ADVANCED STUDIES IN PLANT BIOLOGY : INTEGRATING PLANT SYSTEMS, 2 000.
EDITED BY H. GREPPIN, UNIVERSITY OF GENEVA

SYSTEMIC MODELING OF BIOLOGICAL FUNCTIONS[¶]

APPLICATION TO THE DEVELOPMENT OF THE GENINTER SOFTWARE DEDICATED TO THE COMPILATION OF INTERRELATIONSHIPS BETWEEN GENES AND/OR GENE PRODUCTS

Magali ROUX-ROUQUIE
Biosystémique Modélisation Ingénierie, Génopole-Pasteur, Institut Pasteur, 25-28, rue du Dr. Roux 75724 Paris Cedex 15, France E-mail: mroux@pasteur.fr

INTRODUCTION

The discovery of gene function constitutes one of the major challenges of the genome projects engaged on a large variety of organisms, from bacteria to human.

Until now, this characterization was carried out by the implementation of empirical, physiological, biochemical data, based on the implicit assumption of a bijective mapping between structure/sequence and function. At the beginning of 1980s, approaches referring to the intuitive concepts of ortholog and paralog genes (which received their formal definition later on) have led to the cloning of great functional units like the human lymphocyte antigen (HLA) Complex (1) and to the postulation of mechanist assumptions (2) to explain similarities and differences at functional level between related gene products. In the 1990s, the positional cloning (3), starting from phenotypes has led to gene identification according to the position within the genome, thus, supplementing the device of functional analysis of the genome. During the two preceding decades, more than 10.000 mammal genes have been characterized from the point of view of their structure and of their function.

With the rise of genome projects coupled to strategies for fast highlighting of new genes, considerable amount of sequence data are accumulated to which it is essential to allot a function. The implementation of the traditional methods not being possible for reasons of times and costs (4), data-processing analyses are carried out, in particular, by the search for similarity with known genes. This approach which proves powerful when it falls under a more global approach integrating empirical data (what would relate to only 2% of the data in the database GenBank), reached today its limits in terms of precision and consistency with respect to the biological significance which it makes possible to deliver.
Indeed, it is not rare to meet the situation where two proteins A and B have similar catalytic domains and are seen allotting the same function (acetyltransferase, for example); when a new sequence C is found similar to B - but for another motif/domain which confers, for example, the property to fix itself to the ADN -, C will become by

[¶]A specific workshop has been launched on this topic at the Web site of the European program "Modeling of Complexity" http://mcxapc.org/

transitivity an acetyltransferase even if it carries other properties as transcriptional cofactor ! According to this play of similarities, incomplete data - even erroneous properties - are propagated to new sequences (5).

These reports raise basic problems for the description and the representation of the biological functions. Does the question " *which function is related to such a gene* ? " still has a sense when a same gene can play different roles according to its environment (6). Moreover, the modular arrangement of proteins introduces an additional level of complexity, a motif being able to act on the DNA, a second motif on another protein, a third part still, to catalyze a chemical reaction; worse, the activity of a same domain can be modulated by the nature of its partners (7). Otherwise, isoforms constitute another source of functional diversity; the neurexins constitute a particularly representative example because three genes are sufficient to produce several thousands of distinct proteins by alternative splicing (8). Conversely, unrelated genes deprived of any structural homology, can exert identical functions : comparisons between *Influenza Haemophilus* and *Mycoplasma genitalium* has led to the discovery of non-orthologous genes exerting identical vital functions; thus, ensuring a true functional redundancy. Several hundreds of similar cases have been estimated to exist in the eucaryotic genomes (9). Furthermore, genomes, from bacteria to human, do not show linear relationships between size, gene number and organizational levels : whereas human and mice genomes show identical size (3 Mbases) as well as equivalent gene number (80,000-100,000), two bacteria, *haemophilus influenzae* and *E. coli*) can be found in a ratio of 2, according to gene number (1703 and 4288, respectively), suggesting an essential role to epigenetic constructions in higher organism genesis.

With this respect, it appears not very probable that the analysis of the sequences is sufficient to ensure a direct access to the function. When J. Monod wrote "*the genome entirely defines the function of a protein*", force is to recognize that he modulated the range of this assertion by commenting on the paradox of the epigenetic enrichment which provides a structure with a higher informative contents than that reserved by the genetic determinism (10). However, only the first part of this statement was retained by a broad fraction of the biologist community to sit the bases of a reductionistic approach which only recognizes an explanatory capacity to the structures. With the passage of structural genomics which answers the question " *of what is it made* ? " to the functional genomics which questions "*how does it work* ?", the strategic interest of the functional assertion : "*how does it work* ?" is to suggest new heuristic approaches as well as epistemological implications which it would be advisable to examine.

To achieve this aim, it requires to carefully define the axioms to be used and to check their consistency with the reality to be modeled i. e. : *the functions of living systems*. With this respect, the concepts of *System* and *Living System* must play a pivotal role in the theoretical framework to be used to model biological functions.

BASIC CONCEPTS IN SYSTEMIC MODELING

The concept of `System` was formally introduced by Leibnitz in 1666 as "*a whole of*

elements". This thought is still long-lived and forms a basis for the ambiguities currently in force about the concept of System and resulting in the approach of systems through the mathematical set theory.

Nevertheless, more inclusive definitions have been provided since that, according to Ackoff (1957) (11) who designed a system as "*the unit resulting from interacting elements*" and Rapoport (1968) (12) who defined a system as "*a whole which functions as a unit according to its interactive parts*". This is Von Bertalanffy (13) who first introduced the concept of System in biology, as "*a set of units with relationships among them*". Nevertheless, that is Saussure (1931) (14) who first linked the concepts of System and Organization by defining a system as "*a unit of organized relationships among elements*". This is the definition adopted by Morin who examined the notions and characters related to the concept of System; thus, providing an operational intelligibility of this concept (15), as described below.

Linking Interaction *and* Organization

Interactions are reciprocal actions modifying the behavior or the nature of interacting elements belonging to a system. Starting from disorder, encounters between elements are random but the effects on these elements can produce order according to certain constraints which may depend, notably, on initial conditions and/or intrinsic properties of these elements. When they give rise to the phenomena of organization, interactions become interrelationships consisting of associations, connections, combinations, communications, etc. and constitute the plate-revolving concept between order/disorder and organization. It must be mentioned that these concepts of order/disorder and organization profit from a theoretical framework with thermodynamics of irreversible processes. As matter of fact, the reduction in entropy and the stationary maintenance of entropy deal with organizational development and the increase in entropy resulting in the environment, deals with disorder and disorganization. As a result, the concept of entropy adapted to irreversible systems has led to new concepts such as self-organization.

Linking Organization *and* System

Organization is the layout of interrelationships between elements which confer significant stability or regularity and produces an individual unit equipped with unknown qualities on the level of the components : the system. The system takes the place of simple objects and is substantial ; as a system, it is rebellious to the reduction into its elements.

Linking System *and* Complexity

The concept of system is paradoxical: regarded as a whole, it is homogeneous, considered under the angle of its components, it is heterogeneous. With this respect, a system is a non-elementary global unit made up of various parts in relation. The idea of complex unit shapes as the system has something more than its components considered in a separate or juxtaposed way and cannot be reduced to its parts. This statement can explain why the concept of system was circumvented, even neglected by a science which established its bases on the reducible one, on the simple one.

