Towards the modeling of cell communication and computation using the *shape algebra* of biopolymers*

Sungchul Ji[1] and Gabriel Ciobanu[2]
[1]*Department of Pharmacology and Toxicology, Rutgers University, Piscataway, N.J. 08855, U.S.A. (sji@eohsi.rutgers.edu),*
[2]*School of Computer Science, National University of Singapore, Singapore (gabriel@comp.nus.edu.sg).*

Abstract

One of the critical components of the mechanisms of cell communication and cell computing has been postulated to be space- and time-organized teleonomic (or goal-directed) *shape changes* of biopolymers that are driven by exergonic chemical reactions catalyzed by biopolymer *shape changes* themselves. The generalized Franck-Condon principle resolves the apparent paradox arising when a coupling is attempted between endergonic biopolymer shape changes and the exergonic chemical reactions that are catalyzed by biopolymer shape changes themselves. The agents that induce *shape changes* in biopolymers are referred to as *shape changers,* and *conformons*, defined in 1972-1985 as sequence-specific conformational strains resident in biopolymers, have been identified as *shape changers* of biopolymers. Given a set of space- and time-organized teleonomic shape changes of biopolymers driven by shape changers, conformons, all of the functions of the cell can be accounted for in molecular terms at least in principle. To convert a conceptual model of the cell into a computer model of cell communication and cell computation, it is necessary to encode the conceptual model into a formal model represented in terms of an algebraic language. To this end, we used the process algebra developed by R. Milner et al in order to construct what is here referred to as "shape algebra" that is capable of describing complex and mobile patterns of interactions that occur among biomolecules inside the cell, including proteins and DNA, leading to cell functions.

Keywords: Molecular logic; Cell communication; Cell computing; Cell language; Structural complementarity; Space- and time-dependent shape changes of biopolymers; Teleonomic shape change(r)s of biopolymers; Conformons; Generalized Franck-Condon principle; Labeled deductive systems; Molecular labeled deductive systems; Process algebra; Shape algebra

**This contribution is dedicated to the beautiful memory of SJ's mother, Ms. Bok Nyo Keh, who passed away in Atlantic City, N.J. on September 2, 2002 at the age of 86.*

---

1. Introduction

Modeling cell functions on the molecular level using a formal language has never been attempted before, to the best of our knowledge. A conceptual model of the cell and associated molecular theories (see Table 7 below) have been proposed during the past three decades, but no quantitative computational model of the cell is available as yet. To accomplish such an audacious task, it is deemed mandatory to first develop a *formal language* that is capable of representing the essential features of the complex molecular interactions underlying cellular communication and computation. The main objective of this contribution is to review some of the key concepts and theories that have been developed in constructing the conceptual model of the living cell known as the *Bhopalator* (Ji, 1985, 1991, 2002b), describe the role of the space- and time-dependent shape changes of biopolymers (e.g., proteins, DNA) in coupling various chemical and physical processes underlying cell functions, and discuss our initial version of the "shape algebra" of biopolymers that is aimed at capturing the essence of molecular transformations and interactions that drive cell communication and computation.

2. Modeling relations

According to R. Rosen (1991), a modeling relation is said to exist between a natural system, N, ruled by *causality* and a formal system, F, ruled by logical *inference*, if the following conditions are satisfied:

1) N exhibits orderliness due to *causal* entailment.
2) F exhibits orderliness due to *inferential* entailment.
3) There exists an *encoding* dictionary with which events or phenomena of N can be mapped into propositions of F, leading to the formulation of *hypotheses.*
4) There exists another dictionary that enables *decoding* of propositions of F into phenomena or events of N, leading to the formulation of *predictions.*
5) Direct observations of events or phenomena in N entailed by causality *coincide* or *agree with* the predicted events or phenomena in F entailed by logical inference.

A modeling relation between causal entailment in a natural system and syntactic (or inferential) entailment in a formal system is considered by Rosen to be a concrete embodiment of what he calls *Natural Law,* which makes the following two assertions:

1) The succession of events or phenomena in nature perceived by humans is not entirely arbitrary or whimsical but manifests causal relations.
2) The relations between phenomena can, at least in part, be grasped by the human mind.

The similarity recently uncovered between the language used by the living cell (called *cell language*) and that used by humans (*human language*) (Ji, 1997, 1999a, 201, 2002a; see below) can be viewed as a concrete example of the 'modeling relation' described by Rosen (1991). This similarity connects the *mental world* (see F in Table 1) and the *cellular world* (see N in Table 1). As an

encoding dictionary from cell language to human language, an isomorphism was proposed between the duality of covalent and noncovalent interactions in cell language and the duality of first and second articulations in human language (see Table 2 and Section 5 for more detail). The same isomorphism has been used as a decoding dictionary from human language to cell language, leading to the prediction that DNA embodies three kinds of genes – *lexical*, *grammatical*, and *semantic genes* (Ji, 1999a, 2002a). The first corresponds to structural genes; the second, to the set of physicochemical principles and laws operating throughout the DNA molecule complexed with DNA binding proteins; and the third is postulated to reside in the noncoding regions of DNA and control its space- and time-dependent folding patterns, resulting in the control of the timing of gene expression.

Table 1.
The modeling relation between cell and human languages.

	Natural (N)	*Formal (F)*
System	Molecular language[1] Molecular computing[2]	Symbolic language Symbolic computing
Entailment	Causality: Physical laws Evolutionary rules (*Molecular logic*)	Inference: Logical rules Intuitions (*Human logic*)

[1]Synonymous with cell language, since cells use molecules as signs in communication.
[2]Synonymous with cell computing, since cells use molecules as signs in signal transduction and computation.

3. Cell language

In order for multicellular organisms to be able to maintain their life under given environmental conditions, it is necessary for their component cells to carry out two basic processes: (i) Communication with one another by exchanging appropriate messenger molecules such as neurotransmitters, hormones, growth factors, cytokines, etc, and (ii) execution of the genetic programs encoded in DNA called for, or triggered by, the information received through messenger molecules. The former constitutes *cell communication* and the latter *cell computation*. The system of molecules and their rules of interactions that are necessary for cell communication and cell computation is referred to as *cell language* and the rules embedded in it as *cell logic or molecular logic*, depending on whether the emphasis is on the cell as the smallest unit capable of utilizing cell language or on the molecularity of the components of such a language.

Just as human language is needed for humans to communicate with one another, it is here postulated that cells need a language of their own in order for them to communicate (Ji, 1997, 1999a, 2001, 2002a), to compute, and to construct. Humans have developed specialized systems of signs known as mathematics and formal languages for computation. But cells appear to use the same system of molecular signs and associated rules for the purpose of not only communication but also for molecular computation

(defined as the processing of information according to genetic programs and input signals) and molecular construction (i.e., the production of molecules or processes). In other words, living cells use cell language to perform the triad of *communication, computation,* and *construction*:

Communication
= Exchanging information between cells via 'intercellular molecular messengers'.

Computation
= Transducing extracellular information into intracellular information via signal transduction pathways or vice versa, under the control of genetic information or programs.

Construction
= Building new molecular and supramolecular structures and processes, following the results of cell computation.

For convenience, we will refer to the above three processes as the *c triad*. We postulate that *cells cannot perform, under normal conditions, one without also performing the other two of the c* triad. This postulate may be referred to as *the irreducibility of the c triad.*

Cell language is defined as "a self-organizing system of molecules, some of which encode, act as signs for, or trigger, gene-directed cell processes" (Ji, 1997). Both human and cell languages can be represented as 6-tuples, {*L, W, S, G, P, M*}, where *L* is the alphabet, *W* is the *lexicon*, *S* is an arbitrary set of *sentences*, *G* is a set of *rules* governing the formation of sentences from words (called the *first articulation*) and the formation of words from letters (the *second articulation*), *P* is a set of *physical mechanisms* necessary and sufficient to implement a language, and finally *M* is a set of *objects* or *processes*, both symbolic and material, referred to by words, sentences, and their higher-order structures (Ji, 1999a). In Table 2 cell and human languages are compared with respect to these linguistic elements.

Table 2. A comparison between human and cell languages.

