

**Two operative concepts for the post-genomic era :  
the "mémoire vive" of the cell  
and a molecular algebra**

Simone Bentolila

IGM, University Marne la Vallée - Cité Descartes, 5 Bd Descartes, Champs sur Marne - 77454 Marne la Vallée  
Cedex 2 - FRANCE, *e-mail* : [tolila@infobiogen.fr](mailto:tolila@infobiogen.fr)

The first successes in cloning experiments and stem cell "reprogramming" have already demonstrated the primordial role of cellular working-space memory and regulatory mechanisms, which use the knowledge stored in the DNA database in read mode.

We present an analogy between living systems and informatics systems by considering : 1) the cell cytoplasm as a memory device accessible as read/write; 2/ the mechanisms of regulation as a programming language defined by a grammar, a molecular algebra; 3) biological processes as volatile programs which are executed without being written; 4) DNA as a database in read only mode. We also present applications to two biological algorithms : the immune response and glycogen metabolism.

**Cell signaling, Electronic cell, Gene regulation mechanisms, Formal language,  
Grammar**

## **I - Introduction**

### **1/ Biology and Informatics, two main streams**

It is noteworthy that Turing (Turing 36) and Von Neumann (Von Neumann 45) were interested in human brain function in order to construct the principles of computer architecture by analogy, at a moment in history in which there was great progress in both the electronic and the cognitive sciences. Considerable theoretical progress took place due to development of informatics, in terms of logic, theory of languages and calculability theory. It seems only fair that these theories can, in turn, serve to model the behavior of the cell in terms of the information flow it receives and generates.

In 1953, at about the same time that Von Neumann was developing the bases for modern computer science, Watson and Crick discovered the double helical structure of DNA, and deduced its information storage capacity, using X-ray diffraction data produced by Rosalind Franklin. Shortly thereafter, in 1957, Chomsky (Chomsky 57) was developing a classification of formal grammars which would be widely used in the conception of programming languages, and in the improvement of compilers (Hopcroft & Ullman 79). Meanwhile, in 1961, Jacob and Monod (Jacob & Monod 61) discovered the logic of gene regulation mechanisms during their studies of the lactose operon.

### **2/ Analogies between living systems and informatics systems**

The architecture of informatics systems is based on two main concepts: memory and programming languages. The analogy that will be developed between living systems and informatics systems is based on these concepts. The cell working-space memory is the physico-chemical space in which the processes take place. The elements of the active memory or main memory are: circulating ligands, membrane receptors, cytosol molecules and transcription factors in the nucleus, which are capable of catalytic and regulatory action. They can be seen as a set of variables and positioned pointers in the volatile memory of the cell. These elements of cell memory do not pre-exist to their meaning, i.e. to the variable that they represent. However, these variables may then represent several values during the process, for instance activated or inactivated. These variables may be endogenous or exogenous, local or global.

In fact, the processes are observable when they are taking place in the intra- and inter-cellular spaces which play the role of active memory due to their capacity to maintain and transmit information. We are attempting to learn the underlying language for this program, including syntax, semantics and alphabet. This molecular algebra was first developed to describe the series of operations: "Biological Binding Operators" (Bentolila 1996) that leads to the expression / repression of a gene in a cell. In fact the two main mechanisms are Repressible Systems and Activable Systems, they are symmetric: the repressor or the activator is currently expressed, and a co-factor acts as an external signal who turn off or on the mechanism. This grammar was applied to 4 examples: i) Lactose operon, an instance of a repressible inducible system, moreover a repressible system may be activated. ii) The regulation of metallothionein, an instance of an alternative activable system. iii) Yeast galactose metabolism, a combination of repressible and activable mechanisms. iv) Tryptophan operon, an instance of a repressible system. This grammar has now been expanded in two directions of signal pathway: i) to the chain reaction which leads to the transmission of a signal destined to activate (or inactivate) a metabolic pathway for the delivery of necessary substances; ii) to the integration of the various types of inter-cellular activities in a multicellular organism which is mainly performed by the nervous system, the endocrine system and the immune system. This first grammar was context-sensitive because it requires a Turing Machine which is capable of reading and writing symbols. If we suppose that the semantics of biological binding operators is already implemented (using a database language or any other functional implementation of the compiler), it is sufficient to write a context-free

grammar. Therefore another extension of the present paper is the description of the whole machine including a memory device. Indeed, the implementation of the notion of variable as utilized by the logic of predicates and context-free grammar implies elements of memory.