Complexity versus Complication

The terms of complexity and complication are usually used one for the other whereas their respective meaning has deeply different epistemic implications. J-L. Le Moigne finally released truly the differences between the two concepts (16) : what is complicated can be reduced to the simple, complexity cannot ; a hank is complicated, with a lot of efforts one will be able to clear it up. Accordingly, complicated systems can be broken up into simple elements even if it is at the end of large efforts; in contrast, the elementary principle of a complex system is an implex which refers again to a complex unit.

Three principles account for the operational intelligibility of the concept of system as complex unit (15).

Principle of Emergence *: The whole is more than the sum of the parts*

One calls Emergence the system property which constitutes a new and additional character compared to the qualities of the components considered separately or arranged separately in another system. The organization confers emergent properties to the system which are generated by the way in which it is made up and not only by the properties of its components ; this is this unreducibility in basic units and these relations which make the concept of System, complex. Correlatively, the concept of System is a basic concept because it develops into system of system of system...; such that, it is with the root of complexity. The phenomenal reality of Emergence consists of the new qualities of the system ; they are not logically deductible and are physically irreducible : they are lost if the system dissociates ! Thus, a fabulous systemic architecture is built, the emergent qualities of a system on a lower level becoming basic materials of an higher system in a polysystemic organization : the natural systems constitute an inextricable tangle of systems, from the subatomic level of elementary particles, to the populational level of individuals, passing through the cellular one.

Principle of Constraint : *The whole is less than the sum of the parts*

In addition to the principle of Emergence, any organizational relation exerts restrictions and constraints on the elements or parts which are subjected to it to achieve a specific system, only a limited part of element properties being recruited ; this constitutes the principle of Constraint. From a phenomenal point of view, the constraints of the whole on the parts are due to the organizational character; the constraints of the parts on the whole are due to the material character of the parts. This is reminiscent of a very basic organization underlying biological functions according to the selective involvement of specific domains by modular protein in allosteric interactions.

Principle of Organiza(c)tion : *A system is/has an active organization*

At this point, the fundamental fact is, not only the idea of organization, but the idea of active organization. To say that an organization is active is to say that it generates actions and/or that it is generated by actions with this respect, any active organization can be compared to the organization of a machine. According to its usual meaning, a machine is an manufactured instrument which achieves operations according to its organizational properties. Progress of cybernetics, in particular in the direction of the

operational autonomy of machines, resulted in wondering not only about what does the machine produce but also on what is a machine. One thanks to Wiener (17) not to design a machine like a material instrument but like an physical being with organizing properties The widening of this definition, from the artificial systems to the living systems, makes emerge an unknown basic concept of the artificial systems: the Self. The artificial systems, the industrial machines for example, differ basically from the living machines by their incapacity to reproduce themselves; according to Morin, they have physical being but no Self.

At this point, it is essential to underline the essential characters allowing to give an operational definition to the concept of System. With respect to the principles of Emergence, Constraint, and Organiza(c)tion, a system is a complex unit with active organization among elements; it is at the same time more, less, other that the sum of the parts and its parts are less, possibly more, in any event other than what they would be by themselves. This paradoxical formulation shows the nonsense that there would be to reduce the description of the system in quantitative terms; not only the description must be also qualitative but it must be complex (16).

Linking **Complex System** ***and*** **Living System**

The idea of living system inherited the substantiality of the former living matter and vital principle ruined by the modern biology. But, in spite of life has consistency only at the atomic level, it is also and especially, the product of a vital organization. Consequently, to assess the concept of Life is not only to know the alphabet of the genetic code and the troop of molecular structures which goes with, it is also to know organizational and emergent properties of living beings and to conceive organized and complex units.

The concept of Self

Extended to the living systems, the metaphor of machine - so invaluable to evoke the principle of organisa(c)tion - was examined by Maturana (18) who compared the artificial machines and the living machines by distinguishing (i) the " allopoïetic " systems (the artificial machines) whose operation produces something different from themselves and whose organization remains invariant as long as their product remains the same one (and vice versa); and (ii) the autopoïetic" systems " (the living systems) which are the product of their own operation and whose organization remains invariant as long as they reproduce themselves. The central character of living machines is that they are producing self, organizing self, reorganizing of self, their "*poïesis*" is identified initially with the permanent production of their own being. Self is born in the permanent production and organization from its own being. The idea of self is capital and it is the source of what is specific to the living systems : self-organization (19).

The concept of Goal-directedness (Teleodirectionallity)

But, it is not enough to note the self-organization of the living systems to reach their intelligibility, still is necessary to integrate the functional order in which the various levels which compose them, are arranged. That is particularly true if one considers the systems of regulation in eucaryotes which, although presenting analogies with the procaryotic systems, are infinitely more complexifed. It is necessary to see in this tangle

of systems, an increasing goal-directedness implication (also called teleodirectionnality) of the structural determinants such as it generates increasingly complex functions. It must be reminded that a teleonomic process or behavior is directed by a program and depends on the existence of a goal which is envisaged in the program articulating the behavior. This term of teleonomy was substituted for that of teleology in 1958 in order to name and describe all end-directed systems in opposition to any reference to the concept of teleology impresses of Aristotle finalism . J. Monod used to find this concept profoundly ambiguous and he completed the confusion by choosing to define the essential teleonomic project as consisting in the transmission from generation to generation of the variance content characteristic of the species, mixing up "selective value" and goal-directed activity or behavior controlled by a program (10). Nevertheless, this theoretical concept of teleology /teleonomy seems practically impossible to circumvent even on very lower, intracellular or molecular levels, and imposes an architecture of integrated subsystems on which is based the teleodirectionnality of a larger one (15).

The sole analytical approach does not allow to specify the relation between the molecular components of the genetic program and the plastic, end-directed and self-regulating systems which result from this. One is still limited to a descriptive mode of the style: " if the genetic structure and the surrounding conditions are known, then the appearance of a particular sequence of process could be specified ".To surmount this obstacle, it is necessary to seek other strategies, new theoretical frameworks more adapted to the object of search - Living Systems - than we have just defined as self-organizing complex units.

THE FORMALISM OF COMPLEX SYSTEM MODELING

The formalism is the way to express a concrete system into an abstracted form ; in return, the opposite becomes a interpretation ; the formalism of complex system modeling provides heuristic rules to assess complexity (20). A rapid overview of paradigms on which is based the paradigm of organizing complexity developed by J-L. Le Moigne in the "Dictionary of History and Philosophy of Sciences " (21) would help to focus one special requirements for the modeling of biological functions in living systems.

In the analytical paradigm, who was mainly used to study Mechanics and Energy, the central assumption is a deterministic assumption which accounts for the phenomena by a cascade of causal relations. The key concept is the object : (i) only the structure is explanatory, (ii) the structure is the cause, the necessary and sufficient condition of the effect, (iii) the function is provided by the object. The extraordinary fruitfulness of this paradigm during three centuries has been such, that it seems still the reference of any scientific approach. In the introduction of this paper, we noted the insufficiencies of the analytical paradigm to describe the biological functions. The Cybernetics paradigm restores the concepts of projects, goals, teleology for the study of the behaviors of the objects or the natural and artificial phenomena. Instead of centering the attention on the mechanisms and the structures, it proposes to be unaware of them by locking up them

in a black box while privileging the interpretation of the behaviors. It is not a question any more of explaining the mechanisms by themselves but of understanding or interpreting the behaviors in permanent reference to the projects of the modeled phenomenon. Unfortunately, the cybernetics approach do not allow to manage within a same model the duality of the functioning and evolving object.

