Abstract
The pioneering work of Adleman (1994) demonstrated that DNA molecules in test tubes can be manipulated to perform a certain type of mathematical computation. This has stimulated a theoretical interest in the possibility of constructing DNA-based molecular computers. To gauge the practicality of realizing such microscopic computers, it was thought necessary to learn as much as possible from the biology of the living cell—presently the only known DNA-based molecular computer in existence. Here the recently developed theoretical model of the living cell (the Bhopalator) and its associated theories (e.g. cell language), principles, laws and concepts (e.g. conformons, IDS’s) are briefly reviewed and summarized in the form of a set of five laws of 'molecular semiotics' (synonyms include 'microsemiotics', 'cellular semiotics', or 'cytosemiotics')—the study of signs mediating measurement, computation, and communication on the cellular and molecular levels. Hopefully, these laws will find practical applications in designing DNA-based computing systems. © 1999 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Molecular computer; Cell language; Cell model; Molecular semiotics; Cytosemiotics; Microsemiotics; Conformons

1. Introduction
In 1994, Adleman successfully manipulated a set of $10^{13} - 10^{14}$ molecules of 20-mer DNA sequences to solve an instance of the directed Hamiltonian path problem (Adleman, 1994). This suggests that it may be possible to fabricate one day a molecular computer based on a single copy of DNA duplex. The purpose of this paper is to summarize recent developments in the theoretical biology of the living cell in the hope that some of these principles may be found applicable to designing DNA-based computers, both molecular and macroscopic.

For logical consistency, it is necessary to distinguish two kinds of DNA-based computers: (1) truly molecular computers (TMC) wherein all the computational operations (i.e. input, output, and state transitions) are driven by self-organizing ('self-' meaning 'spontaneous' or 'internally driven') chemical reactions and molecular motors; and (2) quasi-molecular computers (QMC), where most, if not all, of the input and output operations and state transitions are driven 'externally' by macroscopic agents such as human beings or robots. The DNA-based computing system developed by Adleman is an example of QMC, and not TMC, since he had to synthesize 20-mer DNA sequences by himself, carry out splicing and amplification reactions on them, and separate and purify their products using conventional molecu-
lar biological operations on a macroscopic scale. At present, the living cell remains the only TMC known. Therefore, by studying living cells and trying to mimic their mechanisms in full or in part, it may be possible to produce an artificial TMC. Or, failing that, it may be possible at least to construct more efficient QMC’s than are presently available.

2. Cells as truly molecular computers (TMC)

It is instructive to compare the characteristics of the living cell with those of human-made computers as shown in Table 1 (Ji, 1991). The cell can be treated as a molecular machine, having inputs, state transitions, and outputs as indicated. The input device of the cell is provided by receptors which recognize their stereospecific molecular ligands, and the internal state of the cell is determined by gradients of metabolite concentrations, phosphoproteins, and mechanical stresses localized inside the cell, collectively called IDS’s (Intracellular Dissipative Structures, or the dissipative structures of Prigogine) (Ji, 1985, 1991). The language cells use to process information within, and communicate information among themselves was recently characterized and named ‘cell language’ (Ji, 1997a,b, 1999a), or ‘cellese’ (Ji, 1998a) (see below). The current (electron) carriers and conductors in the cell are chemicals which contrast with electron fluxes and wires in human-made computers. Enzymes in the cell turn on or off chemical reactions, just as transistors turn on or off electron flow in computer circuits. It is postulated that the switching functions of enzymes are driven by sequence-specific conformational strains of enzymes known as conformons (Ji, 1974, 1979, 1998b), in analogy to photons (quanta of the electromagnetic radiation) which control the function of transistors. Conformons are not only energy packets but also contain genetic information: Conformons carry both energy and information (more later). Therefore, it may be possible in principle to design biopolymer-based computational circuitry utilizing the mechnochemical and symbolic degrees of freedom inherent in conformons. A similar suggestion was recently made in a paper by