### **3/ Data, program and engine**

Since the discovery of DNA and the genetic code and the identification of a "genetic program" of species, the questions remain: Where is the program? Is DNA the program or the data? What is the engine? Are proteins only the data of the program, or are they the elements of the machine itself?

H. Atlan considers the classical metaphor of the genetic program (Atlan 90, 98). He asserts that the program is located at the cellular level and consists of cellular automata networks and biochemical networks. His goal is to find models for the emergence of structures and functions in the course of evolution.

A. Danchin (Danchin 98) favors the model of the Turing machine and the hypothesis that the program, DNA, contains a logical description of the machine; he considers that a precise understanding of whole genomes will lead to the conclusion that the plan for the cell is in the chromosome.

Section (II) describes the proposed model: 1/ Cell main memory or working-space memory, 2/ DNA database as storage memory, 3/ Regulation mechanisms as a molecular algebra (language : syntax, semantics and alphabet), 4/ A minimal set of genes as the operating system, 5/ Biochemicals as hardware, 6/ Attempt at reprogramming life 7/ Input-output cell management.

## **II - The proposed model**

### **1/ Cell main memory or working-space memory**

The elements of the active memory or main memory at the level of the organism are : circulating ligands, membrane receptors, cytosol molecules (second messengers and molecules involved in signal transduction and enzymes which catalyze metabolic reactions) and transcription factors in the nucleus. The signal is continuous throughout the organism. Intercellular communications are mainly conducted by secreted proteins (eg: hormones, growth factors, cytokines, antibodies) and exogenous ligands (eg: antigen, glucose). These circulating substances transmit a signal to competent cells using trans-membrane proteins as intermediaries; the signal is then relayed to the interior of the cell by transduction by an enzyme cascade (Falke et al 97, Lichstein, Atlan 90, Artavanis-Tsakonas, Matsuno, Fortini 95). The signal may be transmitted to the nucleus by transcription factors which provoke the expression of a gene, or to the cytoplasm where a metabolic pathway may be activated. Without the enzymes to promote a given pathway, reactions would occur, but would proceed so slowly that the products of a given reaction might be degraded before they could serve as substrates for the next reaction of the pathway. The first level of regulation for signal pathways is the transcriptional control and the second is the activation / inactivation of enzymes involved in a cascade or modulated by a ligand; in some cases transcriptional control is a retrocontrol, inhibited by the terminal product of the chain of reactions or by another metabolite. These regulatory mechanisms are extremely and directly sensitive to the exterior environment and are modulated by exogenous ligands, such that the process is data driven.

The Genexpress system (Kolchanov 98, 99) aims to integrate knowledge about expression from various sources compiled in several databases: transcription regulatory regions and expression patterns (TRRD), classification of composite regulatory elements (COMPEL), activities of functional sites and their equilibrium constants (ACTIVITY), physico-chemical properties of sites (B-DNA-VIDEO) and ensembles of interacting genes and their transduction pathway (GeneNet).

R. Hofestadt (1995, 1998) was the first to integrate in the same model two types of regulatory proteins : transcription factors and metabolic enzymes. He has concentrated his study on the flow of materials in the cell; his simulator permits the detection of bottlenecks in a metabolic pathway when there is either a deficiency or an excess of substrate, due to genetic defects.

#### *Modes of activation of an enzyme.*

Molecules interact by contact and chemical interactions; binding between two molecules may produce activation or inhibition of one of the two molecules. This usually involves isosteric modification (for example: inhibition of the catalytic site by an analog of the substrate), allosteric alteration (causing a change in conformation which will activate another site) or covalent modification, essentially by phosphorylation (interconversion).