In the thermodynamic paradigm, the characteristics of the classical thermodynamics which limits the possibilities of making use of it to analyze the living systems are that : (i) thermodynamics studies isolated systems. In theory, the possibility of isolation (closed system) raises doubts and in the field of biology it is wrong as biological systems as open systems (ii) thermodynamics studies states at equilibrium. With regard to the living systems, it should be recalled that all forms of life are not at equilibrium and the passage to such equilibrium means the end of the lifetime; that is enough to understand that the states at equilibrium are not of great help to describe the living organisms ; (iii) the possibility of a difference in entropy between an particular system and the medium in which it is (disproportionate distribution); (iv) thermodynamics does not deal with speeds of the transformations. The dynamic systems met in biology have mechanisms of regulation which ensure their stability; the fight between the entropic tendencies and the mechanisms of regulation end only in the victory of latter only in the dynamic systems. With the emergence of nonequilibrium thermodynamics (irreversible transformations), the concept of time essential to living organisms was introduced by the concept of flow and led to new concepts such as the self-organization and the dissipative structures as mentioned previously. Other concepts enriched this approach, one will quote in particular the non-linear system theory, the theory of fractals…

In " Dictionary of History and Philosophy of Sciences", J-L. Le Moigne appreciates the contributions of these paradigms by noting that they apply to the study of simple systems (decomposable) or to the systems of "low" complexity (disorganized complexity) in opposition to systems of "high" complexity (self-organizing systems) (21).

The Canonical Model of the General System

The axiomatic of conjunctive logic is necessary to instrument Systemic Modeling (SM) as the disjunctive logic justified division as the instrument of analytical modeling. This argumentation is established by the construction of the canonic form of the General System (16, 20). Indeed, to represent a complex phenomenon, one must represent it by an enough general system to give an account of all the types of complexity.

The canonic form of the General System integrates formally the conjunctive axiomatic : the concept of General System emerged by the conjunction of two concepts which are at the origin of modeling : Cybernetics and Structuralism. Cybernetics - as seen previously- is founded on the conjunction of the concepts of active environment and project or teleology. Structuralism is founded on the conjunction of the concepts of operation (to do it; synchronic) and transformation (to become it; diachronic).

The systemic conjunction proposes to hold for inseparable, the operation and the transformation of a phenomenon, the active environment in which it is exerted and projects in relation to which it is identifiable (Figure 1A). One can check, thanks to this definition, that the general System absorbs the three axioms which are the base of systemic modeling: `Operationality`, `Teleological irreversibility`, and

non-separability.

The Canonical Form of Process

To model a complex system is to model a system in action (16, 20). One does not seek to model objects, bodies, as one would do it in analytical modeling. The basic concept of systemic modeling is not the object or the combination of stable objects (the structure) but the action. The characterization of an action or a function can be done recursively, it requires the general concept of Process.

One defines a Process by its exercise and its results : there is Process when there is the modification of a collection of identifiable objects in a reference frame " Time-Space-Form " (Figure 1B). The conjunction of a temporal transfer (a displacement) and a temporal transformation (a change in form) constitutes a Process; one recognizes it with its result: a displacement in a reference frame TSF. Here, the form is taken in the German significance of Gestalt, i. e. what can be distinguished in a sufficiently stable way from its background (from which it is however inseparable). A Process is thus a complex of actions, multiples and even tangled up, which one can always represent in a reference frame TSF. In other words, one can represent a Process by the articulation of the three archetypal functions: Time, Space and Form. These functions are exerted on a collection of unspecified tangible or intangible objects.

Graph of the network of processors

If one agrees to indicate by a Processor Pr, the box by which one represents a Process, it is said that there is a Relation between two processors Pi and Pj when the output of processor Pi is the input of processor Pj ; the Interrelationship IR is then activated (Figure 1C). All the combinations of possible interrelations between N processors can be represented using the structural matrix of the system; the presence of "1" will mean that the interrelationship is activated, the presence of "0" that this interaction is not activated, possibly prohibited (Figure 1D). One can identify feedback interrelationships since some inputs are resulting from some outputs produced before by this processor; such relation informs the system about its state. Thus, there are *a priori* $2N^2$ networks (and thus $2N^2$ graphs) to represent the behavior of a system of N active processors. It is obviously constantly possible to privilege an unspecified square submatrix in the structural matrix by arbitrarily limiting it by selection of the processors which one wishes more specifically to study the internal layout or the total behavior. It is then important not to destroy the interrelationships between this submatrix and the other processors or aggregates of processors. This representation proposes a modeling of the environments, it allows a representation of an opened general system, inter-connected subsets of processors. In addition, the general system could thus be described by one of the possible values of the structural matrix (21).

BEYOND REDUCTIONISM, THE SYSTEMIC APPROACH INTRODUCES NEW PERSPECTIVES IN THE LIFE SCIENCE

At this point, modeling biological functions in living systems shall deal will major concepts of Complexity (Emergence/Constraint), Self-organization [organiza(c)tion and Goal-directedness (teleonomy/teleology) which delineate the central characters of living systems.

According to the complexity of living systems, the individual knowledge of the functioning of each one of their elements (molecular, supramolecular...) is not sufficient to understand and to describe the functioning of the unit. It is moreover necessary to define the relations and connections as well at the topological level th at the qualitative and quantitative levels (principle of Emergence). This property of systems fits well the observations showing that a protein can play distinct roles depending of the partners with which it interacts. Conversely, the functioning of the unit imposes restrictions on its elements so that only one fraction of their properties and their possibilities of action is implemented in the play of the relations which link them (principle of Constraint). The activity of the genome which is expressed differentially according to the cellular type with mechanisms of regulation sophisticated, illustrates perfectly this property and confirms the biological consistency of the conceptual framework offered by the systemic approach.

But, it is not enough to describe relations between the components of a biological system to apprehend the functioning of it. Still, it is necessary that the analysis of these processes integrates the functional order that expresses these phenomena. Moreover, it should be explained how the various structural levels are arranged in order to produce the complex effects illustrating the emergent functional order.

The complete analytical decomposition of the genome for its modeling in network of chemical determinations governing the processes of metabolism and development, raises the question of hypercomplex models difficult to conceive. Systemic modeling, on the basis of the question " what that makes? ", opens new strategies of investigation. It is worried in priority of the functions to ensure, by considering the operated transformations and the links which organize these transformations.

Although it is difficult to simultaneously translate the aim of a system, its components, the interactions between these elements and various other properties of the system, the recourse to the symbolic formalism allowed by systemic modeling, constitutes a major projection for the description of biological systems. In a recent paper, we showed that prototypical organizations for chromatin remodeling may be assessed from a tentative representation taking into account the nature of partners, the effects... (22).

Accordingly, the figure 2A gives an artwork representation of the TGF?? signaling system : TGF?-related factors regulate cell proliferation and differentiation in organisms ranging from insects and worms to mammals. Although only the receptors for TGF?s, activins and BMP have been characterized, all TGF?-related factors are thought to act through a cell surface complex of two types of transmembrane serine/threonine kinase receptors. Following ligand binding, the type II receptor kinase phosphorylates and thereby activates the type I receptor. The SMADs (for C. Elegans Sma and drosophila

Mad) of which approved human symbol is MADH (for Mother Against Decapentaplegic, Homolog) then act as type I-activated signaling effectors. Due to Type I receptor dissociation, activated MADHs form complexes with MADH4 and are translocated to the nucleus where they regulate transcription of specific genes. This model has strong similarities with the Jak/Stat signal transduction pathway from activated cytokine receptors. At this point, MADHs are considered as signaling effectors for the ligand-induced transcriptional responses. A systemic representation takes into account the aspects related to a transformation of `Form` (TGFBRI, TGFBRII, MADH3... are activated by phosphorylation), a variation in `Space` (MADH3 is translocated to the nucleus), and in `Time` (the phosphorylation of MADH3 precedes its translocation in the nucleus).

