antigen | circulating | Expression | B | MHCII + antigen | receptor |
| Th | T receptor | receptor | B | MHCII + antigen | receptor | Expression | Th | cytokine | circulating |
| B | cytokinereceptor | receptor |  | cytokine | circulating | Cell division | B |  |  |
| B | cytokinereceptor | receptor |  | cytokine | circulating | Expression | B | antibody | circulating |
|  | antigen | circulating |  | antibody | circulating |  |  | antigen + antibody | circulating |
| infected cell | antigen | circulating |  | antibody | circulating |  | infected cell | antigen + antibody | receptor |
| macrophage | M receptor | receptor |  | antigen | circulating | Expression | macrophage | MHCII + antigen | receptor |
| macrophage | Fc receptor | receptor |  | antigen + antibody | circulating | destruction antigen + antibody | macrophage |  |  |
| Th | T receptor | receptor | macrophage | MHCII + antigen | receptor | Expression | Th | cytokine | circulating |
| Th | cytokine receptor | receptor |  | cytokine | circulating | Cell division |  |  |  |
| Tc | cytokine receptor | receptor |  | cytokine | circulating | Cell division |  |  |  |
| Tc | T receptor | receptor | infected cell | MHCI + antigen | receptor | Cell destruction |  |  |  |
| K | Fc receptor | receptor | infected cell | antigen + antibody | receptor | Cell destruction |  |  |  |

**References**

1. Atlan H., (1990) The cellular computer DNA: program or data *Bull. math; Biol.*52-3,335-348
2. Bentolila S., (1996) A grammar describing "biological binding operators " to model gene regulation. *Biochimie* 78, 335-350
3. Chomsky N, (1957) Syntactic Structures. *Mouton*
4. Collado-Vides J, (1991) The Search for a Grammatical Theory of Gene Regulation Is Formally Justified by Showing the Inadequacy of Context-free Grammars. *Comput* *Applic Biosci*7(3), 321-326
5. Collado-Vides J, (1991) A syntactic Representation of Units of genetic information.*J Theor Biol* 148, 401-429
6. Collado-Vides J, (1993) A linguistic representation of the regulation of transcription initiation. *BioSystems*29, 87-128
7. Hofestadt R, (1993) A simulation shell to model metabolicpathways. *J Syst Analysis Modeling Simulation*11, 253-262
8. Hofestadt R, Meineke F (1995) Interactive modelling and simulation of Biochemical networks. *Comput Biol Med* 25(3), 321-334
9. Hopcroft J E, Ullman J D (1979) Introduction to automata theory, languages and computation. *Addison Wesley*
10. Searls D B, Dong S (1993) in *Proceedings of the 2nd International Conference on Bioinformatics, Supercomputing, and Complex Genome Analysis* (Lim H A et al, eds) World Scientific, 89-101
11. Trifonov E N, (1993) DNA as a language in *Proceedings of the 2nd International Conferencce of Bioinformatics, Supercomputing, and Complex Genome Analysis*(Lim H A et al, eds) World Scientific, 103-110