#### *Modes of activation of a membrane receptor*

Transmembranous transmission of the signal leads, in one or more steps, to the activation of one or more intracellular protein kinases. The receptors may be protein kinases themselves; examples include receptors for growth factors or insulin. Activation of catalytic receptors may lead to activation of a second messenger, which stimulates the activity of a protein kinase; or receptors may be coupled to a protein kinase, such as cytokine receptors; or there may be receptors with 7 transmembrane domains for which the signal transduction mechanism involves a G protein (the most common); or there may be receptors which are coupled to an ion channel.

#### *Secreted proteins and circulating ligands*

Hormones, growth factors and cytokines are secreted into the general circulation; they attain their target cells by this route. They have high affinity and high specificity for their receptors. Protein hormones as well as hydrophilic mediators of small size cannot traverse membranes by themselves; their receptors are found on the surface of their target cells. Other hormones, such as steroid hormones or thyroid hormones which are hydrophobic, traverse the plasma membrane and bind to their receptors in the cytoplasm of their target cells.

#### *Pheromones*

Pheromones are excreted directly into the air or in a fluid emitted by the organism; the reception of these olfactory chemical messengers occurs at the level of the nasal passages in mammals or the antennas in insects.

### **2/ DNA database as storage memory**

In fact, although DNA is indispensable for the life of the cell, outside of the context of the living cell, DNA is a dead letter. During apoptosis, or programmed cell death, endonuclease enzymes which digest the cell's own DNA are activated, thus rendering the DNA unusable. DNA is a database which is read and interpreted by the cell according to its own identity, its own biochemical context and its environment. The DNA database is a structured database; its grammar is now known. The usefulness of a grammar is now amply acknowledged; the deciphering of the DNA text to identify transcription units and their structure (ie: the motifs of the binding sites for both general and specific transcription factors) by the utilization of languages leads to a global approach which aims to understand the complete structure of the gene — not only the coding region but also the upstream regulatory region and the downstream termination signals — as a "sentence" which must be syntactically correct. This approach has been promoted by the extensive work of D Searls (1993, 1997, 2001) and J Collado-Vides (Collado-Vides 98, 99, Salgado 01). The linguistic approach was also been promoted by Trifonov (1993) in order to identify and characterize words by analysis of biological sequences; for a review of word recognition see Gelfand (1995).

### **3/ Regulation mechanisms as a molecular algebra**

The cell main memory is our field of observation. Neither the genotype nor the observation of the resulting phenotype is sufficient for deciphering the program; it is the ensemble of regulatory mechanisms in progress that are the privileged areas of observation. The program

is not written, the source code does not exist — it is executed at the same time as it writes itself virtually, it ensures its own maintenance (evolution) dynamically. We are attempting to learn the underlying language for this program, including syntax (molecular algebra), semantics (biological processes) and alphabet (binding sites).

#### *Syntax - a molecular algebra*

The basic instruction of the language of regulation mechanisms is the biological binding. We define "Biological Binding Operators" as a molecular algebra which uses two molecules as input operands and gives as a result the modification of one of the two operands. Each elementary binding operation is represented by : 1st operand, 2nd operand, operator, result. Each operand is described by: cell, molecule, type, state (Figures 1 and 2). Very often a cascade of activations leads to a propagation of the signal and will cause the cell to progress to the following state of the process, activating the expression of new genes or transforming substrates into new products.

List of operators and their resulting action :

- BINDING : change in state of a protein (activation / inactivation);
- TRANSDUCTION : is a chain of binding;
- EXPRESSION of a gene : addition of a new protein (circulating, cytoplasmic or receptor);
- METABOLICPATHWAY : addition of a new product.
- CELLDIVISION : addition of a cell;
- CELLDESTRUCTION : suppression of a cell;
- DESTRUCTION : suppression of a circulating molecule;

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

We have applied this grammar to two biological algorithms : a simplified model of the immune response (Figure 1); and regulation of the key enzymes involved in sugar metabolism (Figure 2), which is under hormonal control.