Recent progress in TGF?-signaling analysis provides new insight to explain the sequence heterogeneity of TGF?-responsive promoter sites. As they are translocated to the nucleus, MADHs have been shown to act as transcriptional factors through their ability to directly bind DNA, and to induce transcriptional responses through cooperativity with other transcriptional factors. Figure 2B pictures the cooperative interactions of MADHs with other transcription factors, which drive ligand induced transcription from responsive-promoters. Interaction of MADH2/4 with FAST1 at Mix2 promoter or FAST2 at goosecoid promoter results in activin/TGF?-induced transcriptional activation. Interaction of MADH3/4 with c-FOS/c-JUN allows TGF?-induced transcription from the collagenase I promoter. The similarity in both mechanisms of activation and the proposed cooperativity of MADHs with several DNA-binding transcription factors suggest a prototypical model for transcription activation by MADHs : the heteromeric MADH complex and the cooperating transcription factors (X) interact with promoter sequences ; CREBBP/EP300 can act as activator of MADH2 and MADH3 through direct association. Based on these observations, MADH-response promoters have a double DNA sequence requirement. One sequence confers the specificity to bind transcription factors that cooperate with MADH complex. Another adjacent sequence is required for direct MADH binding and confers MADH selectivity to the first sequence. Thus, only a subset of promoter sequences that bind these cooperating transcription factors, is target for MADH signaling ; this transcriptional cooperative model provides a mechanism for integration of two cross-talking signaling pathways at promoter sequence.

Using the formalism of the Canonical Form of General System, the interrelationships between genes and/or gene products require to be represented as processors (Figure 3A) and the matrix gives a representation of the functional sequence of the events (Figure 3B). With respect to this particular example, one can note it, the structural matrix is very useful to represent the functional order that governs the setting of this particular organization.

Systemic Modeling of Cell Functions

Using the formalism of the systemic representation allow one to describe biological activities in terms of process in a TSF frame : the elementary process is the interrelationship between at least two structures (Figure 4).

These relations could be classified into three types : (i) relations of linear causality in which the former causes involve the effects in a systematic (determinisc) way ; generally a body of causes is combined to produce an effect; (ii) retroactive relations are characterized by a circularity between the events; the anteriority of the cause on the effect disappears and yields the place to the " regulator "; (iii) with the recursive relations, the produced effects are necessary to the processes that generate them.

With this respect, the interrelationship can be named by symbols (the gene symbol), through the nature of the active molecules (ADN, ARN, protein), upstream regulations exerted by distinct structures and/or interacting structures (input) and downstream effects on other structures and/or interacting structures (output). Regulations and effects can be characterized in terms of Space (Localization), Time and Forms (modification of such structure, such relation...). Time can be represented using discrete values if any, or symbolic values by locating the interrelationship/processor within the course of a process (for example: the relationships involved in chromatin remodeling occur during the pre-initiation phase of the transcription process)

APPLICATION TO THE REPRESENTATION OF CELLULAR FUNCTIONS : GENINTER, *A SOFTWARE DEDICATED TO THE COMPILATION OF INTERRELATIONSHIPS AMONG GENES AND/OR GENE PRODUCTS*

As a first step to model living systems using the systemic approach, we are currently developing a software dedicated to knowledge representation of cellular activities.

It must be mentioned that the aim of a system of representation guides the choice of the criteria for the selection of the elements to be represented. For example, if the objective is to represent the three-dimensional structure of proteins, the elements of the representation will concern, in addition to the primary sequence, the domains and the active sites, their modeling in space, the similarity with other proteins in the same species or different species..., but would neglect the role played by these proteins. Although correct, this representation will have a limited rationality since while drawing aside the functional properties, it will not be able to account for (i) the existence of functional differences in proteins presenting the same spatial organization (TIM barrel fold family) or conversely (ii) to identify a same function among proteins presenting distinct 3D organization (serine proteases).

By adapting the rules of design and construction of the complex systems to the representation of the cellular activities, one admits (i) the inference linking the achievement of a biological function to the presence of process and active structures, (ii) the general assumption of self-regulation, (iii) the effects as specific functional

ingredients. Our middle-term goal being to model the network of functioning genes in the paradigm of organizing complexity, our major effort was intended in the description of the interrelationships between cellular elements.

General description of GENINTER

GENINTER was originally dedicated to human genes but it can be used to compile data on genetic and physical interactions of any species. The prototype presented below will be detailed in a forthcoming paper (23).

The pivotal entity in GENINTER is the `Gene`. It is characterized in the `Gene Keyboard` form according to (i) its approved `Symbol(ex: MADH4)`, including the (ii) `Other Symbol (ex: SMAD4, DPC4)` which correspond to the alias currently used in literature, (iii) its `Designation` which explicit the approved symbol [ex: Drosophila Mad (mother against dpp), Homolog of, 4], its `Access Numbers` to GenBank database (`ex :U44378`). In addition, its chromosomal location on human chromosomes (ex: 18q21.1) and associated diseases if any (ex : JPS=juvenile polyposis) is filled in.

In addition, the `Product` of the gene is described according to a `Product Type` (RNA or Protein), its `Access Number` to SwissProt database, as well as other major features like its membership to a known protein (super)`Family`, its current `Subcellular Localization` and its `Tissular Expression` as well as `Isoforms`. The name (`Motif Name`) and the position (`Motif Position`) of any known motifs is detailed with respect to the amino acid length (`Product Length`); further graphic interfaces would be developed to represent motifs within the sequence.

Ten tables, not to mention the tables developed for controlled vocabulary, are linked to the table Gene in the `Gene Keyboard` form (Figure 5).

The `Gene Keyboard` form is roughly usual in gene database; although, in current databases, data are often compiled in a unique table; thus hampering any sophisticated querying of the database.

Description of Interrelationships in GENINTER

The original contribution of GENINTER concerns refined descriptions of interactions in order to develop tools for function analysis.

In the `Interaction` form (Figure 6), one `Type` of structure (DNA, RNA, Protein) is designed as `Self (Symbol Self)` or `Modified Self` if it is the modified (activated) structure which is involved in the `Interaction` (essentially modified proteins : acetylated, phosphorylated, and so on.). The `Self` has `Partners` which have `Type`(DNA, RNA, Protein) and physically interact with the Self. The physical `Interface` between `Self` and `Partner` is compiled as a sequence interval (amino acids or nucleotides depending on the `Type` of the structure). `Other Member` may be involved in the `Interaction` without interacting physically with the `Self`; they are taken into account.

In addition, the `Interaction` form mention the published `Reference` filled in with

links to Medline entries

Specific buttons in the `Interaction` form allow one to move to linked forms :

(i) In the `Regulation` form, the `Regulation Type` refers to controlled vocabulary (*activated by, inhibited by…)* and `Regulation Symbol` to the structure involved in Interaction Regulation. As mentioned for `Self`, the `Regulation Symbol` can be modified (`phosphorylated, acetylated`…) and the `Regulation Interface` may be known, so it is filled in. Interaction `Regulation` may depend on modifications on Localization (`Regulation Localization`), on Process (`Regulation Process`), according to a particular `Cell` Type or `Tissue`. At last, Interaction may be regulated according to a Structure Complex (`Regulation Complex`) instead of an individual structure; thus, the `Regulation Complex` is defined using `Regulation Complex Symbol`, `Regulation Complex Type`, `Modified Regulation Complex Symbol` and `Regulation Complex Interface`.