	Human Language	Cell Language
Alphabet (L)	Letters	4 Nucleotides (or 20 Amino Acids)
Lexicon (W)	Words	Genes (or Polypeptides)
Sentence (S)	Strings of words	Sets of genes expressed coordinately in space and time directed by DNA folds[a] (Sets of proteins interacting non-covalently)
Grammar (G)	Rules of sentence formation	Rules mapping DNA (polypeptide) sequences to folding patterns of DNA (polypeptide) under biological conditions[b]

Phonetics (P)	Physiological structures and processes underlying phonation, audition, and interpretation, etc.	Molecular mechanisms responsible for information and energy transfers and transactions driven by conformons[c] and intracellular dissipative structures (IDSs)[d]
Sematics (S)	Meaning of words & sentences	Gene-directed (or teleonomic) cell processes
First Articulation	Formation of sentences from words	Organization of gene expression in space and time through non-covalent interactions[e] between DNA and proteins (Space- and time-dependent non-covalent interactions among proteins and among proteins, DNA and/or RNA)
Second Articulation	Formation of words from letters	Organization of nucleotides (amino acids) into genes (polypeptides) through covalent interactions [f]

[a]Just as verbal sentences (as written) are strings of words arranged linearly in the *geometric space*, so the cell-linguistic (or molecular) sentences are visualized as strings of gene expressions arranged in the *time space*.

[b]Of all the foldings of DNA and proteins allowed for by the laws of physics and chemistry, only small subsets have been selected by evolution, forming the grammar of cell language.

[c]Sequence-specific conformational strains that carry both free energy (needed for doing work) and genetic information (needed for controlling work). Conformons are thought to provide the ultimate driving forces for all molecular processes on the microscopic level inside the cell (Ji, 1974, 2000).

[d]Space- and time-specific intracellular gradients of ions, biochemicals, and mechanical stresses (e.g., of the cytoskeletal system) that serve as the immediate driving forces for all cell functions on the microscopic and mesoscopic levels (Ji 1991, 1999b).

[e]Also called '*conformational interactions*.' These interactions involve no breaking nor forming of covalent bonds, depending only on the rotation around or bending of covalent bonds, hence implicating smaller energy changes (typically less than a few Kcal/mole) than those of covalent interactions that involve much larger energy changes (10 to 30 Kcal/mole) (Ji, 1997).

[f]Molecular interactions that involve changes in covalent bonds, i.e., changes in valence electronic configurations around nuclei of atoms in molecules.

Table 2, which is largely self-explanatory (see Ji, 1997, 2001, 2002a for more details), demonstrates a striking similarity between human and cell languages, despite their obvious differences. This unexpected finding led to the notion that these two languages are *isomorphic*, in the sense that they both obey a common set of linguistic and semiotic principles (Ji, 1997, 2001, 2002a).

4. Cell computing

Cell computing, defined as state dictated by the genetic information encoded in DNA, in response to molecular stimuli received by the cell through its receptors, is synonymous with what biologists call *signal transduction*, one of the most active contemporary research fields in molecular and cell biology. Mayer and Baltimore (1993) defined signal transduction as "*the conversion of an input signal received at the extracellar face of* transitions that the cell undergoes as *the plasma membrane into an intracellular signal that ultimately alters gene expression in the nucleus",* leading to an output signal into extracellular space. This definition is consistent with the concept of signal transduction widely adopted in the field (Orphanides and Reinberg, 2001). Thus, we can represent signal transduction schematically as follows:

	a		***b***		***c***	
Input Signal	------>	***Intracellular Signal***	------>	***Gene Expression***	----->	***Output Signal*** . . . . . (1)

where ***a*** roughly corresponds to *communication* between the cell and its environment from which the input signal originates, ***b*** to *computation*, and ***c*** to *construction*, i.e., the synthesis of molecules destined to be exported or utilized within the cell where they are made. It is interesting to note that Expression (1) has the same form as "labeled deductive systems (LDS)", a unifying framework for the study of logics and their interactions proposed by Dov Gabbay in the late 1990's (Ohlbach and Reyle, 1999). LDS can be represented as:

$$A \xrightarrow{\ f\ } B \qquad \ldots\ldots\ldots\ldots (2)$$

where *f* is "the 'reason' why A entails B" (De Queiroz and Gabbay, 1999, p.179). Applying Expression (2) to Expression (1), we may identify the entailment, A---> B, as a chemical reaction driven by Gibbs free energy and the 'reason' *f* as an enzyme or a system of enzymes catalyzing that reaction. Since enzymes embody genetic information encoded in DNA, *f* reflects the entailment structures inherent in the genetic network of the cell. If this identification is justified (i.e., if LDS can be extended from traditional logics to 'molecular logic'), it would be reasonable to refer to Expression (1) as a "molecular labeled deductive system", which may be abbreviated as "mlds". Since natural language is a member of LDS and cell language is a member of mlds, and since these two kinds of languages are isomorphic (see Table 2), it may be reasonable to postulate that LDS and mlds are isomorphic. Let us refer to this idea as "the postulate of the isomorphism between LDS and mlds".

5. Molecular sentences and cell computing

The difference between *words* and *sentences* is that the former represents ideas or objects, while the latter represents *judgments* (Martinet, 1967), which may be interpreted as synonymous with *computations*. This view seems to be in agreement with the approach taken by G. Frege (1848-1925) who applied the ma-

thematical notions of *function, arguments*, and *value* to replace the analysis of sentences in terms of *subject, predicate*, and *truth-value* of a sentence (Honderich, 1995). Just as a mathematical function can be viewed as a machine that accepts an input (i.e., as an argument) and computes an output (as a value), so a sentence can be treated as a machine that accepts an input (as either subject or predicate) and computes a truth value (as either a true or a false statement).

Word structures are 'rigid,' being determined by the linguistic group to which individual speakers belong, whereas *sentence* structures (i.e., the word order in a sentence) are flexible and readily altered by individual speakers to reflect their feelings and judgments. This fact is known as double articulation in linguistics (Martinet, 1967 pp. 26-31).

It is critically important that the linguistic analysis of biological systems and processes begin with correct identifications of the *word* and *sentence* analogs of biological systems under consideration. The guiding principle here is suggested to be the dichotomy of *noncovalent* (also called *conformational*) and *covalent* (*configurational*) interactions in molecular and cell biology, in analogy to the dichotomy of the *first* and *second* articulations in linguistics, called *double articulation*. That is, the cell-linguistic analogs of words are suggested to be associated with *covalently bonded structures*, while the cell-linguistic analogs of sentences are postulated to be associated with *noncovalently bonded structures* (e.g., protein folds; see the lower part of Table 2). Therefore, based on the principle of covalent/noncovalent dichotomy, it is suggested that individual polypeptides correspond to words, and complexes of polypeptides (involving conformational interactions within and among component polypeptides known as 'protein-protein interactions') correspond to sentences. As will be pointed out below, a given signal transduction pathway (frequently represented by a set of arrows connecting input signal to gene expression, as in Figure 2 in Orphanides and Reinberg, 2002) can be treated as a system of one or more cell-linguistic sentences, or one or more processes of molecular computations.

Not all the possible combinations of letters and words that are allowed for by the principle of double articulations actually occur in human language. Ferdinand de Saussure (1857-1913) recognized two kinds of constraints, called 'syntagm' and 'paradigm' (Culler, 1991), which limit the variety of well-formed sentences. A *syntagmatic* relation refers to the relation between units that combine to form linguistic sequences such as words and sentences. For example, in a declarative English sentence, a noun phrase is followed by a verb phrase, which is in turn followed by two consecutive noun phrases, if the verb involved is a dative verb:

Subject —> ***Dative Verb*** —> ***Indirect Object*** —> ***Direct Object*** . . . . . .(3)

where the arrows indicate the temporal sequence in which a speaker utters the linguistic units.

A paradigmatic relation is the relation that holds between a particular unit in a given syntagm and other units that can substitute for it in the syntagm. The paradigmatic relations can be represented in a tabular form as shown below:

Table 3. Paradigmatic substitutions in sentences.

	Subject	Dative Verb	Indirect Object	Direct Object
Sentence #1	*I*	*gave*	*her*	*a flower.*
Sentence #2	*He*	*gave*	*me*	*a pencil.*
Sentence #3	*She*	*sent*	*him*	*a letter.*

The syntagmatic relation in signal transduction pathways can be identified with the temporal patterns of protein-protein interactions underlying, for example, the MAP kinase pathway shown below (Marshall, 1994):

Receptor —> MAPKKK —> ***MAPKK*** —> ***MAPK*** —> ***TF*** —> ***DNA*** . . . . (4)

where the arrows indicate activation processes, ***MAPKKK*** *is* mitogen-activated protein kinase kinase kinase, ***MAPKK*** is the substrate for MAPKKK, ***MAPK*** is the substrate for MAPKK, ***TF*** is transcription factor serving as the substrate for MAPK, and ***DNA*** is the region of DNA that is activated by the signal transduction pathway. The obvious similarity between Schemes (3) and (4) may be considered as an indirect evidence for the claim that cell and human languages are isomorphic (Ji 1997).