#### *Semantics of operators - Semantics of processes*

In computer science we have two levels of semantics: the semantics of the compiler which integrates arithmetical and logical operations, and the semantics of the program with its own rules which the programmer applies which is at the level of unrestricted language and Turing machine. To continue the analogy we can say that the semantics of "biological binding" lies in chemical interactions whereas the semantics of the program would be at the biological processes level.

*First level. - Semantics of operators.* In the Arithmetic and Logic Unit (ALU) "binary addition" is built from logical circuits. In fact the passage from the propositional logic of logical circuits to the first order logic of programming languages is dependent on: i) memory which introduces the notion of variables and ii) the introduction of mathematical structures which are supported by the semantics of addition; in fact the elaboration of the truth table for binary addition constitutes the link from one level of complexity to the next. Similarly, all the complexity of the mechanisms of regulation and integration is supported by biochemistry and molecular interactions. But we actually do not know how to build the equivalent of "binary addition", how to compute "Biological Binding Operators"; we actually do not know enough about chemical and physical laws to predict or to compute the result of the binding, in fact sometimes phosphorylation means activation and sometimes inactivation (Ninio 86). So we need the equivalent of the multiplication table given in extension for "Biological Binding Operators".

*Second level - biological processes or biological algorithms.* Some metabolic pathways are reversible and others are irreversible. Intra- and intercellular communication pathways permit control and regulation of the metabolic pathways that all cells of an organism are engaged in. These may be reversible metabolic pathways (generally intermediary metabolism), the

algorithm which control glycogen metabolism is an example (Figure 2), or irreversible pathways that commit the cell to a differentiation path, induced by growth factors. In the glycogen example, the hormones glucagon and insulin act in opposition to inactivate the competing pathway, whereas in the case of growth hormones, cell differentiation is not counteracted. In the control of glycogen metabolism process, the semantics of the algorithm is that two antagonistic hormones are required to reverse processes rapidly when necessary, this is an alternative structure. In the case of the immune system (Figure 1) the algorithm is iterative, and continues to operate as long as antigen molecules are detected.

#### *Alphabet, terminal symbols - binding sites*

This is not a program analogous to a computer program which would be coded in A, T, G, C rather than in 0 and 1. The program is not coded in A, T, G, C but in sequences of biochemical bindings between molecules: enzymes, DNA, co-enzymes, receptors, ligands and other elements which interact in the regulation and transmission of the signal. The terminal symbols for the grammar are the binding sites : DNA sites and the protein domains which interact. Generally a molecule has several activation sites, at least two: a regulatory site and a catalytic site. Whenever possible, the modeling takes the site level into account. This information is not always available for the process being studied.

#### *Digitalization of the signal*

The modeling is logical and syntactic. The computer is a sequential machine with discrete states; a study of the kinetics of reactions shows us that substrates and enzymes must be present in sufficient quantity for the reaction to occur; this is similar to the analogic electric signal which is converted to a discrete signal 0 or 1 based on a significant threshold, so it can be interpreted by the machine.

#### **4/ A minimal set of genes as the operating system**

One of the strong points of the Von Neumann architecture was to separate the control program of the machine (presently known as the operating system), from the specific data processing program. Although both programs use the same logic and are written in the same language, the first one is generic no matter what specific program it executes; it belongs to the machine. The operating system of the cell is constituted of regulation mechanisms which ensure the maintenance of the cellular machinery for gene expression; in fact this is a prerequisite for genomic expression: this cellular machinery is composed of proteins (ribosomal proteins) and RNA (rRNA and tRNA) which must be expressed in both necessary and sufficient quantities. This regulatory program is generic, and universal for all the cells of the organism; it is the kernel of the cellular architecture. It should be noted that this operating system program is written in the same language as the programs which are specific for each cell, i.e. it obeys the same rules of gene regulation and expression governed by biochemical binding laws. Other programs and processes are generic and ensure the proper functioning of the cell. This involves an ensemble of vital functions which are now described as a minimal set of genes. This minimal set of genes (Fraser 95, Hutchinson 99, Cho 99) may be larger or smaller depending on the complexity of the organism: it includes the DNA replication machinery, mechanisms of transport across membranes, entry and exit of essential metabolites, the flux of energy and information essential for inter-cellular communication, genes coding for the enzymes of intermediary metabolism and functions involving auxiliary organelles such as mitochondria (which play a fundamental role in cellular respiration).