(ii) In the `Effect` form, the `Effect Type` refers to a controlled vocabulary (increases, initiates, potentates …) and `Interaction` has effect on structure (`Effect Symbol, Modified Effect Symbol, Effect Interface`), Localization (`Effect Localization`), Process (`Effect Process`), these effect concerns a specific Cell (`Effect Cell`) or `Tissue`. As mentioned in the Regulation form, effects may concern a complex of structures which will be described as `Effect Complex, Effect Type, Modified Effect Complex`, and `Effect Interface`.

(iii) the `Process` form indicates the process in which the interaction is involved (HDAC1:SIN3 interaction is involved in promoter silencing whereas HDAC1 is known as a histone deacetylase, i.e. as a protein modifier) ; the `Process Name` is extracted from a specialized dictionary organized according to specialization and composition Relationships (24)..

(iv) the `Localization` form is organized on the same base as the `Process` form ; it indicates the localization of the interaction (the interaction between TGFBR1 and MADH3 occurs at the inner face of the plasma membrane whereas MADH3 is known as cytoplasmic). The `Localization Name` is extracted from a specialized dictionary of Macromolecular Structures (24).

(v) The `Experiment` form is filled in using an taxonomy yet under development.

Thus, it required a set of 24 tables linked to the table `Interaction` in order to consistently described interrelationships between genes and/or products.

Querying : Visualization of interaction data

Interactions and underlying concepts (partnerships, effects, regulation, etc.) being explicitly stored and structured in specific tables, it will be possible to query over interactions. The vocabulary being structured, the result of any query will be then more accurate. We are currently refining terms of the controlled vocabulary using a statistical and a lexical approach (23).

Future developments.

GENINTER is an important step toward the explicit representation of interactions among genes and/or gene products. Its querying facilities will be powerful and quite helpful. In addition, we plan to develop classification according to interrelationships in order to improve function assignment to gene sequences.

REFERENCES

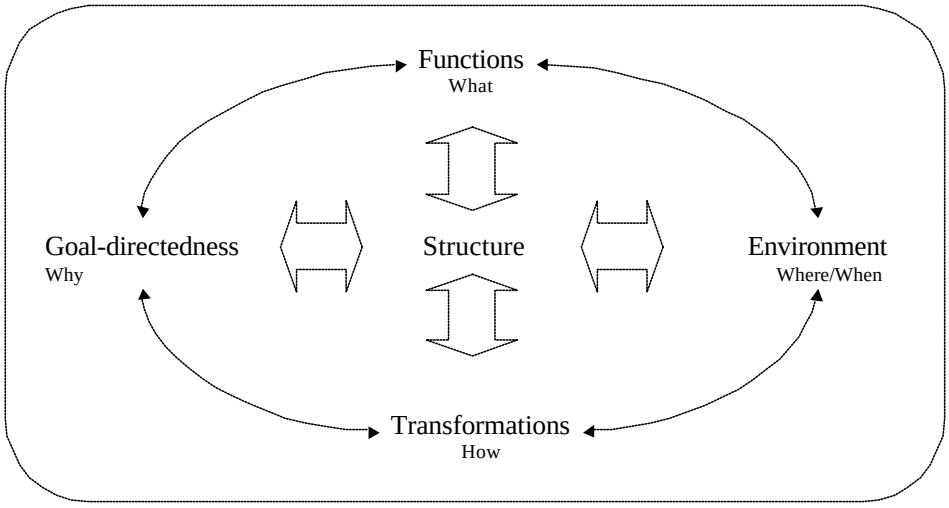
1. Roux M, Auffray C, Lillie JW, Boss JM, Cohen D, DeMars R, Mawas C, Seidman JG, Strominger JL (1983) *Genetic mapping of a human class II antigen beta-chain cDNA clone to the SB region of the HLA complex*. Proc. Natl. Acad. Sci USA 80:6036-6040.
2. Roux M, Auffray C, Lillie JW, Korman AJ, Strominger JL. (1983) *Homology matrix comparison of human and Murine class II antigens.* In « Gene Expression » , UCLA Symposium on Molecular and Cellular Biology, D. Hamer and Rosenberg M (Eds), Liss, New York, Vol 8, pp. 481-490
3. Collins FS (1995) *Positional cloning moves from perditional to traditional.* Nat Genet 1995 9:347-50
4. Fields, S. (1997) *The future is function*. Nature genetics 15, 325-327.
5. Pennisi, E. *Keeping genome databases clean and up to date.* (1999) Science 286, 447-450.
6. Jeffery CJ (1999) *Moonlighting proteins*. Trends Biochem. Sci. 24:8-11
7. Perissi, V., Dasen, J. S., Kurokawa, R., Wang, Z., Korzus, E., Rose, D. W., Glass, C. K., Rosenfeld, M. G. (1999) *Factor-specific modulation of CREB-binding protein acetyltransferase activity.* Proc. Natl. Acad. Sci. USA. 96 : 3653-3657.
8. Missler, M. Südhof, T. C. *Neurexins : three genes and 1001 products* (1998) Trends Genet. 14, 20-26.
9. Koonin, E. V., Mushegian, A. R., Bork, P. (1996) *Non-orthologous gene displacement*. Trends Genet. 12 : 334-336
10. Monod J. (1970) *Le hasard et la nécessité.* Editions du Seuil, Paris. p108.
11. Ackoff R. L. (1957) *Eléments de recherchje opérationnelle*, Dunod, Paris.
12. Rapoport (1968) *General Systems Theory*. International Encyclopedia of the Social Science, vol 15. The Free Oress, New York.
13. Von Bertalanffy L. (1973) *Théorie Générale des Systèmes*. Dunod, Paris.
14. De Saussure F. (1931) *Cours de Linguistique Générale*. Payot, Genève.
15. Morin E. (1977) *La Méthode. I. La nature de la nature*. Editions du Seuil, Paris.
16. Le Moigne J-L. (1994) La Théorie du Système Général. Théorie de la Modélisation. PUF, 4e édition, Paris.
17. Wiener (1962) *Cybernétique et Société*. Union Générale d'Editions, Paris.
18. Maturana H.(1974) Stratégies cognitives. L'Unité de l'Homme 2 : Le Cerveau Humain. (E. Morin, M. Piattelli-Palmarini), Editions du Seuil, Paris.
19. Atlan H. (1974) *On a formal definition of Organization*. J. Theor. Biol., 45 : 1-9.
20. Le Moigne J-L. (1993) *Formalism of Systemic Modeling*. Some Physicochemical and Mathematical Tools for Understanding of Living Systems. Greppin, Bouzon, Degli-Agosti Editors, University of Geneva.
21. Le Moigne J-L. (2000) Dictionary of History and Philosophy of Sciences. PUF, Paris.
22. Roux-Rouquié M. Chauvet M-L Munnich A. and Frézal J. (1999*). Human genes involved in chromatin remodeling in transcription initiation, and associateddiseases : An overview using the GENATLAS database.* Molecular Genetics and Metabolism. 67 : 261-277.
23. Bravais E. Roux-Rouquié M. et al. (2000) *The GENINTER prototype : Logic scheme,*

controlled vocabulary and querying. Manuscript in preparation.
24. Capponi C. Page M. Bravais E. and Roux-Rouquié M. (2000) *GENINTER, a database dedicated to the compilation of interactions among genes and gene products.* JOBIM.