The paradigmatic relations in the MAP kinase pathway that obtain among signal transducing proteins is summarized in Table 4:

Table 4. Paradigmatic relations among signal transducing proteins in the MAP kinase cascade. Data from (Marshall, 1994).

	MAPKKK	MAPKK	MAPK
Nemtode, Drosophila, Vertebrates	*raf*	*MAPKK*	*MAPK*
S. pombe mating response	*byr2*	*byr1*	*spk1*
S. cerevisiae mating response	*STE11* *BCK1*	*STE7* *MPK1/2*	*FUS3, KSS1* *MPK1*

Osmotic regulation	?	*PBS2*	*HOG1*

It is noteworthy that both cell types and cell functions (see the 1st column) determine the nature of paradigmatic substitutes (see the elements of Table 4, which represent various proteins). Conversely, it may be stated that the nature of signal transducing proteins constituting a signal transduction pathway determines the meaning (semantics) of signal transduction (see the 1st column).

As indicated in Table 1, it is postulated that there exists a modeling relation between *cell computing* and *symbolic computing,* as there is between *cell language* and *symbolic (or human) language* (see the 1st row). Since human logic underlies both human language and symbolic computation (Akama, 1997; Havas, 1992), it is reasonable to assume that there exists cell (or molecular) logic that supports cell language and cell computing (see the 2nd row). In the next section, the concept of molecular logic is formulated.

6. Molecular (or cell) logic

It is convenient to describe *molecular logic* utilizing Aristotle's doctrine of four causes:

Material Cause. The living cell is postulated to be the smallest autonomous material system that can implement logical processes such as molecular computing (Ji, 1999b), just as the brain is the smallest anatomical unit that can perform human reasoning. The chemical processes that drive logical processes in the cell divide into two classes – exergonic (i.e., Gibbs free energy-releasing) *chemical reactions* that provide free energy necessary for molecular logical activities, and *biopolymer shape changes* that provide control over such chemical reactions. These two kinds of processes mutually affect each other and are tightly coupled, just as matter and spacetime affect each other in general relativity (Smith and Welch, 1991; see the cell force/gravitational force analogy described on pp. 114-119 in Ji, 1991). We can visualize this coupling using the metaphor of mechanical gears. Exergonic chemical reactions constitute *chemical reaction gears,* and the biopolymer shape changes accompanying catalyzed chemical reactions constitute *biopolymer gears*. The motion of one kind of gears is obligatorily coupled to the motion of the other kind. We have begun to express such a coupling between chemical reactions and biopolymer shape changes using process algebra (see Section 8, and Ciobanu and Ji, 2002). Unlike macroscopic mechanical gears whose shapes are fixed in space and time, biopolymer gears are deformable and hence can assume different shapes in space (due to variations in local pH, ionic strength, ionic compositions, and local mechanical stresses and strains) and time (due to temporal control exerted by genetic information encoded in monomer sequences) in the process of effectuating genetically determined effects or goals. Thus, the material cause of molecular logic can be characterized as goal-directed (or teleonomic) space- and time-dependent (or - organized) shape changes of biopolymers (proteins, RNA & DNA) that both catalyze and are driven by exergonic chemical reactions.

Formal Cause. Molecular logic can be viewed as a system of *rules* (selected by evolution) and the *laws* of physics and chemistry that is obeyed by space- and

time-dependent shape changes of biopolymers in the course of effectuating teleonomic molecular processes inside the cell. Therefore, goal-realizing, space- and time-dependent biopolymer shapes are 5-dimenional, the first three degrees of freedom specifying geometry, the fourth specifying time, and the fifth reflecting an internal degree of freedom that is needed to specify gene-controlled, biologically meaningful processes. This fifth dimension encodes biological *information*, a parameter unique to living systems.

Efficient Cause. The fundamental mechanism underlying molecular logic is postulated to be *nonstochastic* (also *called goal-directed*, *informational*, or *teleonomic*) space- and time-dependent shape changes of biopolymers driven by exergonic chemical reactions that biopolymers themselves catalyze. Given teleonomic space- and time-dependent biopolymer shape changes, all of the living processes inside the cell can be completely accounted for in a manner consistent with the laws of quantum and statistical mechanics (Ji, 1991, 2000, 2002b). But there is subtle difficulties to overcome here: Since the free energy that drives biopolymer shape changes cannot be released from exergonic chemical reactions until biopolymers actually catalyze those reactions which in turn require their endergonic shape changes, where do biopolymers obtain free energy to drive their initial endergonic (Gibbs free energy consuming) teleonomic space- and time-dependent shape changes needed for catalysis? One solution to this dilemma was proposed in (Ji, 1974), based on the "generalized Franck-Condon principle", according to which biopolymers can transiently "borrow" thermal energies from their environment to catalyze exergonic chemical reactions. These thermal energies are then paid back to the environment, within time ϑ, the turn-over time of enzymes (MacClare, 1971; Ji, 2000) in order not to violate the second law of thermodynamics. As long as enzymes do not retain thermal energies longer than, or equal to, ϑ, it can be shown that the second law is not violated by such transient borrowing by biopolymers of thermal energies for their catalytic actions (MacClare, 1971; Ji, 1974, 2000).

Final Cause. The final cause of human logic is to assure correct inferences so as to reach truth. Analogously, it is suggested here that the final cause of molecular logic is to correctly implement genetic programs so as to contribute to the survival of the cell. It is possible that the goals of human reasoning, e.g., the search for truth, have emerged, through biological evolution, from the goal of molecular reasoning assuring the survival of the cell or the maintenance of life.

Molecular logic thus characterized can be compared with *human logic* at 10 different levels as shown in Table 5:

(1) *Logic Machine*: The brain is the anatomical unit executing logical thinking. Similarly, it is assumed that the cell is the smallest material system capable of molecular logical processes (Ji, 1999b).

(2) *Signs Used*: Charles S. Peirce (1839-1914), the father of modern *semiotics* (the study of signs) and a pioneer in modern logic, distinguished three categories of signs (often defined as *something that stands for something else*) based on the nature of the relation between a sign and its referent. Thus, portraits are *iconic signs* (due to similarity), smokes are *indexical signs* for fire (due to causality), and

written marks are *symbolic signs* having arbitrary relation to their referents (due to convention). Signs used in human logic are *symbolic signs* but those used in molecular logic are simultaneously symbolic (as carriers of *information*) and indexical (as carriers of *free energy*).

(3) *Purpose*: As already indicated, the purposes of human logic and molecular logic are search for truth and survival, respectively.

(4) *Syntax:* The grammar of human logic is "formation rules" and that of molecular logic can be identified with the laws of physics and chemistry whose implementation is guided by teleonomic information stored in biopolymer structures.

(5) *Lexicon*: The lexical units of human logic are elements of the axiom system, while those of molecular logic can be identified with molecules acting as symbolic signs such as hormones, cytokines, and genes.

(6) *Sentences*: Sentence equivalents of *human logic* are known as *well formed formulas,* and those of *molecular logic* can be identified with *noncovalent complexes of biopolymers* based on postulates of the cell language theory (Ji, 1997, 2001, 2002a). Noncovalent complexes are also called *metabolons* (Srere, 1987), *modules* (Hartwell, et. al., 1999), and *hyperstructures* (Norris, et al., 1999),

(7) *Context/Background*: The context or background which makes human logical reasoning possible is human language. Similarly, the context or background which makes molecular logic possible may be cell language, a self-organizing system of molecules acting as information and/or energy carriers effectuating molecular computing and communication (Ji, 1997, 2001, 2002a).

(8) *Limit*: G∠del's theorem provides a limit to human logic. Thermal fluctuations and the Heisenberg uncertainty principle limit the accuracy and reliability of molecular logic (see the *biological uncertainty principle* on p. 118 in Ji, 1991).

(9) *Scale*: The scale of human logic (as measured by the size of signs used) is *macroscopic* while that of molecular logic is *microscopic.* This difference may have profound consequences in biology and philosophy (Ji, 2001, 2002a).

