Available online at www.sciencedirect.com
SCIENCE DIRECT®
BioSystems 70 (2003) 165–181
BioSystems
www.elsevier.com/locate/biosystems

Conformon-driven biopolymer shape changes in cell modeling

Sungchul Ji [a,*], Gabriel Ciobanu [b]

[a] *Department of Pharmacology and Toxicology, Rutgers University, Piscataway, NJ 08855, USA*
[b] *School of Computer Science, National University of Singapore, Singapore, Singapore*

Received 27 December 2002

This paper is dedicated to the memory of S.J.'s mother, Ms. Bok Nyo Keh, who passed away in Atlantic City, NJ on 2 September 2002 at the age of 86, as this manuscript was being born.

Abstract

Conceptual models of the atom preceded the mathematical model of the hydrogen atom in physics in the second decade of the 20th century. The computer modeling of the living cell in the 21st century may follow a similar course of development. A conceptual model of the cell called the *Bhopalator* was formulated in the mid-1980s, along with its twin theories known as the *conformon theory* of molecular machines and the *cell language theory* of biopolymer interactions [Ann. N.Y. Acad. Sci. 227 (1974) 211; BioSystems 44 (1997) 17; Ann. N.Y. Acad. Sci. 870 (1999a) 411; BioSystems 54 (2000) 107; Semiotica 138 (1–4) (2002a) 15; Fundamenta Informaticae 49 (2002b) 147]. The conformon theory accounts for the reversible actions of individual biopolymers coupled to irreversible chemical reactions, while the cell language theory provides a theoretical framework for understanding the complex networks of dynamic interactions among biopolymers in the cell. These two theories are reviewed and further elaborated for the benefit of both computational biologists and computer scientists who are interested in modeling the living cell and its functions. One of the critical components of the mechanisms of cell communication and cell computing has been postulated to be space- and time-organized teleonomic (i.e. goal-directed) *shape changes* of biopolymers that are driven by exergonic (free energy-releasing) chemical reactions. The generalized Franck-Condon principle is suggested to be essential in resolving the apparent paradox arising when one attempts to couple endergonic (free energy-requiring) biopolymer shape changes to the exergonic chemical reactions that are catalyzed by biopolymer shape changes themselves. Conformons, defined as sequence-specific mechanical strains of biopolymers first invoked three decades ago to account for energy coupling in mitochondria, have been identified as *shape changers*, the agents that cause *shape changes* in biopolymers. Given a set of space- and time-organized teleonomic shape changes of biopolymers driven by 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, it is necessary to represent the conceptual model in an algebraic language. To this end, we have begun to apply the process algebra of Milner [Communicating and Mobile Systems: The π-calculus, Cambridge University Press, Cambridge, 1999] to develop what is here called the "shape algebra," capable of describing complex and mobile patterns of interactions among biomolecules leading to cell functions.
© 2003 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Cell modeling; Cell computing; Cell language; Molecular logic; Time- and space-dependent shape change(r)s of biopolymers; Conformons; Generalized Franck-Condon principle; Shape algebra

1. Introduction

The advent of the microarray technique in molecular biology in the mid-1990s has been revolutionizing

* Corresponding author.
E-mail addresses: sji@eohsi.rutgers.edu (S. Ji), gabriel@comp.nus.edu.sg (G. Ciobanu).

0303-2647/$ – see front matter © 2003 Elsevier Science Ireland Ltd. All rights reserved.
doi:10.1016/S0303-2647(03)00038-8

molecular and cell biology (Eisen et al., 1998; Holter et al., 2000; Wyrick and Young, 2002; Cooper, 2002; Nakanish et al., 2001). With this method, biologists can now measure the intracellular concentrations of tens of thousands of mRNA molecules quickly and simultaneously. However, the mechanisms underlying the regulation of mRNA levels inside the cell are so complex that it is almost impossible to analyze microarray data meaningfully without any viable theoretical model of the living cell itself. Therefore, one of the major benefits of modeling the cell (communication and computation) is to enable biologists to interpret microarray data in a biologically meaningful way, just as the model of the hydrogen atom proposed by Bohr in 1913 allowed physicists to interpret absorption and emission spectral data of atoms in a quantitative manner (Moore, 1963; Corney, 1977).