<table>
<thead>
<tr>
<th>Parameters</th>
<th>The cell</th>
<th>The computer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Input</td>
<td>DNA, substrates, receptor ligands</td>
<td>Input data</td>
</tr>
<tr>
<td>2. State transitions</td>
<td>Changes in IDS’s</td>
<td>Changes in internal states</td>
</tr>
<tr>
<td>3. Output</td>
<td>IDS-driven cell processes</td>
<td>Output data</td>
</tr>
<tr>
<td>4. Language used</td>
<td>Cell language</td>
<td>Human language</td>
</tr>
<tr>
<td>5. Current carrier</td>
<td>Chemicals</td>
<td>Electrons</td>
</tr>
<tr>
<td>6. Current conductor</td>
<td>Chemicals</td>
<td>Wires</td>
</tr>
<tr>
<td>10. Current controller</td>
<td>Enzymes</td>
<td>Transistors</td>
</tr>
<tr>
<td>11. Control mechanism</td>
<td>Conformons</td>
<td>Photons</td>
</tr>
<tr>
<td>12. Energy source</td>
<td>Chemicals</td>
<td>External voltage</td>
</tr>
<tr>
<td>13. Memory</td>
<td>Biopolymers, IDS’s</td>
<td>Flip-flops capacitors</td>
</tr>
<tr>
<td>14. Structural rigidity</td>
<td>Thermally fluctuating</td>
<td>Thermally robust</td>
</tr>
<tr>
<td>15. Physical size</td>
<td>Microscopic</td>
<td>Macroscopic</td>
</tr>
<tr>
<td>16. Programs stored in</td>
<td>Biopolymers</td>
<td>Softwares</td>
</tr>
<tr>
<td>17. Programmability</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>18. Self-reproducibility</td>
<td>Yes</td>
<td>Not yet</td>
</tr>
<tr>
<td>19. Model</td>
<td>Bhopalator</td>
<td>Turing machine</td>
</tr>
</tbody>
</table>

*a For detailed explanations, see text.
b Intracellular dissipative structures such as intracellular calcium waves (Sawyer et al., 1985), transmembrane gradients of ions, and phosphorylation state of proteins.
c The language used by the cell that is found to be isomorphic with human language (Ji, 1997a,b, 1999a).
d Sequence-specific conformational strains of biopolymers (e.g. DNA supercoils) (Benham, 1996a,b) that carry both free energy and genetic information.

e Because of the near-infinite number of conformational states available to biopolymers which determine their biological activities, and because biopolymers in turn determine cell states, it is impossible to describe all the possible internal states that a cell can exist in. This would logically lead to the conclusion that the behaviors of the cell cannot be programmed in the conventional sense (Conrad, 1988).

f The universal construction machine proposed by von Neumann (Arbib, 1988) can self-reproduce on paper but has not yet been physically implemented. Self-reproduction in the symbolic space does not guarantee self-reproduction in the material space. For a related topic, see the symbol-matter complementarity discussed by Pattee (1995).