(10) *Complementarity*: The principle of complementarity first enunciated by N. Bohr (Pais, 1991) may be operative in both human and molecular logics. Human logic is characterized by the mind/body complementarity, in agreement with the thoughts of Aristotle (384-322 B.C.), Spinoza (1632-1677), and Merleau-Ponty (1907-1961) (Dillon, 1997). In contrast, molecular logic embodies the information/energy complementarity (Ji, 1995; see also the *von Neumann-Pattee principle of matter-sign complementarity* in Ji, 1999a). The essential points of the comparison between human and molecular logics are collected in Table 5.

Table 5. A comparison between human logic and molecular (or cell) logic.

Parameter	**Human Logic**	**Molecular Logic**
1. *Machine*	The Human Brain	The Cell
2. *Signs Used*	Abstract Symbols (acting as Symbolic Signs)	Molecules (acting both as Symbolic and Indexical Signs)
3. *Purpose*	Correct Inference leading to Truth	Correct Implementation of Genetic Programs leading to Life
4. *Syntax*	Formation Rules	Laws of Physics and Chemistry implemented under the constraints of teleonomic Information encoded in biopolymers
5. *Lexicon*	Elements of Axiom System	Informational Aspect of Individual Biopolymers (in contrast to the energetic aspect)
6. *Sentences*	Well Formed Formulas	Noncovalent Complexes of Biopolymers
7. *Context*	Human Language	Cell Language
8. *Limits*	G∠del's Theorem	Thermal Fluctuations Heisenberg Uncertainty Principle
9. *Scale*	Macroscopic	Microscopic
10. *Complementarity*	Mind/Body	Information/Energy

The hypothetical entity exhibiting the properties of *information* and *energy* as its complementary aspects is known as *gnergy* (Ji, 1991, 1995, 2000, 2002a,b). The discrete units of gnergy are called 'gnergons', the best studied example of which being *conformons,* defined as sequence-specific conformational strains of biopolymers that carry both free energy (to do work) and genetic information (to control work) (Ji, 1974, 1991, 2000). Available data indicate that the average conformon in proteins carries 8-16 Kcal/mole of free energy and 40-200 bits of information (Ji, 2000; Benham, 1996). The corresponding values for the average conformon in DNA have been estimated to be 500-2500 Kcal/mole and 200-600 bits of information, respectively. According to the conformon theory of molecular machines, chemical reactions produce conformons, which then drive

shape changes of biopolymers in space- and time-dependent manner, leading to biological or teleonomic functions (Ji, 2000; Ciobanu and Ji, 2002).

All molecular machines inside the cell, from simple enzymes to complex ones such as ion pumps and molecular motors are driven by conformons produced by exergonic chemical reactions catalyzed by biopolymers. Since biopolymer shape changes are intrinsic to any operational cycle of molecular machines, they too must be driven by conformons. Thus the sequence of events from exergonic chemical reactions to endergonic teleo-nomic functions of biopolymers can be represented as follows:

Chemical Reactions → *Conformons in*

Biopolymers → *Biopolymer Shape*

Changes → *Teleonomic Functions* . . . (5)

Therefore it follows that biopolymer shapes also exhibit two complementary aspects, namely, *informational* and *energetic/material*, in agreement with the gnergy principle which states that the ultimate reality is a complementary union of information and energy (Ji, 1995).

Table 6. Biopolymer shapes as carriers of information and energy, or gnergy.

Gnergy	*Molecular Shape*
Information	*Nonstochastic* (i.e., rule-governed) *behaviors*, e.g., timing of catalysis and ligand binding interactions controlled by biopolymers.
Energy	*Stochastic* (i.e., law-governed) *behaviors,* e.g., enzyme catalyzed chemical reactions, and ligand binding interactions driven by Gibbs free energy decreases.

7. Interactions between complementary shapes of biopolymers in cell communication and computation

The heart of human language is *categorization*, by which a very large number of *cases* is classified into a much smaller number of groups known as *words* (Ellis, 1993). Analogously it is postulated here that the most basic molecular process underlying cell language needed for cell communication and computation is the ability of biopolymers to stereoselectively interact with (or recognize) their ligands through *complementary binding*. A similar idea was recently expressed by Bar-Ziv et al. (2002). Human language is built on a mapping of *words* to *cases.* It seems that cell language is based on a mapping of *biopolymer shapes* to their *complementary ligands* (see Expression (11) below). So, *categorization* in human language is analogous to *complementary binding* in cell language. It should be noted here that a ligand binding to a biopolymer could be another biopolymer, small molecular weight biochemicals (e.g, ATP), or inorganic ions (e.g., K^+, inorganic phosphate, etc.).

Complementary bindings between biopolymer shapes so basic to cell language are predicated on two necessary conditions, which together constitute a sufficient condition:

i) *The* ***principle of structural complementarity.*** The structures of the binding sites of a biopolymer and its ligand must be *complementary* to each other like a key in a lock or a male and a

female (abstracted into the notion of the Yin and Yang doctrine in Taoist philosophy), and

ii) The ***principle of sufficient binding force***. The complementary binding surfaces of a biopolymer and its ligand must be sufficiently attractive to each other to result in a significant dissipation of Gibbs free energy upon binding under physiological conditions. Otherwise thermal fluctuations will prevent the binding, despite the complementarity between the shapes of the binding sites involved.

We can represent the complementary binding interaction between a biopolymer A and its ligand B (which can be assumed here to be a small molecular weight species without losing generality) as follows:

$$A + B <----> A'/B' \quad \ldots\ldots\ldots (6)$$

where A' and B' represent the shapes of A and B after binding (indicated by the slash symbol) which differ from the original shapes of A and B, due to the conformational deformations experienced by A and B upon mutual binding. Please note that Process (6) is written as a reversible reaction but can be made to proceed irreversibly in either direction, depending on the sign and magnitude of the Gibbs free energy change accompanying (6). The following facts must be kept in mind when considering Process (6):

i) Due to the restriction placed by the Generalized Franck-Condon principle (Ji, 1974, 2000), A and B cannot bind directly but must first undergo reversible conformational transitions to A' and B', respectively, through thermal fluctuations, before a productive binding interaction can take place between A' and B'. Strictly speaking, this so-called Franck-Condon mechanism (Ji, 1974, 2000) contradicts the main tenet of the 'induced-fit hypothesis' of Koshland (Stryer, 1995) that conformational changes of A and B to A' and B', respectively, follow, rather than precede, the binding interactions. We can represent the Franck-Condon mechanism of binding interactions in terms of the following three component processes:

$$A <-----> A' \quad \ldots\ldots (7)$$

$$B <-----> B' \quad \ldots\ldots\ldots (8)$$

$$A' + B' <-----> A'/B' \quad \ldots\ldots\ldots (9)$$

Notice that the sum of Processes (7), (8) and (9) leads to Process (6). Processes (7) and (8) are often referred to as "pre-equilibration" in chemical kinetics, and Process (9) represents the binding process proper. Pre-equilibration processes are usually endergonic (i.e., free energy consuming, or the net Gibbs free energy change is positive, due to the fact that A' is less stable than A), while the binding reaction is exergonic (i.e., free energy releasing, or the net free energy change is negative, because A' and B' attract each other). If the free energy change accompanying Process (9) is much more negative (i.e., spontaneous) than the sum of the free energy changes of Processes (7) and (8) is positive (i.e., nonspontaneous), the overall free energy change of Process (6) can be negative and hence (6) can proceed spontaneously from left to right.

ii) The symbol A represents the shape of a biopolymer. But, due to the flexibility of biopolymers under physiological conditions, a biopolymer molecule constantly undergoes conformational

transitions from one conformational isomer (called 'conformers') to another. *Conformers* should not be confused with *conformons*, packets of free energy and genetic information localized in sequence-specific sites within conformers (Ji, 2000). Therefore A must be interpreted as the conformation averaged over an ensemble of n conformers, each having slightly different energies, e_i. The probability of occurrence of the i^{th} conformer, $P_i(e_i)$, is then given by the Boltzmann distribution law,

$$P_i(e_i) = g_i e^{-ei/kT} \quad \ldots\ldots\ldots\ldots\ldots \quad (10)$$

where g_i is the i^{th} degeneracy (i.e., the number of different conformers having the same energy e_i), k is the Boltzmann constant, and T is the absolute temperature (Andrews, 1963). Equation (10) states that, as the energy content of a conformer, e_i, increases, the probability of a biopolymer assuming that conformation decreases exponentially, under the condition that the degeneracy, g_i, remains constant.