Modeling the living cell 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 the associated molecular theories have been proposed during the past three decades, but no quantitative computational model of the cell exists at present. To construct such a model, 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 objectives of this contribution is (a) to describe the key concepts and theories on which the *conceptual model* of the living cell, the *Bhopalator* (Ji, 1985, 1991, 2002b), is based; (b) to elaborate on 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 (c) to present our preliminary version of the "shape algebra" of biopolymers that is aimed at capturing the essence of molecular transformations and interactions that drive cell *communication*, *computation* and *construction*.

2. Modeling relations

To successfully model a system as complex as the living cell, it may be useful to have a theory of modeling itself. According to Rosen (1991), a *modeling relation* exists between a natural system, N, and a formal system, F, if the following five 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 states that:

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, 2001, 2002a; see Section 3) can be viewed as a specific example of the 'modeling relation' described above. The similarity between human and cell languages connects the *mental world* (F in Table 1)

Table 1
The modeling relation between cell and human languages

Natural (N)	Formal (F)
System	
Molecular language[a]	Symbolic language
Molecular computing[b]	Symbolic computing
Entailment	
Causality (or physical laws)	Inference (or logical rules)
Evolutionary rules	Rationality
(molecular logic)	(human logic)

[a] Synonymous with 'cell language,' since cells use molecules as signs in communication.

[b] Synonymous with 'cell computing,' since cells use molecules as information carriers in signal transduction and computation. For the definition of 'cell computing,' see Sections 3 and 5.

Table 2
A comparison between human and cell languages

	Human language	Cell language	
		DNA language	Protein language
Alphabet (L)	Letters	4 Nucleotides	20 Amino acids
Lexicon (W)	Words	Genes	Polypeptides
Sentence (S)	Strings of words	Sets of genes expressed coordinately in space and time directed by DNA folds	Sets of proteins interacting noncovalently
Grammar (G)	Rules of sentence formation	Rules mapping DNA sequences to folding patterns of DNA under biological conditions	Rules mapping polypeptide sequences to 3-D foldings of polypeptides
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 and intracellular dissipative structures (IDSs)	
Semantics (S)	Meaning of words and 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 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 into genes through covalent interactions	Organization amino acids into polypeptides through covalent interactions

and the *cellular world* (N in Table 1). As the 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 3). The same isomorphism has been used as a decoding dictionary from human language to cell language, which led 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 primarily in the noncoding regions of DNA that control the space- and time-dependent folding patterns of DNA, leading to the control of the timing of gene expression. An indirect empirical evidence supporting a functional role of noncoding regions of DNA was provided by Amano et al. (1997) who showed a linear correlation between the number of noncoding bases and the number of transcription factors in a given genome, specifically among multicellular organisms but not unicellular ones (Ji, 2002a).

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: (a) communication with one another by exchanging appropriate messenger molecules such as neurotransmitters, hormones, cytokines, growth factors, etc., and (b) execution of the genetic programs stored in DNA that are called for, or triggered by, the information carried by intercellular messenger molecules. The former constitutes *cell communication* and the latter *cell computation*. "*Cell computing*" is defined here as the input-state transition-output processes performed by the cell in order to accomplish specific biological functions or goals. This definition is consistent with the definition of *molecular computing* given by Conrad (1992). The system of molecules and their rules of interactions that are necessary for cell communication and cell computation is referred to as *cell language* (defined below) and the rules embedded in it as *cell logic or molecular logic* (see Section 6), 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 reasonable to assume that there exists a language that cells use to communicate among themselves. Such a language was referred to as *cell language* and its postulated characteristics were described elsewhere (Ji, 1997, 1999a, 2001, 2002a). Humans have developed specialized systems of symbolic signs for mathematics and formal languages for computation. Similarly cells appear to use a system of molecular signs (or *indexical signs* in semiotic terminology of C.S. Peirce (1839–1914)), (Buchler, 1955) and associated rules for the purpose of not only *molecular communication* but also *molecular computation* (i.e. the processing of information according to genetic programs stored in DNA and input signals received from cell environment) (Conrad, 1982) and *molecular construction* (i.e. the production of molecules or molecular processes). In other words, it is postulated here that living cells use cell language to perform the triad of *communication*, *computation*, and *construction* on the molecular level:

Communication	Exchange of 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 stored in DNA.
Construction	Building new molecular and supramolecular structures or processes, following the results of cell computation.