#### **5/ Biochemicals as hardware**

The information substrate is biochemical. The program, the data and the machine itself are of the same substance—these are the biochemical constituents which obey biochemical and physical laws. The program contains the organizational plan of the machine; however the machine must pre-exist the execution of this program. The machine is the cell itself; all these constituents are reproducible, provided that the machine exists. The interaction between the cell and its DNA is intrinsic; one cannot function without the other. To postulate that there is a machine and a program does not take evolution into account and the fact that the machine

constructs itself dynamically. The machine does not exist before the program; not only the program and the machine are of the same substance, but the machine and the program evolve together.

The subtle but effective distinction that informatics makes between machine and program, wired instructions and coded instructions, does not exist in the logic of living systems; as in the first computers, the program is hard-wired. More correctly, only the machine exists; all the instructions are hard-wired but this is a machine which constructs itself, or more exactly reconstructs itself dynamically as a function of the program which it must execute. It is not the program which is coded in the DNA, but rather the basic knowledge which interacts with the volatile short term memory of the cell in a given state, using a molecular algebra which defines biochemical binding between molecules and induces a certain dose of determinism.

The software is volatile, the reproducibility of biological processes is due to the pre-existence of the cell body as a machine and the DNA as a database, as well as the stability of the internal physico-chemical context of the cell and the organism, as well as the stability of the external environment.

### **6/ Attempt at reprogramming life**

The year 1996 marked the first successes in animal cloning. Since 1998 biologists have been learning how to isolate and cultivate "undifferentiated" and "pluripotent" stem cells (Lagasse 01, Krause 01, Kehat 01, Ourednik 01, Jackson 01). In both cases, this involves obtaining the memory of the cell cytoplasm and the nucleus in a state corresponding to the desired stage: the initial state of the newly-fertilized egg in cases of cloning and transfer of a nucleus already containing the two strands of DNA or the state of a differentiated cell from a target tissue in case of re-programming of stem cells for re-implantation. According to the analogy that we are proposing, the problematic amounts to loading the memory correctly, and initializing the proper program which will read the DNA database. In order to change the destiny of the stem cell, one has to "load the memory registers" as was done with the first computers. In vitro, this means adding the products which will launch the desired cell differentiation pathway; in vivo, this means putting the cell in a tissue context in which signals from neighboring cells will initiate the process. In both cases one is amazed at the plasticity and faculties of adaptation and cooperation of cells and tissues. In the case of nuclear transfer, the cytoplasmic environment is determinant (Rideout 01). It is remarkable that the ovum can reverse the epigenetic modification imposed on the genome during differentiation to recreate a state of totipotency; these observations suggest that the initial transcriptional activity of the donor nucleus is controlled predominantly by the egg cytoplasm. In the case of grafts of stem cells it is the tissue environment that is determinant (Krause 01); indeed the local environment stimulates the expression of an ensemble of genes, which provokes morphologic changes, with environmental factors influencing the cell proliferation and differentiation. However, the failure of cloning experiments (Rideout 01) demonstrates the limits and difficulties of this artificial re-programming. It is as if the hard disk stopped by mistake: the reading heads were not necessarily correctly positioned (on the DNA) and the information (molecules) from the ovum cytoplasm may not always be sufficient to re-initialize the DNA pointers. Errors are often due to time factors; gametogenesis is a complex sub-program. Indeed, normal development depends upon a precise sequence of changes in the configuration of the chromatin and on the methylation state of the genomic DNA in the mature gametes, and is different for sperm and ovum.