One unique property of conformational isomers (i.e., conformers), in contrast to covalent isomers, is that conformers can undergo rapid transitions from one conformational state to another, due to thermal fluctuations, whereas covalent isomers are relatively stable, because they cannot be interconverted without breaking and forming covalent bonds (see Table 4 in Ji, 1997). The precise number n of the conformers available to a biopolymer is not known in most cases, but the careful investigations performed by H. Frauenfelder (1997) and others on conformational transitions of myoglobin induced by light-activated desorption of carbon monoxide from the heme iron of myoglobin indicate that the number n of biologically significant conformers may be in the range of thousands, if not millions.

iii) When we say that a biopolymer A has n biologically significant conformational states or conformers, we mean that each of the n conformers accessible to A can, if selected, participate in some biologically significant process inside the cell, such as ligand recognition/binding, catalysis, transmembrane movement of ions, and translational movement along a molecular track. Let us assume that biopolymer B mediates a biological function f_i when it undergoes a conformational transition from the i^{th} to the $(i + 1)^{th}$ conformers, B_i to B_{i+1}, that catalyzes a chemical reaction, c_i ---> c_{i+1}, driven by the Gibbs free energy change, g_i. The i^{th} elementary or atomic biological function (f_i) can then be represented as a 5-tuple as shown in Expression (11), with each term having the meaning defined in Expression (12):

$$f_i = (c_i, c_{i+1}, g_i, B_i, B_{i+1}) \quad \ldots\ldots\ldots \quad (11)$$

$$\begin{array}{c} B_i <-> B_{i+1} \\ c_i <----------> c_{i+1} \\ f_i => g_i \end{array} \quad \ldots\ldots\ldots\ldots (12)$$

Expression (12) indicates that the conformational transition from B_i to B_{i+1} is coupled to the transformation of a chemical reactant c_i to product c_{i+1} which leads to gene-directed (or teleonomic) function, f_i, with a concomitant dissipation of gibbs free energy, g_i. It is important to point out that biological function, f_i, is obligatorily coupled to dissipation of Gibbs free energy, g_i, as indicated by the unidirectional arrow shown below the double-headed arrow in (12). Most importantly, Expression (12) embodies what may be called the "molecular gear hypothesis", the essence of which being that shape changes in biopolymer B and

chemical changes catalyzed by such shape changes are obligatorily coupled, through a pair of complementary binding interactions, between B_i and c_i at the beginning of a work cycle and between B_{i+1} and c_{i+1} at the end. We can visualize such a coupling in terms of two wheels equipped with complementary sets of gear teeth that rotate in synchrony. As indicated earlier, the direction of Process (12) is entirely dependent on the sign of the Gibbs free energy, g_i, proceeding from left to right when $g_i = G_{i+1} - G_i < 0$, and in the opposite direction, when $g_i = G_{i+1} - G_i > 0$.

Finally, it should be pointed out that any observable biological functions, F, can be regarded as the sum of a set of elementary (or atomic) biological functions, f_i, that are coupled in space and time:

$$F = \varphi_{i,w} f_i \quad \ldots\ldots\ldots\ldots\ldots\ldots (13)$$

where the index i ranges from 1 to W, the total number of the atomic biological functions, f_i.

8. Shape algebra of biopolymers

The main idea of this section is to provide a formal approach able to express the interaction among molecular processes, taking care that shape-changes of biopolymers are responsible for all goal-directed molecular processes inside the cell, including cell shape changes themselves that affect many cell functions (Chen et al., 1997). An important concept offered by quantum physics to biology is the notion of complementarity (Pais, 1991), generalized in terms of *information* and *energy* (Ji, 1991, 1995, 2002b). Since interaction is so important, the authors have proposed a theory of interacting shapes, where we use the term ***shape*** to refer to a complementary union of *energy* and *information* (as light is viewed as a complementary union of particles and waves).

Thus the notion of shape is used to catch the indistinguishable *information* and *energy* at the level of functioning bioploymers in the living cell. However, the description of a cell or molecular system in terms of shapes may be difficult. Instead, it may be easier to consider an equivalent description in terms of interacting molecular (sub)processes, taking into account of the stimulus and response actions coexisting in the cell space.

Although molecular information is rapidly accumulating, it is difficult to analyze it, since it is dense, disparate, and without an appropriate formal tool. In recent years, various approaches from mathematics and computer science have been adapted for the representation of molecular processes in biology. The use of *process algebras* (pi-algebra) introduced in computer science for specification of the concurrent communicating processes is quite new. The pi-calculus is a widely accepted model of interacting systems with dynamically evolving communication topology (Milner, 1999). We think that an appropriate version of pi-calculus might be an adequate formalism for describing the biomolecular processes. As far as we know, the first papers using the pi-calculus in describing molecular processes were (Ciobanu, 2000) and (Ciobanu, 2001). In (Ciobanu, 2000), the pi-calculus is used to describe the DNA methylation, suggesting a conformational transformation. In (Ciobanu, 2001) are defined the so-called molecular structures and it is proved that they have the same expressive power as the pi-calculus (which have the same computational power as Turing machines). A quite technical approach considering shared

resources (covalent bonds and structures) is described in (Ciobanu and Rotaru, 2002). The use of the pi-calculus to describe the dynamics of the Na pump is presented in (Ciobanu, et al. 2002). The Albers-Post mechanism of the Na^+/K^+ ATPase is translated into a model which can describe molecular interactions, conformational transformations, and ion transportation occurring in the pumping process. In this Section, we briefly present a *shape algebra*, i.e. a version of pi-calculus able to describe the interaction between molecular shapes according to the principles of structural complementarity and of sufficient binding free energy discussed in Section 7. The pi-calculus that was introduced by Milner, Parrow, and Walker as an attempt to describe mobile concurrent processes (Milner, 1999) allows for dynamic reconfiguration among processes and is able to describe mobile systems, thereby providing a conceptual framework and mathematical tools. The pi-calculus has a well-defined semantics and an appropriate algebraic theory.

The computational world of shape algebra contains just *processes* and *interaction channels* (also called *shapes*). The shape algebra considers a shape as a pair (i,e) of information (i) and energy (e), and it takes interaction as a primitive. There are two basic entities in *shape algebra*: the *shapes*, named x, y, … , and *molecular processes* P, Q, … that interact through them. Also, there are two types of atomic actions, called *shape guards* or *prefixes*: the input guard x(y) is used to receive a shape for y along the channel x, and the output guard x<z> is used to send the shape z along the channel x. Interaction is established by a non-deterministic matching which dynamically binds "senders" to eligible "receivers". Even though there are many pairs that can satisfy the matching condition, only a single receiver gets the commitment of the sender. Thus processes can interact by using complementary shapes. A shape received in one interaction can be used in another; by receiving a shape, a process can interact with processes that previously could not interact because of lack of suitable (complementary) shape for interaction, but can now interact because they share the same shape (in complementary forms). Starting with atomic shapes and simpler processes, complex processes can be constructed in many ways. The process expressions are defined by guarded processes, parallel composition P|Q, non-deterministic choice P + Q, and replication !P. Over the set of processes is defined a *structural congruence relation*; this relation provides a static semantics of some formal constructions. The structural congruence deals with aspects related to the structure of the processes. In shape algebra, the evolution of a process is described by a reduction relation over processes which is called reaction. This reaction relation contains those transitions that can be inferred from a set of rules.

We present in this section a monadic version of the shape algebra, meaning that exactly one shape is communicated in an interaction. Let X be an infinite countable set of shapes. The elements of X are denoted by x,y,z … . The terms of this formalism are called processes and processes are denoted by P, Q, R … .

Definition. The processes are defined over the set X of shapes by using the prefixes

p ::= x <z > | x(y) | [x=y] p.

The processes are defined by the following grammar:

P::= 0 | p.P | P + Q | P | Q | !P.

Processes evolve by performing interactions, and these interactions are given by their prefixes p. The input and output prefixes x(y) and x<z> represent the passive and the active complementary shapes, receiving and sending a shape during an interaction. The output prefix x<z> is the active shape of type x (male) and sends a shape z; an input prefix x(y) represent the passive shape of type x (female) and waits a shape that will substitute the bound variable y. The match prefix [x=y] p.P can evolve as p.P if x and y are the same, and do nothing otherwise. 0 is the empty process. P + Q represents a nondeterministic choice of P or Q. P | Q represents the interaction of P and Q. A replicated process !P denotes a process that allows to generate arbitrary instances of P for interaction. The replication !P can be expressed by recursive equations of parametric processes as well.