For convenience, we will call the above three processes *the c triad*. Available experimental data indicate that, under normal conditions, these three cellular processes are tightly coupled so that they occur together but, under pathological conditions (e.g. presence of inhibitors or uncouplers), only a part of the component processes of the c triad may occur leading to unphysiological consequences such as cancer and malformations of organs. We will refer to this postulate 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 *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 conceptual 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.

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 sequences (or strings) of gene expressions arranged or emitted in *time*.

Of all the possible folding patterns of DNA and proteins allowed for by the laws of physics and chemistry, only small subsets have been selected by evolution based on their functional characteristics, thus constituting the grammar of cell language.

Conformons are defined as sequence-specific conformational strains that carry both free energy (needed for doing work) and genetic information (needed for controlling work). Conformons thus defined are postulated to provide the ultimate driving forces for all molecular processes inside the cell (Ji, 1974, 2000).

Intracellular dissipative structures (IDSs) are gradients of ions and molecules or mechanical stress gradients sustained by cytoskeletal systems, all confined within the cell volume. These are the microscopic cellular versions of the concept of "dissipative structures" first defined on the macroscopic scale by Prigogine and his school (Babloyantz, 1986). Experimental evidence indicates that it is IDSs, not polypeptides, that serve as the proximal or immediate physical causes for all cell functions (Ji, 1991).

One of the major postulates of the cell language theory is that the dichotomy of *covalent* and *noncovalent* interactions ubiquitously found in the cell corresponds to the *double articulation* in human language. Noncovalent interactions that do not involve breaking or forming covalent bonds are also called '*conformational interactions*,' and these are thought

to be analogous to the first articulation in human language. Conformations are defined as "non-identical arrangements of the atoms in a molecule obtainable by rotation about one or more single bonds" (Eliel et al., 1966). Because conformational interactions involve only the rotation around or bending of covalent bonds, they implicate smaller energy changes (typically less than a few Kcal/mol) than covalent interactions, which involve much larger energy changes (10–30 Kcal/mol) due to breaking and/or forming chemical bonds (Ji, 1997). Covalent interactions involve changes in valence electronic configurations around nuclei of atoms in molecules and are thought to correspond to the second articulation in human language.

Table 2 (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, more generally, *semiotic* principles (Ji, 1997, 2001, 2002a).

4. Cell computing

Cell computing, defined as state transitions of the cell controlled by both the genetic information encoded in DNA and the environmental inputs received via receptors, is synonymous with what biologists call *signal transduction*, one of the most active research fields in contemporary molecular and cell biology. Mayer and Baltimore (1993) recently defined signal transduction as "*the conversion of an input signal received at the extracellular face of 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 used in the field (Orphanides and Reinberg, 2002). Thus, we can represent signal transduction schematically as follows:

$$\text{Input Signal} \xrightarrow{a} \text{Intracellular Signal} \xrightarrow{b} \text{Gene Expression} \xrightarrow{c} \text{Output Signal} \quad (1)$$

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 formal structure as "labeled deductive systems (LDS)" in computer science, a unifying framework for the study of logics and their interactions proposed by Gabbay in the late 1990s (Ohlbach and Reyle, 1999). LDS can be represented as

$$A \xrightarrow{f} B \quad (2)$$

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, $A \rightarrow B$, 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 systems" (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 would be reasonable to conclude that LDS and MLDS are also isomorphic. Let us refer to this conjecture as "the postulate of the isomorphism between LDS and MLDS."

5. Molecular sentences and cell computing

The difference between *words* (or more precisely "non-logical concepts" of Mahner and Bunge (1997, p. 52)) and *sentences* (or "propositions," Mahner and Bunge, ibid.) is that the former represents ideas or objects, while the latter represents *judgments* (Martinet, 1967), which may be regarded as synonymous with *computations*. This view seems to be in agreement with the approach of G. Frege (1848–1925) who applied the mathematical 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 (i.e. order of letters in a word) 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 (within grammatical constraints, of course) 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, which is hypothesized to be analogous to or isomorphic with the dichotomy of the *first* and *second* articulations in linguistics (see *double articulation* in Table 2). That is, the cell-linguistic analogs of words are *covalently bonded structures*, while those of sentences are 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, while 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 Orphanides and Reinberg, 2002) can be treated as a system of one or more cell-linguistic or molecular sentences, or one or more processes of molecular computations.

Not all 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 delimit the variety of well-formed words and 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:

$$\text{Subject} \rightarrow \text{Dative verb} \rightarrow \text{Indirect object} \rightarrow \text{Direct object} \quad (3)$$

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

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.