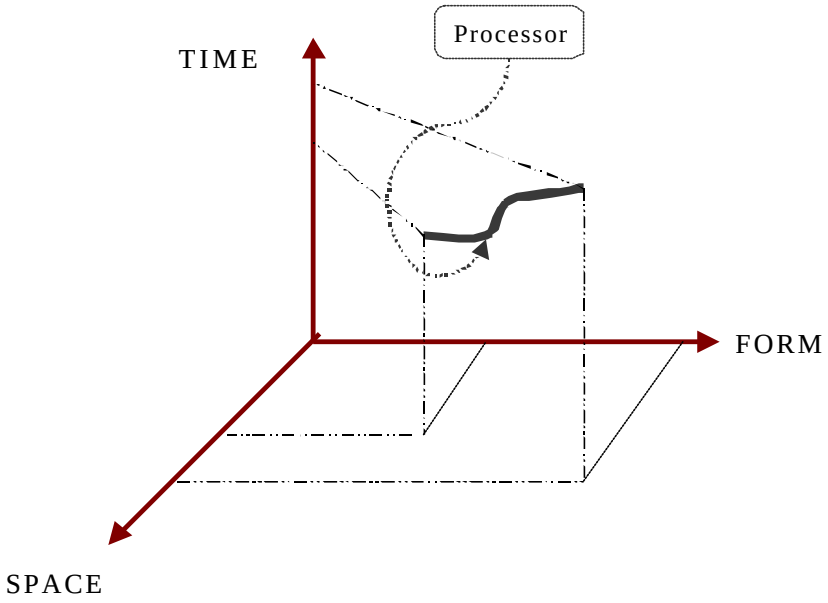
ACKNOWLEDGMENTS.

We thank Jean Louis LE MOIGNE for invaluable discussions and François KOURILSKY for encouragement and advice.

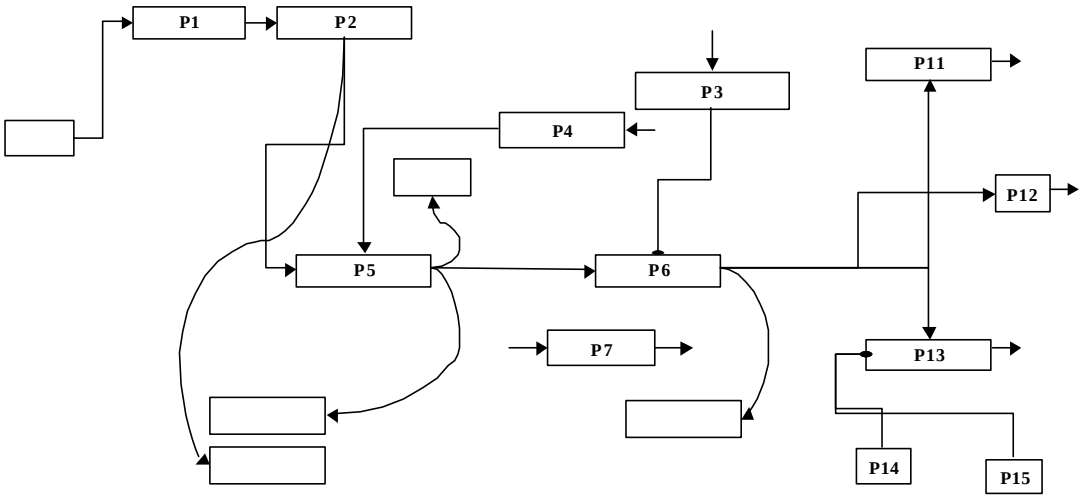
A


Functions
What
Goal-directedness
Why
Structure
Environment
Where/When
Transformations
How

B


Processor
TIME
FORM
SPACE

C



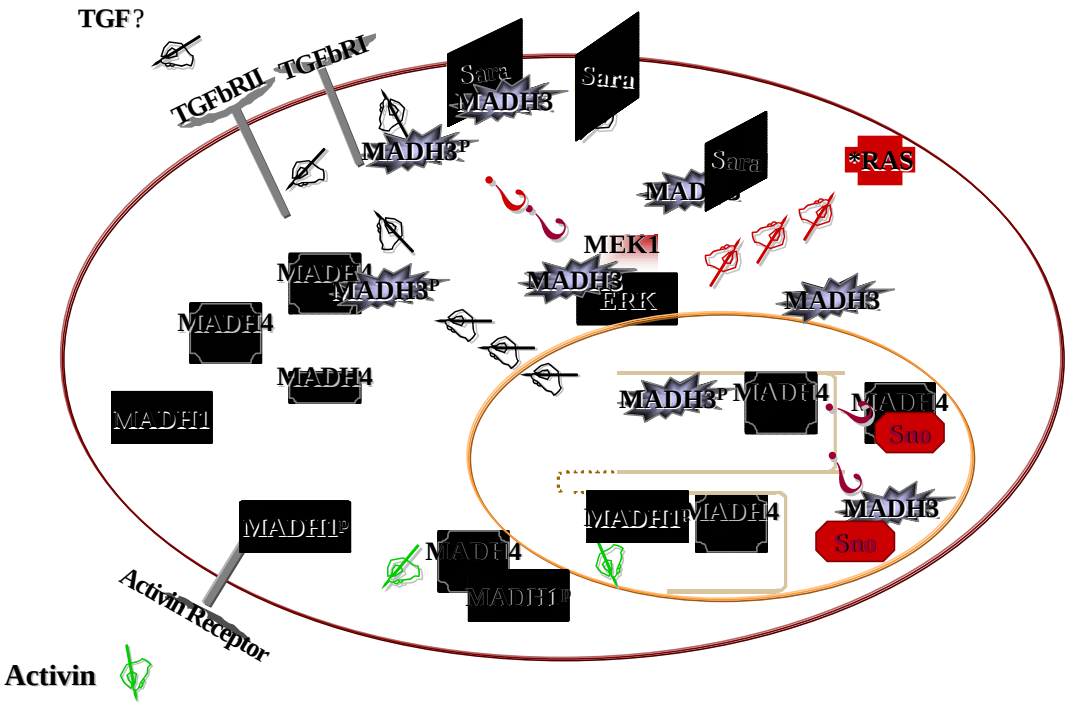
D

INPUT

OUTPUT	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
P1		1													
P2					1										
P3						0									
P4					1										
P5						1									
P6											1	1	1		
P7															
P8															
P9															
P10															
P11															
P12															
P13															
P14													0		
P15													0		

FIGURE 1. The formalism of complex system modeling.
(A) Canonical Model of the General System ; (B) Canonical Form of Process (C) Graph of Processor Network : → activated interrelationship, ⊸ prohibited interrelationship, empty box : gate-element ; (D) Structural Matrix (see text).