The interaction x<z>.P | x(y).Q is a formal translation of a unidirectional complementary interaction between a biopolymer and its ligand described by Process (6) of Section 7. This interaction is possible according to the fact that the parts have the same (type of) shape x. An interaction is actually defined by an active part (male, sender) x<z>.P and a passive part (female, receiver) x(y).Q, and it can be represented by the following transition:

x<z>.P | x(y).Q → P | Q{z/y}.

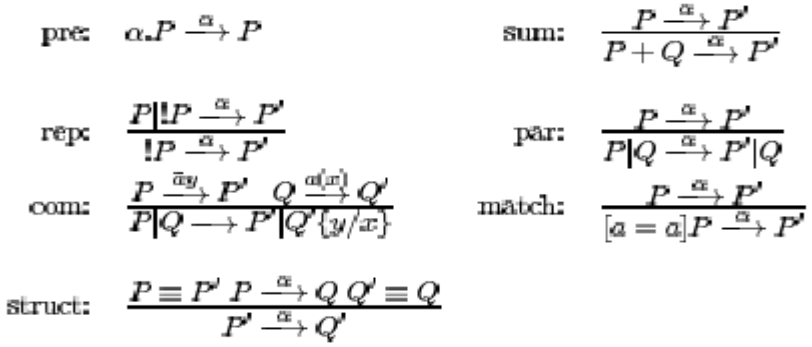
This is a synchronous interaction: an output prefix cannot interact without the simultaneous availability of an input prefix. Technically speaking, as in logic, the prefix x(y) binds the name y; we denote by fn (P) the set of the names with free occurrences in P. We denote by P{v/u} the result of simultaneous substitution in P of all free occurrences of the name u by the name v, using alpha-conversion wherever necessary to avoid name capture. A structural congruence relation is defined over the set of processes; this relation provides a static semantics of some formal constructions. We denote by $=_\alpha$ the standard alpha-conversion of the lambda-calculus (a well-known term-algebra with the same computational power as the Turing machines).

Definition. The relation ≡ over processes is called structural congruence and it is defined as the smallest congruence over processes which satisfies

[x=x] p.P ≡ p.P	P ≡ Q if P $=_\alpha$ Q	!P ≡ P	!P
P + 0 ≡ P	P + Q ≡ Q + P	P + Q) + R ≡ P + (Q + R)
P	0 ≡ P	P	Q ≡ Q	P	(P | Q)	R ≡ P	(Q | R)

The evolution of the interacting processes is described by a reduction relation over processes called reaction. This reaction relation contains those transitions which can be inferred from a set of rules. If we now consider α as a prefix indicating the shape of an interaction, then we have the following definition for the reduction relation.

Definition. The reduction relation over processes is defined as the smallest relation → satisfying the following rules:



It is very useful to be able to compare the behaviour of two systems. We introduce two behavioural equivalences. These behavioural equivalence are based on the important notion of bisimulation originating from process algebra in computer science. There are several versions of bisimilarity; one of them is called open bisimilarity. Its definition is given by using the labeled transition system defined by the reduction rules.

Definition. A relation *S* defined over processes is called an open simulation if for all P and Q, whenever P *S* Q then for all substitutions σ the following holds if Pσ → P', there exists Q' so that Qσ → Q' and P' *S* Q'. *S* is an open bisimulation if both *S* and S^{-1} are open simulations. Two processes are called open bisimilar, and we denote this by P ~ Q, if there exists an open bisimulation *S* that relates them P *S* Q.

We have many results related to this algebra. One result is interesting from a computational point of view: This algebra is able to translate the lambda-calculus. As a consequence, this interaction algebra has the same computational power as the Turing machines, i.e. it is able to express all the computable functions. Other results refer to the bisimulations (e.g. the bisimulations are equivalence relations, every process is strongly bisimilar to a summation, strong bisimulation is a congruence, weak bisimulation is a congruence, congruence properties of replication, unique bisimilar solutions of equations). These results are similar to those described by Robin Milner (Turing Award) in his recent book (Milner, 1999). Various properties of the systems described by shape algebra can be checked automatically by studying the bisimilarity between two processes, namely the model and its specification. More helpful in the verification process is a bisimilarity called *weak open bisimilarity*. It allows the basic verification technique based on temporal logic for proving properties about the concurrent interacting systems with a finite state space. This means that we can have a sophisticated software tool based on a very powerful logic called mu-calculus able to translate a description of a system given by shape algebra into a finite state transition system, and then to verify the properties of this finite state transition system. Modeling and verifying with this logic and some of its proper subsets have been thoroughly investigated in the literature in recent years (Clarke, Grumberg and Peled, 1999). Model checking of the pi-calculus processes is discussed in (Dam, 1996). In addition to this logic, the Mobility Workbench (Victor and Moeller, 1994) supports open bisimulation checking. Using these tools, (Ciobanu et al., 2002) describes the

sodium-potassium exchange pump. This molecular system is concerned not only with phenomena related to distribution, cooperation, but also with mobility and adaptability. The pi-calculus and its molecular version here called *shape algebra* provide a suitable framework whose primitive is a shape-based interaction and able to explain the coupling phenomenon of ion transport and ATP hydrolysis. The coupling phenomenon could be described in an elegant way, emphasizing the communication of the two regions of the protein (the ATP binding and phosphorylation domain and the cation co-ordination domain). Using a more realistic stochastic version of the pi-calculus. The above-mentioned paper describes the molecular interactions and conformational transformations in an explicit way. We manipulate formally the changing con-formations and describe the corresponding dynamic systems using discrete mathematics instead of the usual described in more details, step by step. Moreover, using a sophisticated software tool, it is possible to verify various useful properties of the described systems.

A detailed presentation of the shape algebra and its results requires more space than we have here; it will be presented in another paper in preparation. However we have presented the main ingredients of this algebra, and it is hoped that the reader can gain an idea about its capabilities in representing molecular interactions mediated by biopolymer shapes.

9. A historical analogy between the modeling of the cell in biology and the modeling of the atom in physics

There is a possible historical analogy that can be drawn between the atomic physics of the 19^{th}-20^{th} century and cell biology of the 20^{th}-21^{st} century as summarized in Table 7. This table is largely self-explanatory due to the extensive footnotes attached to it:

Table 7. A comparison between physics and biology with a special attention given to the role of the microarray technique in the biology of the 21^{st} century.

Parameter	*Physics*	*Biology*
Time	19^{th} – 20th	20^{th} – 21^{st}
Object of Study	**The Atom**	**The Cell**
Experimental Technique	Atomic Absoprtion/Emission Spectroscopy	Microarray Technique (1995)
Data	Atomic Line Spectra	Genomic Data Transcriptomic Data Proteomic Data Metabonomic Data Phenomic Data

Regularities	Lyman Series Balmer Series Ritz-Paschen Series Brackett Series Pfund Series	Genetic Code Gene Clusters (Eisen et al., 1995; Lee et al., 2002) Characteristic Modes (Holter et al., 2000) Circe Effect (Jencks, 1975) Conformational Substates (Frauenfelder, 1987; Niehaus et al., 1997)
Model	**The Bohr Model** (1913)	**The Bhopalator**[1] (1983) (Ji, 1985, 1991, 2002b)
Concepts	**Quantum** (1900)	**The Conformon**[2] (1972) (Ji, 1974, 2000) **IDSs**[3] (1985) (Ji, 1991) **Cell Language**[4] (1997) (Ji, 1997, 2001, 2002a)
Theories	**Quantum Mechanics** (1925)	**Conformon Theory of Molecular Machines**[5] [(Ji, 1974, 2000) **Cell Language Theory**[6] (Ji, 1997, 2001, 2002a) **Molecular Information Theory**[7] (Ji, 2002c)
Mathematics	Probability Theory	Molecular Shape Algebra[8] Ciobanu and Ji, 2002)

[1]A molecular model of the living cell proposed at a meeting held in Bhopal, India, in 1983.

[2]Sequence-specific conformational strains of biopolymers that carry free energy and genetic information. *Conformons* are postulated to drive all molecular machines (by causing their shape changes), including simple enzymes, ion pumps, myosins, dyneins, kinesins, and topoisomerases.

[3]The acronym for ***i****ntracellular* ***d****issipative* ***s****tructure****s,*** such as transmembrane ion gradients, cytoplasmic calcium ion waves, and cytoskeletal mechanical strains that require dissipation of free energy for their maintenance. IDSs are thought to integrate the activities of individual molecular machines into the *supramolecular machines* called the cell.