g A theoretical model of the living cell based on the concepts of self-organization, conformons, and IDS’s, proposed in Bhopal, India in 1983 (Ji, 1985).
Conrad and Zauner (1998). The energy source in the cell is again chemicals, while that in computers is an external voltage source. The cell can store information in two contrasting ways—in stable sequence patterns of biopolymers (i.e. the Watson–Crick form of genetic information (Ji, 1988)) or in dynamic, dissipative concentration gradients (i.e. the Prigoginian form of genetic information (Ji, 1988)). The latter is also called gradients (i.e. the Prigoginian form of genetic information (Ji, 1988)) or in dynamic, dissipative concentration gradients (i.e. the Prigoginian form of genetic information (Ji, 1988)) or in dynamic, dissipative concentration gradients (i.e. the Prigoginian form of genetic information (Ji, 1988)) or in dynamic, dissipative concentration gradients (i.e. the Prigoginian form of genetic information (Ji, 1988)). The latter is also called IDS’s, or intracellular dissipative structures, exemplified by transmembrane or cytoplasmic gradients of ions (Sawyer et al., 1985). Computers store information in flip-flops and capacitors. Probably the most important difference between cell- and human-made computers is the fact that the former requires, and the latter is adversely affected by, thermal fluctuations. Because of the functional requirement for thermal fluctuations, the cell’s computational processes must operate on the microscopic scale, whereas human-made computers must work on the macroscopic scale to avoid short circuiting by thermal electrons. Computational programs in the cell are stored primarily in DNA (in the forms of the lexical, syntactic, and semantic genes (Ji, 1999a)), whereas those in human-made computers (in most cases) are located in software. The first comprehensive theoretical model of the cell known as the Bhopalator was proposed over a decade ago (Ji, 1985), which may be compared with the Turing machine in computer science. In passing, it should be pointed out that the recent proposal of Bennett (1982) that the complex of RNA polymerase and DNA can be viewed as a molecular Turing machine contrasts with the basic postulate presented here that the cell is the smallest molecular computer in existence (see the Fifth Law of Molecular Semiotics discussed in Section 9). If this view is correct, Bennett’s molecular Turing machine can only be viewed as a component of the molecular computer, just as a CPU is a component of man-made computers.

As evident in Table 1, the cell is a TMC, since it has the following three characteristics: (1) Molecularity—all the moving parts of the cellular computer that are responsible for input, output, and state transitions are thermally fluctuating molecules; (2) Internal Energy Source—all molecular motors are driven by the mechanical energy (i.e. conformons) stored inside molecular motors themselves which is ultimately derived from ATP and its equivalents; and (3) Computability—cells are endowed with the capacity to perform computation, namely, information processing according to a set of instructions stored in DNA.

A recent review of the extensive experimental data that exist in the literature concerning the phenomenon of the so-called programmed cell death (apoptosis) has led the present author to conclude that cells have evolved to carry out genetic programs of the general type, (xyz), which, when translated into human language, reads as follows: ‘When you are in state X and receive signal Y, then do Z’. The symbol x indicates the genes coding for enzymes that determine the internal state, X, of the cell (e.g. phosphorylation states of proteins, and cytoplasmic concentration gradients of ions, namely IDS’s (Ji, 1985, 1991)), y represents the genes coding for the receptors for signal Y (e.g. TNF, FasL), z stands for genes coding for enzymes necessary to execute the results of cellular computation (e.g. proteolysis) (Ji, 1997b), and the parentheses, ( ), symbolize the ‘spatiotemporal genes’ postulated to be located in noncoding regions of DNA and control the spatiotemporal evolution of gene expression (Ji, 1991, 1997b, 1999a). The temporally ordered expression of the set of genes, xyz—which automatically leads to the generation of unique IDS’s—corresponds to a cell-linguistic sentence (Ji, 1997b).

Therefore, it is suggested here that the minimum set of postulates necessary and sufficient to enable cells to carry out cellular computation consists of: (1) conformons that drive molecular motors; (2) intracellular dissipative structures (IDS’s) that are necessary to couple events between biopolymers and their environment on the one hand and between inside and outside of the cell on the other; and (3) cell language that enables cells to communicate in space and time. The first two postulates refer to the material aspect and the last one to the symbolic aspect of the cell—the two complementary aspects of all self-reproducing systems (see the ‘von Neumann–Pattee principle’ of matter–symbol complementarity discussed in Section 9).
3. Conformons in proteins and DNA

It is now well established that DNA can store free energy as mechanical (i.e. conformational) strains (Wang, 1996; Benham, 1996a,b). For example, one negative turn of a double-stranded DNA around its helical axis causes the separation of about 10 base pairs, thereby storing free energy as conformational strains in the amount of 5–6 Kcal/mol. Furthermore, Benham (1996a,b) has shown that conformational strains tend to localize in regulatory regions and the 3'-end of structural genes and rarely, if at all, inside structural genes. Such sequence-specific conformational strains were predicted to exist in biopolymers and referred to as 'conformons' (Green and Ji, 1972; Ji, 1974, 1985, 1990, 1998b). Conformons are molecular entities that contain both free energy and genetic information postulated to be necessary and sufficient for driving all purposeful, goal-directed molecular processes inside the cell. The free energy (8–16 Kcal/mol in proteins and 500–2500 Kcal/mol in DNA) and information (40–200 bits in protein and 200–600 bits in DNA) contents of conformons have been estimated from experimental data obtained from the literature (Ji, 1998b).