A



B

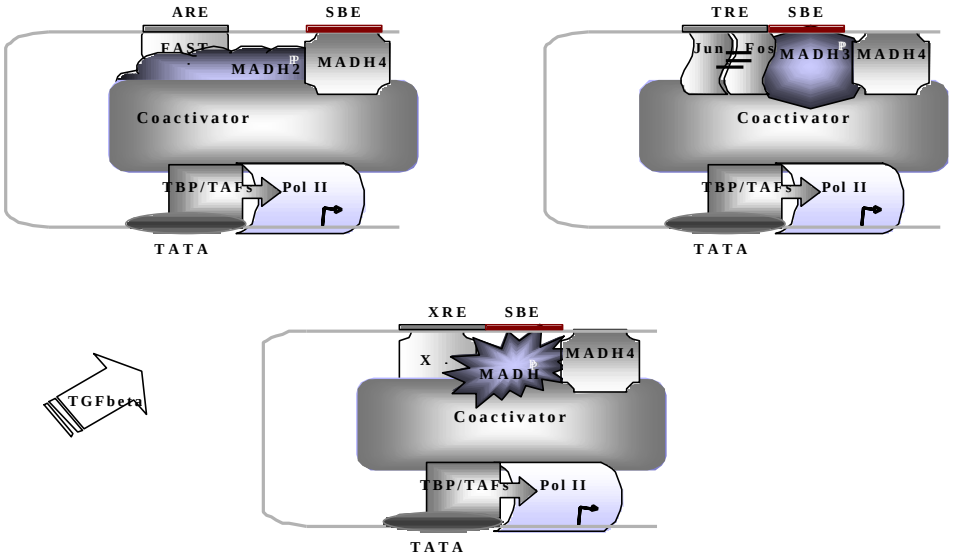
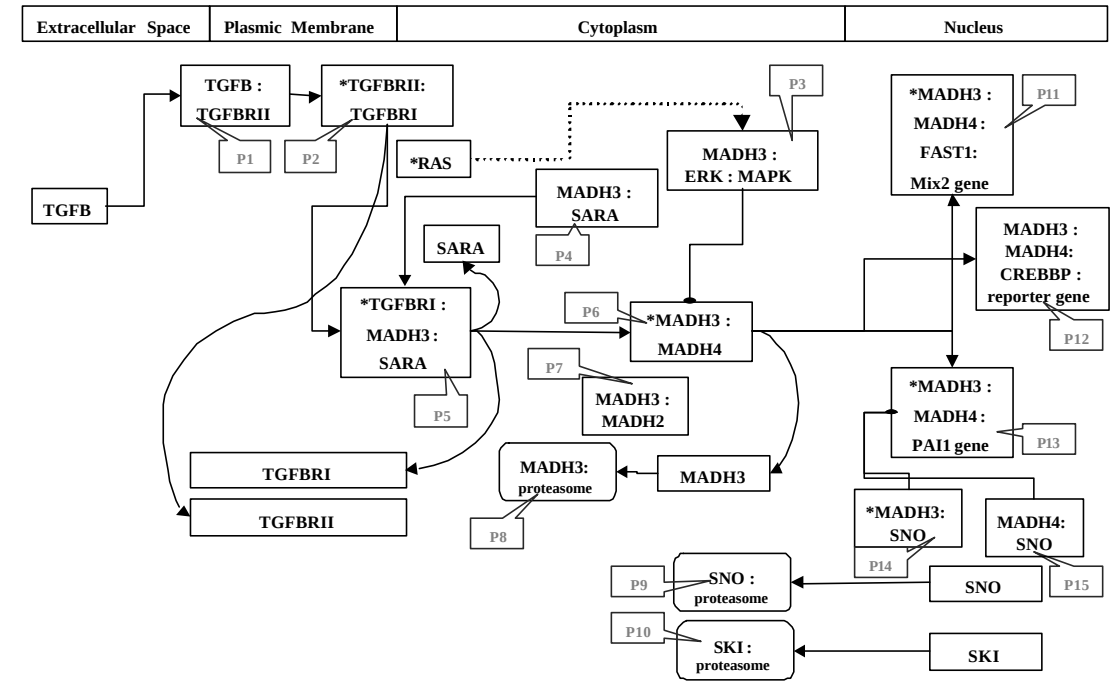


FIGURE 2.
Artwork representation of TGF?/activin signaling organization (simplified)
(A) see text ; (B) SBE : SMAD (MADH) Binding Element, ARE : Activin Response Element, TRE : TPA- Response Element (AP1 binding site); XBE : promoter element that binds transcription factor X.

A



B

INPUT

OUTPUT	P1	P2	P3	P4	P5	P6	P7 P8	P9	P10	P11	P12	P13	P14	P15
P1 P2		1			1									
P3						0								
P4					1									
P5						1								
P6 P7										1	1	1		
P8														
P9														
P10 P11														
P12														
P13														
P14												0		
P15												0		

FIGURE 3. (A) TGF?-dependant interrelationships
(→ activated interrelationship, ⊸ prohibited interrelationship : the targeted interrelationship cannot occur , gate-elements : see legend in figure 1C)
(B) matrix representation.

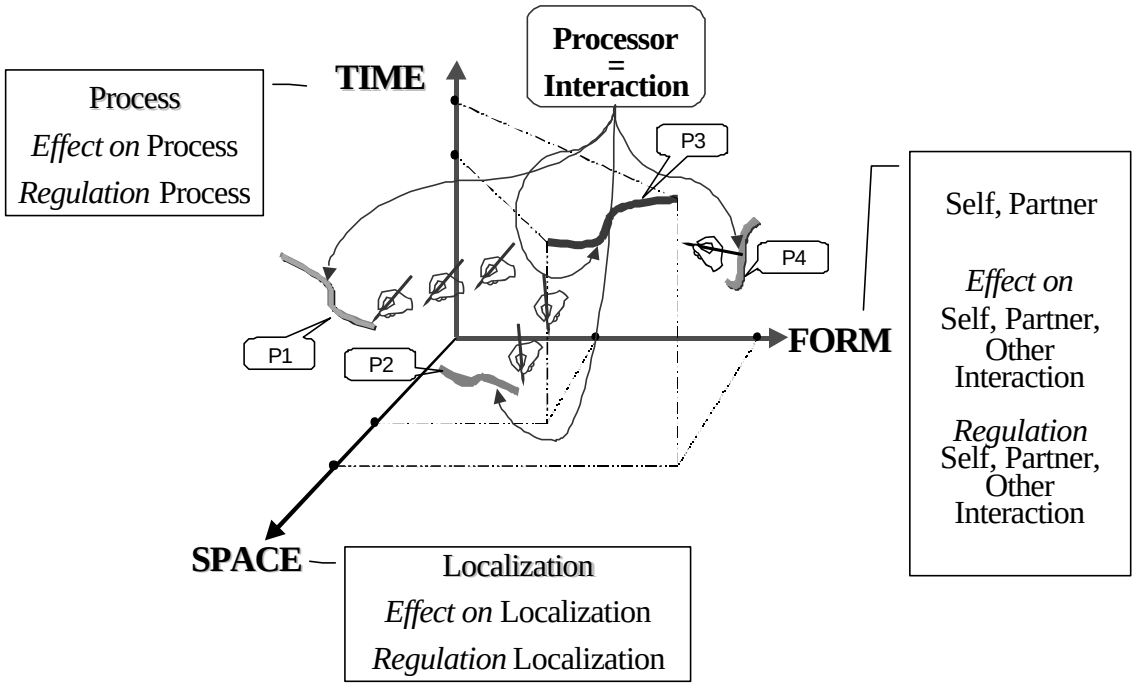


FIGURE 4. Systemic Modeling of Interrelationships between genes and/or gene products.
Processors P1 and P2 regulate P3, P3 has effect on P4.