[4]The molecular language in which living cells communicate among one another and process information or compute within themselves.

[5]This theory states that all molecular machines are driven by *conformons,* conformational strains of biopolymers that provide both the free energy to do work and the genetic information to control work.

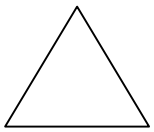
[6]According to the cell language theory, *genes* in cell language are analogous to words in human language and *space- and time-dependent teleonomic conformational changes of biopolymers* are analogous to uttered *sentences* in human language.

[7]Molecular information is defined as the ability of molecular systems to reduce molecular uncertainty, given the requisite free energy. Molecular uncertainty is in turn defined as the negative of the logarithm of molecular probability. MIT provides plausible mechanisms for the production, storage, transduction, and transmission of molecular information.

[8]A mathematical theory of molecular interactions based on space- and time-dependent shape changes of biopolymers as basic mechanisms of living processes.

The essence of the analogy between the atomic physics of the 19th-20th century and cell biology of the 20 th-21st century is summarized in the following two diagrams:

Atomic Spectroscopy



Regularities (Lyman series, Balmer series, etc.)

Bohr's Atomic Model

Figure 1. The role of atomic spectroscopy (absorption and emission) in the development of Bohr's model of the hydrogen atom.

Physicists had accumulated a large amount of experimental data on atomic absorption and emission spectra throughout the 19th century and, by the end of that century, discovered unexpected regularities embedded in these spectra. These regularities could be expressed in terms of simple algebraic formulas known as Lyman series, Balmer series, Ritz-Paschen series, etc., all of which had the following surprisingly simple mathematical form (Moore, 1963):

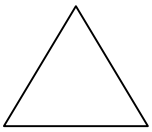
$$< = K\,(1/m^2 - 1/n^2) \quad \ldots\ldots\ldots\ldots (14)$$

where < is the frequency of a spectral line, K is a constant, m and n are two positive integers, whose numerical values varied with different series. As is well known, the simple regularity embodied in this formula remained mysterious until 1913, when Niels Bohr proposed a structural model of the hydrogen atom, based on which he was able to derive Eq. (14) from first principles, determine the numerical value of K using experimental data, and characterize the meanings of m and n (i.e., the lower and upper energy levels, respectively, of the hydrogen atom between which the electron undergoes transitions upon absorbing or emitting light quanta).

Of course, Figure 1 omits the important fact that, in order for Bohr to be able to come up with his atomic model, he needed more than just spectral data of atoms: He needed the concept of the quantum invoked by M. Planck in 1900 to account for blackbody radiation and the empirical knowledge that the hydrogen atom consisted of one proton and one electron. In other words, Bohr needed both the "holistic data" of spectral frequencies and the "reductionistic data" on atomic components and their physical properties.

I think it is reasonable to assert that there exists a historical analogy between atomic physics depicted in Figure 1 and the cell biology as represented in Figure 2:

Microarray Experiments



Regularities (GeneClusters, Characteristic Modes, Genetic Networks, etc.)

The Cell Model (e.g., the Bhopalator)

Figure 2. The predicted role of the microarray techniques in the formulation of a quantitative (computer) model of the living cell.

Just as the atomic spectroscopy provided a holistic experimental method for atomic physicists of the last century (Figure 1), so the microarray technique will be providing for biologists a holistic molecular biological method to study the living cell (Figure 2). The microarray methods (with which biologists can now study thousands of genes simultaneously) contrast the traditional biochemical and biophysical methods, most of which are "reductionistic" in the sense that they allow biologists to study individual genes and their products rather than the whole genome. Again, just as the holistic spectral data (quantitatively encoded in various named series) *plus* the reductionistic data on atomic components allowed Bohr to construct his theoretical model of the atom, so it may be predicted that biologists will need (for the construction of a viable computer model of the living cell) both the holistic data derived from microarray experiments and the traditional reductionistic data already available in the literature. To the best of my knowledge, the Bhopalator model of the cell proposed in (Ji, 1985) appears to be the only comprehensive theoretical model of the cell currently available in the biological literature. The validity of the Bhopalator may be checked against microarray data, but, before this can be done, it would be necessary for biologists first to extract "regularities" hidden in the enormous amount of the digital data generated by microarray experiments. Searching and documenting such regularities may take years and even decades of intensive and sophisticated statistical analyses, just as it took so many years for physicists to discover regularities among atomic line spectral data (Table 7). Microarray experiments are currently riddled with theoretical problems. For example, increases in mRNA levels measured with microarray techniques may not always be associated with corresponding increases in the rate of gene expressions, as is now widely assumed (Ji and Ohman-Strickland, 2002). Problems such as this highlight the urgent need for the biological community to become theoretically sophisticated in order to adequately deal with the complexity inherent in the system under investigation, namely the living cell. It took a group of brilliant physicists decades, if not centuries, to formulate a viable mathematical model of the hydrogen atom, a material system that is simpler than the living cell by a factor of at least 10^{15} (based on the volume ratio between the hydrogen atom and the average living cell). So, it may take decades, if not centuries, for biologists to come up with a mathematical/computer model of the living cell, it at all.

The regularities predicted to be found in microarray data will either (1) invalidate the Bhopalator and lead to an entirely different model of the cell, or (2) modify the Bhopalator in such a way as to accommodate the regularities (and even predict some of them).

Another facet that Figure 1 does not indicate (but Table 7 does) is the fact that the Bohr atom as proposed initially was incomplete, violating the electromagnetic wave equation of Maxwell, and the attempt to remedy this defect led to the birth of modern quantum mechanics through the efforts of *Heisenberg, de Broglie*, *Schroedinger*, *Pauli, Dirac,* and *Born* during the decade that followed the announcement of Bohr's atomic model in 1913. If Figure 2 is right, the Bhopalator model of the cell (if found to be compatible with microarray data) may similarly contribute to the emergence of fundamental theories in biology that will

account for life on the cellular level and beyond.

10. Conclusions

The set of physical *laws* and evolutionary *rules* that enable cells to communicate and compute has been referred to as *cell language* (Ji, 1977, 2001, 2002a). Just as *categorization* is considered to be the heart of human language, *complementary binding* between a biopolymer and its ligand has been postulated to be the heart of cell language and hence of cell communication and cell computing. Space- and time-dependent organization of complementary binding interactions driven by the Gibbs free energy decrease accompanying chemical reactions leads to biological functions. The coupling between biopolymer shape changes and chemical reaction intermediates has been visualized in terms of the coupling between two gears, one representing the biopolymer shape changes and the other the coupled chemical reaction. These considerations led to the suggestion that biological functions are a 5-tuple, the first three terms characterizing the chemical reactions that drives an orderly series of biopolymer shape changes, which are represented by the fourth and the fith terms. The process algebra of Milner et al. (1999) has been adopted to formulate what has been here referred to as 'shape algebra'. This new algebra seems to provide an efficient formal language for describing complex and dynamic biopolymer shape changes driven by exergonic chemical reactions that underlie teleonomic functions of the cell, including cell communication and cell computing.

Acknowledgements

One of the present authors (SJ) gratefully acknowledges the financial supports received from Dr. John L. Colaizzi, Dean, and Dr. R. Snyder, Chairman of the Department of Pharmacology and Toxicology, at The Ernest Mario School of Pharmacy, Rutgers University, Piscataway, N.J., that enabled him to attend the Workshop on Membrane Computing held in Curtea de Arges, Romania, in August, 2001. He is also grateful for the generous financial support provided by the Mathematical Institute of the National University of Singapore, which allowed him to travel to Singapore to present two papers at the Workshop on Modeling in Molecular Biology organized by Dr. G. Ciobanu in June, 2002.

References

Akama, S., Ed., 1997. Logic, Language and Computation. Kluwer Academic Publishers, Dordrecht.

Andrews, F. C., 1963. Equilibrium Statistical Mechanics. John Wiley & sons, Inc., New York, p. 31.

Babloyantz, A., 1986. Molecules, Dynamics & Life: An Introduction to Self-Organization of Matter. John Wiley & Sons, New York.

Bar-Ziv, R., Tlusty, T., Libchaber, A., 2002. Protein-DNA computation by stochastic assembly cascade. Proc. Nat. Acad. Sci. 99 (18), 11589-11592.