4. Intracellular dissipative structures

There are a large number of chemical reactions that are known to self-organize (i.e. 'spontaneously organize', or 'organize without any external energy input') in space and time (e.g. oscillating chemical concentrations inside a test tube, or formation of geometric patterns of chemical concentrations in a petri dish) (Babloyantz, 1986). All self-organizing chemical reaction–diffusion systems are named as 'X-ators', where X is the name of the city that is somehow connected with the mechanism of the reaction named. For example, the theoretical model that simulates the gross features of the Belousov–Zhabotinsky reaction is known as the Brusselator because of the theoretical contributions made by Ilya Prigogine and his group in Brussels (Prigogine, 1977, 1980).

Self-organizing chemical reactions can also occur inside the living cell, such as intracellular gradients or oscillations of calcium ion concentrations (Sawyer et al., 1985). In addition, these so-called IDS’s appear to drive or be coupled to cell functions such as chemotaxis and phagocytosis. Based on these observations, it was postulated in 1983 at a meeting in Bhopal, India that: (1) living cells can be treated as self-organizing chemical reaction–diffusion systems (hence the name, the Bhopalator); and (2) the ultimate form of expression of genes is not polypeptides as is commonly believed but IDS’s (Ji, 1985, 1991).

5. Cell language (Cellese)

Language is a means of communication. Since all cells must communicate to survive, develop, and function normally, they must possess a language of their own. Such a language was named 'cell language' (Ji, 1997a,b) or 'cellese' (Ji, 1998a) and defined as:

A self-organizing system of molecules, some of which encode, act as signs for, or trigger, gene-directed cell processes.

Cell language so defined has many similarities to human language (Table 2), leading to the conclusion that cell and human languages are isomorphic (i.e. similar in principle) (Ji, 1997a,b).

Double articulation (also called duality) is one of the most fundamental characteristics of all human languages (Martinet, 1960; Lyons, 1993). The principle of double articulation allows humans to create an almost infinite number of novel and meaningful sentences based on finite sets of lexicon and grammatical rules. It is because of this principle that humans can exhibit the so-called rule-governed creativity (Lyons, 1992)—another basic aspect of human language. Cell language is believed to possess double articulation as well—the first articulation being identified with the spatiotemporal organization of gene expression through the control of DNA folding patterns primarily through conformational (also called non-covalent) interactions; and the second articulation, with the linear arrangement of nucleotides to
Table 2  
A comparison between human and cell languages

<table>
<thead>
<tr>
<th></th>
<th>Human language</th>
<th>Cell language</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Alphabet</strong></td>
<td>Letters</td>
<td>Four nucleotides (or 20 amino acids)</td>
</tr>
<tr>
<td><strong>2. Lexicon</strong></td>
<td>Words</td>
<td>Genes (or polypeptides)</td>
</tr>
<tr>
<td><strong>3. Sentences</strong></td>
<td>String of words</td>
<td>Spatiotemporally organized expression of 'strings' of genes (or protein complexes)</td>
</tr>
<tr>
<td><strong>4. Grammar</strong></td>
<td>Rules of sentence formation</td>
<td>Rules mapping the physicochemical properties of nucleotides to folding patterns of DNA in response</td>
</tr>
<tr>
<td><strong>5. Meaning</strong></td>
<td>Things, ideas, concepts, and judgements referred to</td>
<td>IDS's produced by spatiotemporally organized expression of sets of genes, which drive gene-directed</td>
</tr>
<tr>
<td><strong>6. Mechanisms</strong></td>
<td>Physiological processes of phonation, audition,</td>
<td>cell processes</td>
</tr>
<tr>
<td></td>
<td>cognition, etc.</td>
<td>Self-organizing chemical reactions and actions of molecular motors (e.g. RNA polymerase)</td>
</tr>
</tbody>
</table>

*a* See text for more details.

form structural genes through covalent (or configurational) interactions (Table 3). Thus, the double articulation in cell language is critically dependent on the duality of molecular interactions—namely, noncovalent (or conformational) and covalent (or configurational) interactions.