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 in Table 3.

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):

$$\text{Receptor} \rightarrow \text{MAPKKK} \rightarrow \text{MAPKK} \rightarrow \text{MAPK} \rightarrow \text{TF} \rightarrow \text{DNA} \quad (4)$$

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 to which TF binds. The obvious similarity between Schemes (3) and (4) may be considered as providing an indirect support 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 are summarized in Table 4.

The symbols in italics represent proteins that are analogous to words forming molecular sentences. The question mark in at the bottom of the first column

Table 4
Paradigmatic relations among signal transducing proteins in the MAP kinase cascade

	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*

Data from Marshall (1994).

indicates that the predicted protein was not yet found as of the time of the writing of this manuscript.

It is noteworthy that both cell types and cell functions (see the first column) determine the nature of paradigmatic substitutes (see the elements of Table 4, representing 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 first 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 first two rows, Table 1). Since human logic underlies both human language and symbolic computation (Akama, 1997; Havas, 1992), it would seem reasonable to conclude that there exists cell (or molecular) logic that supports both cell language and cell computing (see the second row, Table 1). In the next section, the concept of molecular/cell logic is discussed.

6. Molecular (or cell) logic

It is convenient to describe *molecular logic* utilizing Aristotle's doctrine of four causes, although they are not absolutely independent of one another.

1. *Material cause*. 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 space–time affect each other in general relativity (Smith and Welch, 1991; also 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 can express such a coupling between chemical reactions and biopolymer shape changes using process algebra (see Section 8). 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. In other words, the shape of a biopolymer is determined both by environmental information due to its conformational flexibility and by genetic information that affects its covalent structure (Ji, 2002b). 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 and DNA) that both catalyze, and are driven by, exergonic chemical reactions.
2. *Formal cause*. The living cell has been 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. 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 five-dimensional, 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 what is here called biological or genetic *information*, a parameter unique to living systems.

3. *Efficient cause*. The fundamental mechanism by which molecular logic is implemented is postulated here to be *nonstochastic* (alternatively 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 thermodynamics, and quantum and statistical mechanics (Ji, 1991, 2000, 2002b).

   But there are subtle difficulties to overcome when one considers the molecular details of the mechanism of coupling endergonic biopolymer shape changes to exergonic chemical reactions for which such shape changes are required for catalysis. Since the free energy that drives endergonic biopolymer shape changes cannot be released from exergonic chemical reactions until biopolymers actually catalyze those reactions, where do biopolymers obtain free energy to drive their initial endergonic 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, in order not to violate the second law of thermodynamics (McClare, 1971; Ji, 2000). As long as enzymes do not store thermal energies longer than, or equal to ϑ, the second law is not violated by such transient borrowing by biopolymers of thermal energies (McClare, 1971; Ji, 1974, 2000).

   Conrad (1982, 1992, 1999) and Conrad and Zauner (1998) also considered biopolymer shape changes as fundamental to molecular computing. Conrad formulated a comprehensive model of cell computing, called the *self-assembly model*, based on pattern recognizing power of biopolymer shapes. But he did not propose any molecular mechanisms for coupling biopolymer shape changes to biochemical reactions that they catalyze, as was done in Ji (1974) and reviewed above. Since all biopolymer shape changes are reversible while cell and molecular computing is not (e.g. hormone-induced synthesis of a protein cannot be reversed), biopolymer shape changes alone, without being coupled to appropriate irreversible exergonic chemical reactions, cannot be said to carry out any molecular computation and hence the expression "conformation-driven computation" frequently employed in Conrad (1999) may be judged as inaccurate and hence best replaced by the "thermodynamically correct" expression, "conformon-driven computation."

   It is interesting to note that logical operations (e.g. $+$, $-$, x, /, etc.) performed by base elements in human-made computers are unidirectional or irreversible, while the operations of base elements of cell computers, namely biopolymer shape changes, are bi-directional or reversible, as already mentioned. This difference may constitute a fundamental barrier for constructing a common theory of computing (Kazic, 1999) that can be applied to both human-made and molecular computers. It was postulated elsewhere (Ji, 1999b) that the cell is the smallest molecular computer in existence. The validity of this postulate appears enhanced by the fact that the ultimate arbiter of the direction of all the chemical reactions going on inside the cell is the living cell itself (i.e., an organized system of biopolymers and chemical reactions), not just biopolymers, neither individually nor in groups.
4. *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

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	Gödel's theorem	Thermal fluctuations Heisenberg uncertainty principle
9. Scale	Macroscopic	Microscopic
10. Complementarity	Mind/body	Information/energy

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 and multicellular organisms.