Benham, C. J., 1996. Duplex destabilization in superhelical DNA is predicted to occur preferentially at regulatory regions. J. Mol. Biol., 255, 425-434.

Chen, C. S., Mrksich, M., Huang, S., Whitesides, G. M., Ingber, D. E., 1997. Geometric Control of Cell Life and Death. Science 276, 1425-1428.

Ciobanu, G., 2000. Formal Description of the Molecular Processes. In C.Martin-Vide, Gh.Paun (Eds.): Recent Topics in Mathematical and Computational Linguistics, Editura Academiei, Bucharest 82-96.

Ciobanu, G., 2001. Molecular Structures. In: C. Martin-Vide, V. Mitrana (Eds.), Words, Sequences, Languages: Where Computer Science, Biology and Linguistics Come Across. Kluwer, Dordrecht, 299-317.

Ciobanu, G., Ciubotariu, V., Tanasa, B., 2002. A pi-calculus model of the Na pump, Genome Informatics 2002 (to appear).

Ciobanu, G., Ji, S., 2002. Shape Algebra: A Mathematical Representation of the Conformon Theory of Molecular Machines. In: G. Ciobanu (Ed.), Modeling in Molecular Biology (in preparation).

Ciobanu, G., Rotaru,M., 2002. Molecular Interaction. J. Theoret. Computer Science, Elsevier (to appear in Oct., 2002).

Clarke, E.M., Grumberg, O., Peled, D., 1999. Model Checking. MIT Press, Cambridge.

Cohen, M. C., Cohen, S., 1996. Cytokine Function: A Study in Biological Diversity. Am. J. Clin. Pathol., 105(5), 589-98.

Culler, J., 1991. Ferdinand de Saussure, Revised Edition, Cornell University Press, Ithaca.

Dam, M., 996. Model Checking Mobile Processes. Information and Computation 129(1), 35-51.

De Queiroz, J. G. B., Gabbay, D. M., 1999.Labelled Natural Deduction. In: Ohlbach, H. J., Reyle, U. (Eds.), Logic, Language and Reasoning, Kluwer Academic Publishers, Dordrecht, pp. 173-250.

Dillon, M. C., 1997. Merleau-Ponty's Ontology. Second Edition. Northwestern University Press, Evanston.

Eisen, M. B., Spellman, P. T., Brown, P. O., Botstein, D., 998. Cluster analysis and display of genome-wide expression patterns. Proc. Nat. Acad. Sci. (USA) 95, 14863-14868.

Ellis, J. M., 1993. Language, Thought, and Logic. Northwestern University Press, Evanston, pp. 27-44.

Frauenfelder, H., 1987. Function and dynamics of myoglobin. Ann. N. Y. Acad. Sci. 504, 152-167.

Hartwell, L. H., Hopfield, J. J., Leibler, S., Murray, A. W., 1999. From molecular to modular cell biology. Nature 402 (6761 Suppl): C47-52.

Havas, K. G., 1992. Thought, Language and Reality in Logic. Akademiai Kiado, Budapest.

Holter, N. S., Mitra, M., Maritan, A., Cieplak, M., Banavar, J. R., Fedoroff, N. V., Fundamental patterns underlying gene expression profiles: Simplicity from complexity. Proc. Nat. Acad. Sci. 97 (15), 8409-8414.

Honderich, T. (Ed.), 1995. The Oxford Companion to Philosophy. Oxford University Press, Oxford, pp. 296-297.

Jencks, W., 1975. Binding energy, specificity, and enzymic catalysis: The Circe Effect. Adv. Enzymol. 43, 219-410.

Ji, S., 1974. A General Theory of ATP Synthesis and Utilization. Ann. N. Y. Acad. Sci., 227, 211-226.

Ji, S., 1985. The Bhopalator: A Molecular Model of the Living Cell based on the Concepts of Conformons and Dissipative Structures. J. theoret. Biol. 116,

399- 426.
Ji, S., 1991. Biocybernetics: A Machine Theory of Biology. In: Ji, S. (Ed.), Molecular Theories of Cell Life and Death, Rutgers University Press, New Brunswick, pp. 1-237
Ji, S., 1995. Complementarism: A Biology-Based Philosophical Framework to Integrate Western Science and Eastern Tao. In: Psychotherapy East and West: Integration of Psychotherapies. Korean Academy of Psychotherapists, Seoul, Korea, 1995, pp. 518-548.
Ji, S., 1997. Isomorphism between cell and human languages. BioSystems, 44, 17-39.
Ji, S., 1999a. The Linguistics of DNA: Words, Sentences, Grammar, Phonetics, and Semantics. Ann. N. Y. Acad. Sci. 870, 411-417.
Ji, S., 1999b. The cell as the smallest DNA-based molecular computer. BioSystems, 52, 123-133.
Ji, S., 2000. Free energy and information contents of conformons in proteins and DNA. BioSystems, 554, 107-130.
Ji, S., 2001. Isomorphism between Cell and Human Languages: Micro- and Macrosemiotics. In: S. Simpkins and J. Deely (Eds), Semiotics 2000: "Sebeok's Century". Legas, Ottawa, 357-374.
Ji, S., 2002a. Microsemiotics of DNA. Semiotica, 138 (1/4), 15-42.
Ji, S., 2002b. The Bhopalator: An Information/Energy Dual Model of the Living Cell (II). Fundamenta Informaticae 49, 147-165.
Ji, S., 2002c. Molecular Information Theory. In: Ciobanu, G. (Ed.), Modeling in Molecular Biology (in preparation).
Ji, S., Ohman-Strickland, P., 2002. Iincreases in mRNA levels measured with microarray techniques do not necessarily mean increases in corresponding gene expressions (in preparation).
Lee, K., Ouyang, M., Georgopoulos, P., Ji, S., 2002. Imputations, *core clusters*, and *superclusters*: A novel method for clustering DNA microarray data. Bioinformatics (to be submitted).
Lyons, J., 1992. Language and Linguistics. Cambridge University Press, Cambridge, 228-235.
Lyons, J, 1993. Semantics I. Cambridge University Press, Cambridge, pp. 238-242.
Martinet, A., 1960, Elements of Genera Linguistics. The University of Chicago Press, Chicago.
McClare, C. W. F., 1971. Chemical Machines, Maxwell's Demon and Living Organisms. J. Theoret. Biol., 30, 1-34.
Marshall, C. J., 1994, Curr. Opin. Genet. Develop. 4, 82-89.
Mayer, B., Baltimore, D., 1993. Signaling through SH2 and SH3 domains. Trends Cell Biol., 3**,** 8-13.
Milner, R., 1999. Communicating and mobile systems: the B-calculus Cambridge University Press, Cambridge.
Moore, W. J., 1963. Physical Chemistry. Third Edition. Prentice-Hall, Inc., Englewood Cliffs, N.J., pp. 474-477.
Nathan, C., Sporn, M., 1991. Cytokines in Context. J. Cell Biol. 113(5), 981-985.
Nienhaus, G. U., Mηller, J. D., McMahon, B. H., and Frauenfelder, H., 1997. Exploring the conformational energy landscape of proteins. Physica D 107, 297-311.
Norris, V., et al., 1999. Hypothesis: Hyperstructures regulate bacterial Structure and the cell cycle. Biochimie 81, 915-920.
Ohlbach, H. J, Reyle, U. (Eds.), 1999. Logic, Language and Reasoning, Kluwer Academic Publishers,

Dordrecht.
Pais, A., 1991. Niels Bohr's Times, in Physics, Philosophy, and Polity. Clarendon Press, Oxford.
Rosen, R., 1991. Life Itself: A comprehensive Inquiry Into the Nature, Origin, and Fabrication of Life. Columbia University Press, New York.
Smith, H. A., Welch, G. R., 1991. Cytosociology: A Field-Theoretic View of Cell Metabolism. In: Ji, S. (Ed.) Molecular theories of Cell Life and Death, Rutgers University Press, New Brunswick, pp. 282-323.
Srere, P. A., 1987. Complexes of Sequential Metabolic Enzymes. Ann. Rev. Biochem. 56, 89-124.
Stryer, L., 1995. Biochemistry. Fourth Edition. W. H. Freeman and Company, New York, p. 218.
Victor, B., Moeller, F., 1994. The Mobility Workbench - A tool for the pi-Calculus. In: D. Dill (Ed.), Proceedings CAV'94: Computer-Aided Verification. Lecture Notes in Computer Science, Vol.818, Springer Verlag, Berlin, 428-440.