### 6. From sequence space to function space

To account for the phenomenon of *Homo sapiens*, from DNA to language (including formal languages, mathematics, and semiotics (Eco, 1979; Deely, 1994)), it is necessary to invoke at least five spaces—(1) *Sequence Space*, (2) *Shape Space*, (3) *Chemical Reaction Space*, (4) *Function (or Biosemiosis) Space*, and (5) *Symbol (or Anthroposemiosis) Space* (see Fig. 1). The arrows indicate mappings, correlations, or mechanisms.

The mapping of the sequence space to the shape (i.e. 3-dimensional structural, or conformational) space is one-to-

n, where n is an immensely large number. For example, if one considers a polypeptide of just 100 amino acids, the minimum number of possible conformations is $2^{100}$ or $\sim 10^{30}$, assuming that each added amino acid residue can exist in at least two distinct conformations. If one makes the reasonable estimation that it takes the protein at least $10^{-9}$ s to undergo one conformational transition, it would take the protein $(10^{30})(10^{-9}) = 10^{21}$ s or $10^{12}$ years to undergo all possible conformational transitions. This is $\sim 10^2$ times the age of our Universe! Two conclusions can be drawn from this order-of-magnitude calculation: (1) The number of possible conformations available to a protein is almost infinite; and (2) it will take a protein an almost infinitely long time to find its minimum free energy conformation.

Although individual sequences have a *one-to-almost infinite mapping* to the shape space, it is possible to reduce the conformational degrees of freedom dramatically, through the mechanism of intermolecular conformational (i.e. noncovalent) interactions (now more popularly known as ‘protein–protein interactions’), leading to a near one-to-one mapping as will be explained in the next section.

The chemical reaction space can be divided into two subspaces—the *low activation-energy reaction space (LAERS)*, and the *high activation-energy reaction space (HAERS)*. The former consists of all the non-enzymic chemical reactions that occur spontaneously under physiological conditions, while the latter is comprised of those chemical reactions that do not occur at all under the same conditions. A subset of HAERS possesses the important property that these reactions do proceed spontaneously when bound stereospecifically to their conjugate enzymes. These reactions are called ‘enzymic reactions’, or
Table 3
Double articulation or duality of human and cell languagesa

<table>
<thead>
<tr>
<th>Human language</th>
<th>Cell language</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. First articulation</td>
<td>Organizing words to form sentences</td>
</tr>
<tr>
<td>2. Second articulation</td>
<td>Organizing gene expressions in space and time through predominantly noncovalent (i.e. conformational) interactions</td>
</tr>
<tr>
<td></td>
<td>Organizing letters to form words</td>
</tr>
<tr>
<td></td>
<td>Organizing nucleotides into DNA through covalent interactions (i.e. bond formations)</td>
</tr>
</tbody>
</table>

a See text for more details.

‘enzyme-catalyzed chemical reactions’. Of the numerous possible mappings allowed for by the principles of chemistry and physics, biological evolution appears to have selected predominantly a one-to-one mapping between the shape space and the chemical reaction space, through mechanisms similar to the one explained in the next section. This would make it possible, at least in principle, to realize nearly one-to-one mapping between the sequence space and the chemical reaction space.