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 (Buchler, 1955; Corrington, 1993), the father of modern *semiotics* (the study of signs) and a pioneer in modern logic, distinguished three categories of signs (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 have dual characters—symbolic signs (as carriers of *information*) and indexical signs (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 functionally significant *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*: Gödel'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). In addition, molecular logic may be subject to imprecision, thus qualifying molecular words and sentences as fuzzy sets (Zadeh, 1965).

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 hypothetical physical entity exhibiting *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* (see Section 3) (Ji, 1974, 1991, 2000). Available data indicate that the average conformon in proteins carries 8–16 Kcal/mol of free energy and 40–200 bits of information (Ji, 2000; calculations based on data from Benham, 1996). The corresponding values for the average conformon in DNA have been estimated to be 500–2500 Kcal/mol 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 teleonomic functions (Ji, 2000).

All molecular machines inside the cell, from simple enzymes to complex ones such as ion pumps and molecular motors are thought to be driven by conformons produced by exergonic chemical reactions catalyzed by biopolymers. Since biopolymer shape changes are intrinsic to any operational cycle of molecular machines, it seems logical to conclude that they too are driven by conformons. Thus, the sequence of events from exergonic chemical reactions to endergonic teleonomic functions of biopolymers may be schematically represented as follows:

$$\text{Chemical reactions} \rightarrow \text{Conformons} \rightarrow \text{Biopolymer shape changes} \rightarrow \text{Teleonomic functions} \quad (5)$$

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).

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

At 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). Such structural complementarity may be best characterized in terms of *possibility theory* based on multivalued (or fuzzy) logic rather than traditional *probability*

theory based on binary logic (Klir and Harmanec, 1996).

(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 any generality) as follows:

$$A + B \leftrightarrow A'/B' \tag{6}$$

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 imposed 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), according to which conformational changes of A and B to A′ and B′, respectively, follow, rather than precede, the binding interactions, unlike in the Franck-Condon mechanism shown below:

$$A \leftrightarrow A' \tag{7}$$

$$B \leftrightarrow B' \tag{8}$$

$$A' + B' \leftrightarrow A'/B' \tag{9}$$

Notice that the sum of Processes (7)–(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 as written.

(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 a set or an ensemble of n conformers, each having slightly different energies, e_i. The probability of occurrence of the ith conformer, $P_i(e_i)$, is then given by the Boltzmann distribution law,

$$P_i(e_i) = g_i\, \mathrm{e}^{-e_i/kT} \tag{10}$$

where g_i is the ith 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). Eq. (10) states that as the energy content of a conformer, e_i, increases, the probability of a biopolymer assuming that conformation decreases exponentially, as long as 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 thermal motions are not energetic enough to break

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 Frauenfelder (1987) and others (Nienhaus et al., 1997) 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 functionally 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 belonging or 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 ith to the $(i+1)$th conformers, B_i to B_{i+1}, that catalyzes a chemical reaction, $c_i \rightarrow c_{i+1}$, driven by the Gibbs free energy change, g_i. The ith 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):

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

$$c_i \overset{B_i \leftrightarrow B_{i+1}}{\underset{f_i \rightarrow g_i}{\longleftrightarrow}} c_{i+1} \tag{12}$$

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 bi-directional arrow in (12). As already indicated in Section 6, biopolymer shape changes, which are bi-directional (or reversible), can be made to proceed unidirectionally only when coupled to exergonic and irreversible chemical reactions that dissipate Gibbs free energy by the amount, g_i. 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 already mentioned in Section 6. The direction of Process (12) is entirely dependent on the sign of the Gibbs free energy change, 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 expressed as the sum of a set of elementary (or atomic) biological functions, f_i, that are coupled in space and time and hence context-dependent (Conrad, 1999; Ji, 2002):

$$F = \Sigma_{i,w} f_i \tag{13}$$

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 biopolymers 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 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). 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, 2001). In Ciobanu (2000), the pi-calculus is used to describe the DNA methylation, using conformational transformations. 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; the pi-calculus is a general computational model able to simulate the lambda-calculus (Milner, 1999), thereby achieving Turing completeness (all computable functions are definable in it). 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. (in press). The Albers-Post mechanism of the Na^+/K^+ ATPase is translated into a model which can describe molecular interactions, conformational transformations, and ion transport 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 $x, y, \ldots$, and *molecular processes* $P, Q, \ldots$ 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 $x\langle z\rangle$ 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 used in one interaction can cause a shape change that allow another interaction; in this way, a process can interact with processes that previously could not interact because of lack 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 $x, y, z \ldots$. The terms of this formalism are called processes and processes are denoted by $P, Q, R, \ldots$.