### **7/ Input / output cell management**

The existence of membranes is fundamental for the evolution of metabolism and for cellular differentiation and identity: in fact, a critical concentration of matter and energy is necessary for the establishment of a distinctive differentiated metabolism that determines the cell's identity. However, the cell lives by interacting with its environment; its metabolism requires exchanges and circulation, both membrane proteins and secreted proteins are necessary to ensure communication and a dynamic state.

## **Discussion**

The logic of living systems goes well beyond the model of informatics systems, they allow modeling of observable functioning, but do not explain the fact that the machine regenerates itself; this modeling cannot answer questions about the origin of life and evolution of species. The conditions which led to the origin of life have been lost; nature no longer knows how to create a cell or a strand of DNA. There has been no creation of a cell for 3.5 billion years; cells in a given tissue divide and cells differentiate in the course of development. DNA is no longer created, instead there is replication, mutation, recombination, transposition and evolution of new proteins from old ones. We know how to synthesize a DNA strand *in vitro*, but we do not know how to synthesize a cell in a given state with all its components (cloning experiments use a mature ovocyte). There is continuity and transmission of memory from one cell to another. The active memory of cells is a continuous process in living systems; the machine never stops.

As in electronics in which few basic components (logical operators) are used, few biochemical mechanisms of activation have been observed, although biological functions are extremely varied. Evolution occurs by the elaboration of complex edifices from simple components. We can observe synonyms: several enzymes may perform the same function, and homonyms: enzymes which are very similar (sequence homology or combination of domains) but which perform different functions. The large majority of mediators belong to families of molecules which have diversified from a limited number of different types of molecules during the course of evolution. For instance it seems that the phylogenetic diversity of endocrine controls has been produced by duplications and the evolution of duplicated ancestral hormone-receptor pairs rather than by the appearance of new pairs. This is a sign that the program is written and saved at the same time that the function takes place.

## **Acknowledgments**

I thank the staff at Généthon and Infobiogen, for their support and Susan Cure for help in writing the English manuscript. I also thank Dr Maxime Crochemore at the University of Marne la Vallée for fruitful discussions

## References

- Artavanis-Tsakonas, Matsuno K., Fortini M.E. (1995) Notch Signaling. *Science*, **268**, 225-232.
- Atlan H., Koppel M. (1990) The cellular computer DNA: program or data. *Bull. math. Biol.*, **52(3)**,335-348.
- Atlan H, Cohen IR. (1998 ) Immune information, self-organization and meaning. *Int . Immunol.*, **10(6)**, 711-7.
- Barkai N., Leibler S. (1997) Robustness in simple biochemical networks. *Nature*, **387**, 913-917.
- Bentolila S. (1996) A grammar describing "biological binding operators " to model gene regulation. *Biochimie*, **78**, 335-350.
- Bozinovski S. (1996) Neuro-genetic agents, genetic code, and flexible manufacturing metaphor. *CMPSCI Technical Report 96-01. Computer Science Departement, University of Massachusetts at Amherst*
- Bray D.(1995) Protein molecules as computational elements in living cells. *Nature*, **376**, 307-312.
- Chomsky N. (1957) *Syntactic Structures*. Mouton, Paris.
- Collado-Vides J., Gutierrez-Rios R.M., Bel-Enguix G. (1998 ) Networks of transcriptional regulation encoded in a grammatical model. *Biosystems*, **47(1-2)**, 103-18.
- Collado-Vides J, Hofestadt R, Mavrovouniotis M, Michal G. (1999) Modeling and simulation of gene regulation and metabolic pathways. *Biosystems*. **49(1)**, 79-82.
- Danchin A. (1998) *La barque de Delphes*. Editions Odile Jacob, Paris.
- Douglas P. et al. (2000) Modeling Transcriptional Control in Gene Networks - Methods, Recent Results, and Future Directions. *Bull. math. Biol*, **62**, 247-292.
- Falke J.J. et al. (1997) The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annu Rev Cell Dev Biol*, **13**, 457-512.
- Fraser C.M. et al. (1995) The Minimal Gene Complement of *Mycoplasma genitalium*. *Science*, **270**, 397-403.
- Gelfand M.S. (1995) Prediction of function in DNA sequence analysis. *J Comput. Biol.*, **2**, 87-115.
- Hofestadt R., Meineke F. (1995) Interactive modelling and simulation of Biochemical networks. *Comput . Biol .Med.*, **25(3)**, 321-334.
- Hofestadt R. (1998) A integrative molecular information system. *Medinfo*, **9(1)**, 361-4.
- Hopcroft J.E., Ullman J.D. (1979) *Introduction to automata theory, languages and computation*. Addison Wesley, USA.
- Hutchinson C.A. et al. (1999) Global Transposon Mutagenesis and a Minimal *Mycoplasma* Genome. *Science*, **286**, 2165-2169.
- Jacob F., Monod J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, **3**, 318-356.
- Jackson K.A. (2001) Regeneration of ischemic cardiac muscle and vascular endothelium by adult stem cells. *J Clin Invest*, **107(11)**, 1395-1402.
- Kehat et al. (2001) Human embryonic stem cells differentiate into myocytes with structural and functional properties of cardiomyocytes. *J Clin Invest*, **108(3)**, 407-414.
- Kolchanov N.A. et al. (1998) Genexpress: a computer system for description, analysis and recognition of regulatory sequences of the eukaryotic genome. *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology, AAI Press, Menlo Park, California, USA*, 95-104.
- Kolchanov N.A. et al. (1999) Integrated databases and computer systems for studying eukaryotic gene expression. *Bioinformatics*, **15(7-8)**, 669-86.