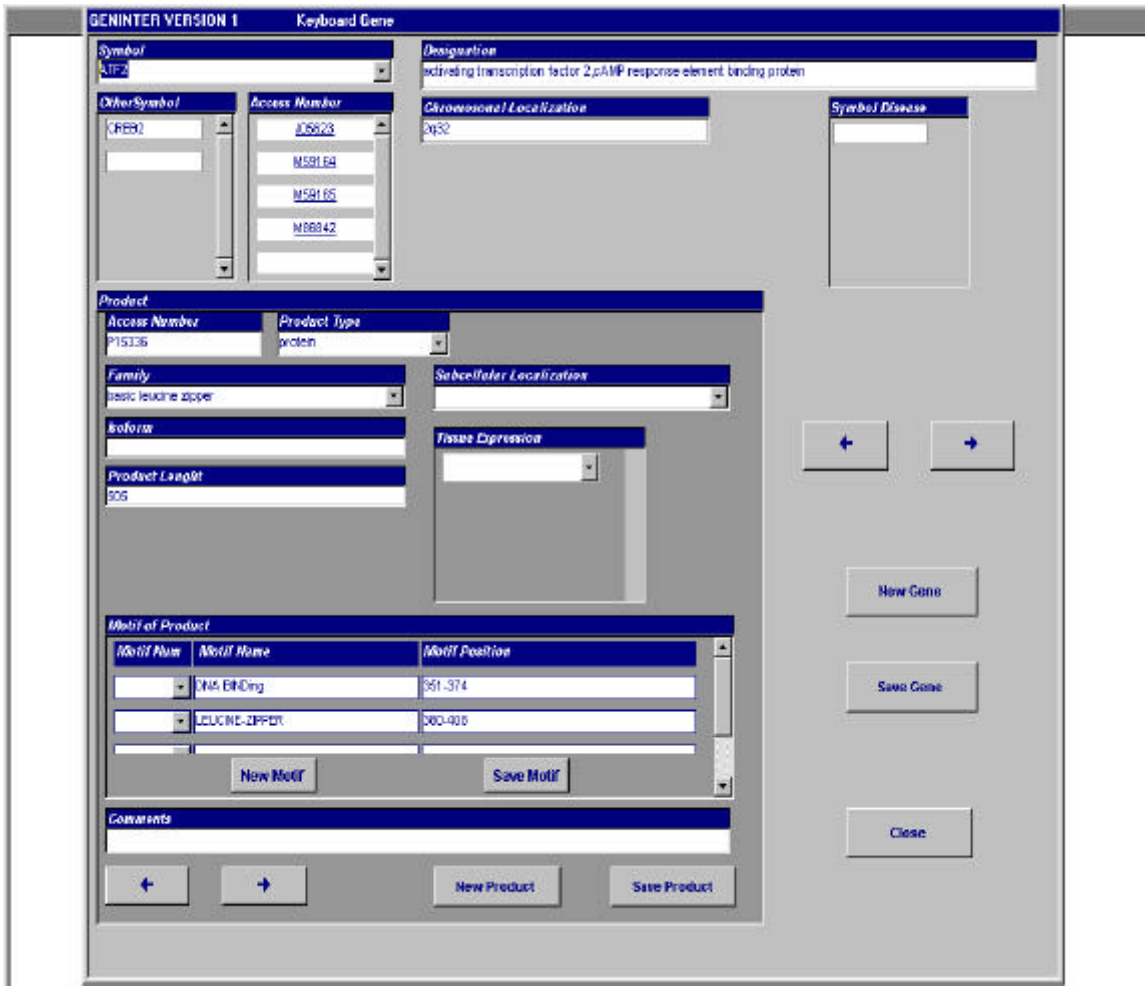


FIGURE 5. Keyboard Gene Form in GENINTER (see text for comments)

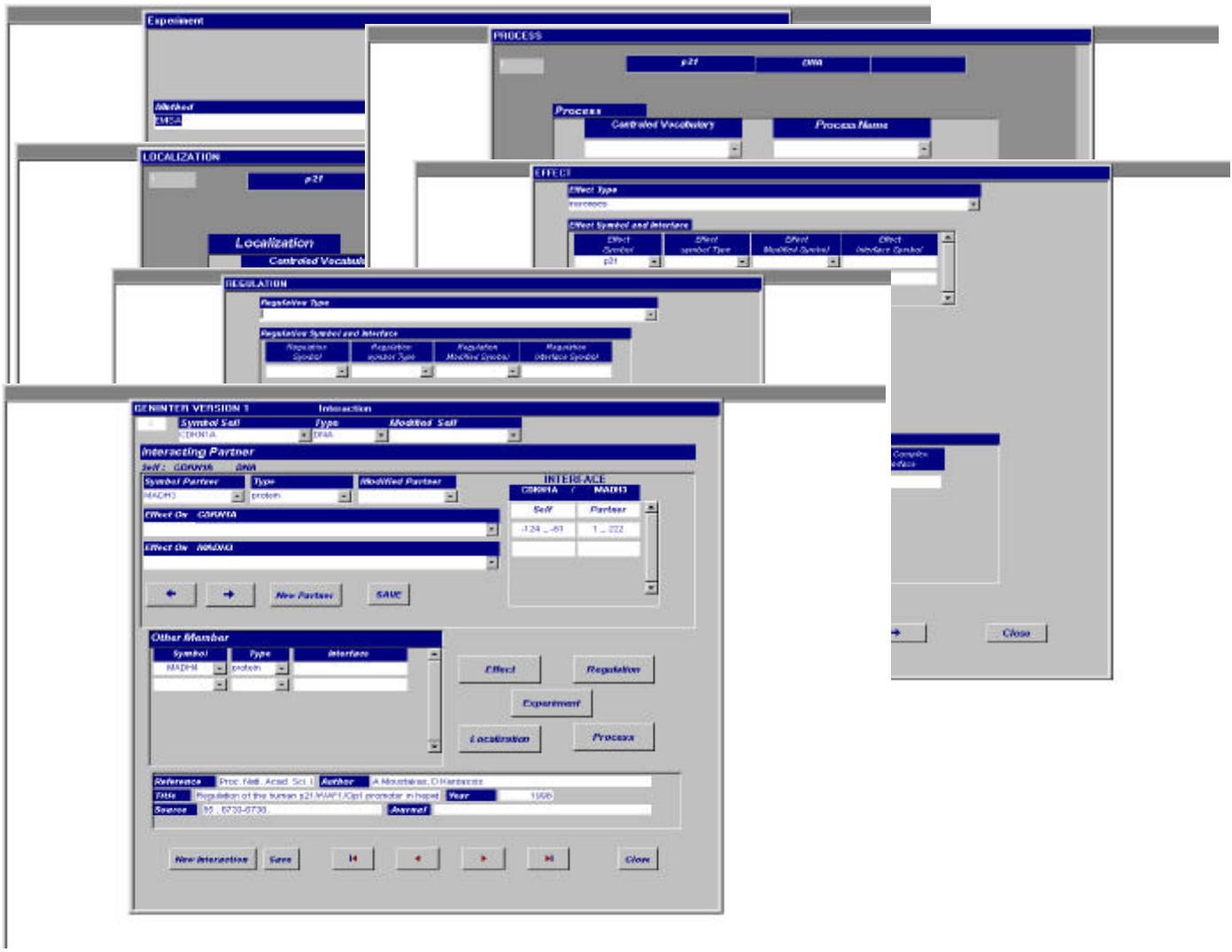


FIGURE 6. Interaction Form in GENINTER.