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

$$p ::= x\langle z\rangle | x(y) | [x = y]p.$$

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 $x\langle z\rangle$ represent the passive and the active complementary shapes, receiving and sending a shape during an interaction. The output prefix $x\langle z\rangle$ 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 $x\langle z\rangle.P|x(y).Q$ 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) $x\langle z\rangle.P$ and a passive part (female, receiver) $x(y).Q$, and it can be represented by the following transition:

$$x\langle z\rangle.P|x(y).Q \rightarrow P|Q\{z/y\}.$$

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 $\equiv$ over processes is called structural congruence and it is defined as the smallest congruence over processes which satisfies Tables 6 and 7.

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.

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 $\rightarrow$ satisfying the following rules:

$$\text{pre}: \ \alpha.P \xrightarrow{\alpha} P \quad \text{sum}: \ \frac{P \xrightarrow{\alpha} P'}{P + Q \xrightarrow{\alpha} P'}$$

$$\text{rep}: \ \frac{P|!P \xrightarrow{\alpha} P'}{!P \xrightarrow{\alpha} P'} \quad \text{par}: \ \frac{P \xrightarrow{\alpha} P'}{P|Q \xrightarrow{\alpha} P'|Q}$$

$$\text{com}: \ \frac{P \xrightarrow{\bar{\alpha}y} P' \ Q \xrightarrow{\alpha(x)} Q'}{P|Q \rightarrow P'|Q'(y/x)} \quad \text{match}: \ \frac{P \rightarrow^{\alpha} P'}{[\alpha = \alpha]P \xrightarrow{\alpha} P'}$$

$$\text{struct}: \ \frac{P \equiv P' \ P' \xrightarrow{\alpha} QQ' \equiv Q}{P' \xrightarrow{\alpha} Q'}$$

It is very useful to be able to compare the behavior of two systems. We introduce two behavioral equivalences. These behavioral equivalence are based on the

Table 7
The definition of structural congruence

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

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 $P\sigma \rightarrow P'$, there exists Q' so that $Q\sigma \rightarrow Q'$ 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 $P \sim Q$, 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 such as Mobility Workbench (Victor and Moeller, 1994), 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 et al., 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. (in press) describes the sodium–potassium channel. 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 coordination 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 conformations 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. Conclusions

The set of physical *laws* and evolutionary *rules* that enable cells to *communicate*, *compute*, and *construct* has been referred to as *cell language* (Ji, 1974, 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, computing, and construction. Space- and time-dependent organization of complementary binding interactions driven by Gibbs free energy decreases 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 Reactions. 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 fifth terms. The process algebra of Milner (1999) has been adopted to formulate what has been here referred to as 'shape algebra.' This new algebra, when fully developed, may provide an efficient formal language for describing complex and dynamic biopolymer shape changes driven by conformons that underlie teleonomic functions of the cell, including *cell communication*and *cell computing*.

Acknowledgements

One of the present authors (S.J.) 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, NJ, 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.

Amano, N., Ohfuku, Y., Suzuki, M., 1997. Genomes and DNA conformation. Biol. Chem. 378, 1397–1404.

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

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

Bar-Ziv, R., Tlusty, T., Libchaber, A., 2002. Protein–DNA computation by stochastic assembly cascade. Proc. Natl. Acad. Sci. U.S.A. 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.

Buchler, J., 1955. Philosophical Writings of Peirce. Dover Publications, Inc., New York, pp. 98–119.

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: Martin-Vide, C., Paun, Gh. (Eds.), Recent Topics in Mathematical and Computational Linguistics. Editura Academiei, Bucharest, pp. 82–96.

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

Ciobanu, G., Ciubotariu, V., Tanasa, B., in press. A pi-calculus model of the Na pump. Genome Informatics.