- Krause D.S. et al. (2001) Multi-Organ, MultiLineage Engraftment by a Single Bone Marrow-Derived Stem Cell. *Cell*, **105**, 369-377.
- Kupiec J.-J. (1997) A darwinian theory for the origin of cellular differentiation. *Molecular and General Genetics*, **255**, 201-208.
- Lagasse E. et al. (2001) Toward Regenerative Medicine. *Immunity*, **14**, 425-436.
- Lassaigne R., de Rougemont M (1993) *Logique et fondements de l'informatique*. Hermes, Paris.
- Lassaigne R., de Rougemont M. (1996) *Logique et complexité*. Hermes, Paris.
- Lichstein D., Atlan H. (1990) The "cellular state": the way to regain specificity and diversity in hormone action. *J Theor Biol*, **145**, 287-294.
- Ninio J., Bokor V. (1986) Stratégie d'adaptation moléculaire. *La vie des sciences, CRAS, série générale*, **3**, 121-136.
- Ourednik V. et al. (2001) Segregation of Human Neural Stem Cells in the Developing Primate Forebrain. *Science*
- Rideout III W.M. et al. Nuclear Cloning and Epigenetic Reprogramming of the Genome. *Science* , **293**, 1093-1098.
- Salgado H. et al. (2001) RegulonDB (version 3.2): transcriptional regulation and operon organization in Escherichia coli K-12. *Nucleic Acids Res.* **29(1)**, 72-4.
- 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.
- Searls D.B. (1997) Linguistic approaches to biological sequences. *Comput Appl Biosci*, **13(4)**, 333-44.
- Searls D.B. (2001) Reading the book of life. *Bioinformatics*, **17(7)**, 579-580.
- Sonigo P. (1990) Design and trials of AIDS vaccines. *Immunology Today*, **11**, 465-471.
- Trifonov E. N. (1993) DNA as a language in *Proceedings of the 2nd International Conference of Bioinformatics, Supercomputing, and Complex Genome Analysis* (Lim H A et al, eds) World Scientific, 103-110.
- Turing A.M. (1936) On computable numbers, with an application for the Entscheidungsproblem. *Proc. Lond. math. Soc.*, **42**, 230-265
- Von Neumann J. (1945) First Draft of a Report on the EDVAC . W-670-ORD-4 926.