Ciobanu, G., Rotaru, M., 2002. Molecular interaction. J. Theoret. Comp. Sci. 289 (1), 801–827.

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

Conrad, M., 1982. Molecular computing as a link between biological and physical theory. J. Theoret. Biol. 98, 239–252.

Conrad, M., 1992. Molecular computing: the lock-key paradigm. Computer 25, 11–20.

Conrad, M., Zauner, K.-P., 1998. DNA as a vehicle for the self-assembly model of computing. BioSystems 45, 59–66.

Conrad, M., 1999. Molecular and evolutionary computation: the tug of war between context freedom and context sensitivity. BioSystems 52, 99–110.

Cooper, S., 2002. Cell cycle analysis and microarrays. Trends Genet. 18 (6), 289–290.

Corney, A., 1977. Atomic and Laser Spectroscopy. Clarendon Press, Oxford, pp. 1–16.

Corrington, R.S., 1993. An Introduction to C.S. Peirce. Rowman & Littlefield Publishers, Inc., Lanham, MD.

Culler, J., 1991. Ferdinand de Saussure (Rev. Ed.). Cornell University Press, Ithaca.

Dam, M., 1996. Model checking mobile processes. Inform. Comput. 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 (2nd ed.). Northwestern University Press, Evanston.

Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Nat. Acad. Sci. U.S.A. 95, 14863–14868.

Eliel, E.L., Allinger, N.L., Angyal, S.J., Morrison, G.A., 1966. Conformational Analysis. Interscience Publishers, New York.

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 (Suppl. 6761), C47–C52.

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., 2000. Fundamental patterns underlying gene expression profiles: simplicity from complexity. Proc. Natl. Acad. Sci. U.S.A. 97 (15), 8409–8414.

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

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, 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 54, 107–130.

Ji, S., 2001. Isomorphism between cell and human languages: micro- and macrosemiotics. In: Simpkins, S., Deely, J. (Eds.), Semiotics 2000: “Sebeok’s Century”. Legas, Ottawa, pp. 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.

Kazic, T., 1999. Elements of a more comprehensive theory of computing. BioSystems 52, 111–122.

Klir, G.J., Harmanec, D., 1996. Generalized information theory: recent developments. Kybernetes 25 (7/8), 50–67.

Mahner, M., Bunge, M., 1997. Foundations of Biophilosophy. Springer-Verlag, Berlin.

Martinet, A., 1967. Elements of General 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. MAP kinase kinase kinase, MAP kinase kinase and MAP kinase. Curr. Opin. Genet. Dev. 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 π-calculus. Cambridge University Press, Cambridge.

Moore, W.J., 1963. Physical Chemistry (3rd ed.). Prentice-Hall, Englewood Cliffs, NJ, pp. 474–477.

Nakanish, T., Oka, T., Akagi, T., 2001. Recent advances in DNA microarrays. Acta Medica Okayama 55 (6), 319–328.

Nienhaus, G.U., Mηller, J.D., McMahon, B.H., Frauenfelder, H., 1997. Exploring the conformational energy landscape of proteins. Physica D 107, 297–311.

Norris, V., Alexandre, S., Bouligand, Y., Cellier, D., Demarty, M., Grehan, G., Gouesbet, G., Guespin, J., Insinna, E., Le Sceller, L., Maheu, B., Monnier, C., Grant, N., Onoda, T., Orange, N., Oshima, A., Picton, L., Polaert, H., Ripoll, C., Thellier, M., Valleton, J.-M., Verdus, M.-C., Vincent, J.-C., White, G., Wiggins, P., 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.

Orphanides, G., Reinberg, D., 2002. A unified theory of gene expression. Cell 108, 439–451.

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 (4th ed.). W. H. Freeman and Company, New York, p. 218.

Victor, B., Moeller, F., 1994. The mobility workbench—a tool for the pi-calculus. In: Dill, D. (Ed.), Proceedings CAV’94: Computer-Aided Verification. Lecture Notes in Computer Science, vol. 818. Springer-Verlag, Berlin, pp. 428–440.

Wyrick, J.J., Young, R.A., 2002. Deciphering gene expression regulatory networks. Curr. Opin. Genet. Dev. 12 (2), 130–136.

Zadeh, L.A., 1965. Discussion: probability theory and fuzzy logic are complementary rather than competitive. Technometrics 37 (3), 271–276.