The mapping between the chemical reaction space and the function space seems to vary from many-to-one (i.e. many enzymic reactions together effectuate a particular metabolic conversion, such as glycolysis, arachidonic acid cascade, oxidative phosphorylation, etc.) to one-to-many (e.g. ATP hydrolysis reaction driving many different functions such as active transport, muscle contraction, anabolic reactions, DNA supercoiling, etc.). The function space can also be called the biosemiosis or biosemiotic space, since all living processes are mediated by molecular signs, including DNA and proteins (Sebeok, 1976; Ji, 1997a,b, 1999a) and hence can be identified with biosemiosis (Deely, 1994).

The final mapping—from the function to the symbol space—is predominantly, if not exclusively, confined to the brain of Homo sapiens. Hence, we can refer to it as the anthroposemiosis space (Eco, 1979; Deely, 1994). Whether the emergence of anthroposemiosis has contributed to, is neutral with respect to, or is even detrimental to, the survival of the human species is not known. It is surprising to find that the symbol space appears to be mapped back onto the sequence space (i.e. the structures of human language and DNA are similar and correlated), as evident in the recently recognized isomorphism between cell and human languages (Ji, 1997a,b, 1999a). This suggests to this author that a full understanding of the structure and function of DNA may not be possible until and unless the principles of human language (Harris, 1993) and semiotics (Sebeok, 1976; Eco, 1979; Deely, 1994) are correctly assessed and applied to the molecular biology of DNA (Ji, 1997a, 1998a, 1999a,b).

7. ‘Collapse’ of the conformational degrees of freedom by intermolecular interactions

Though the number of potential conformations belonging to a sequence of a relatively small number of monomers (whether amino acids or nucleotides) is almost infinite as pointed out above, the actual number of conformations associated with a protein under physiological conditions of temperature, ionic strength, pH, etc. appears to be very small, most likely due to the selection imposed on the conformations of proteins by the microenvironmental conditions in the cell during their genesis. Let us denote the set of such biologically selected conformations belonging to sequence \( s_i \) with the symbols \( C(s_i) \). Then, two such sets of conformations, \( C(s_1) \) and \( C(s_2) \), belonging to sequences \( s_1 \) and \( s_2 \) can be depicted as two bases of inverted circular cones located on the shape space with their apexes resting on the sequence space (Fig. 2). The difference between the two sequences (as measured in terms of some metric such as the Hemming distance) is designated as \( d \). As \( d \) is reduced, or as \( s_1 \) and \( s_2 \) becomes similar in sequence structure, the
Fig. 1. The five spaces that are required to account for *Homo sapiens* from DNA sequences to human language. See text for detailed explanations.

Two circular bases will begin to overlap. When the two polymers interact noncovalently (i.e. conformationally), there are two extreme possibilities:

**Case 1.** The structure of the two sequences are such that the interacting surfaces exhibit little affinity for each other, leading to little or no binding free energy.

**Case 2.** The two interacting surfaces are complementary to each other, resulting in high-affinity binding interactions with large free-energy changes.

In the former case, the conformations belonging to the intersection of the two bases, namely \( C(s_1, s_2) = C(s_1) \cap C(s_2) \), will undergo little or no alterations, whereas, in the latter case, they will undergo net conformational changes as the two molecules interact, so that the probability of observing the altered conformations (or conformers) under physiological conditions will be high. Such a reduction or ‘collapse’ of the conformational degrees of freedom of biopolymers is reminiscent of, and may be fundamentally related to, the phenomenon of the collapse of the wave function upon measurements widely discussed in quantum mechanics (Cushing, 1994).

Therefore, it is possible, in principle, to design (or select) two sequences in such a manner that the membership (or cardinality) of the intersection set, \( C(s_1, s_2) \), is made arbitrarily small by adjusting the magnitude of \( d \). The stringent stereoselectivity (i.e. molecular recognition) exhibited by many protein–protein or protein–DNA interactions so universal in molecular biology may be best understood in these set-theoretic terms. This mapping implicates two diametrically opposed processes—the variation resulting from the conformational degrees of freedom inherent in flexible biopolymers and the selection of sequences (by cells) based on their ability to generate conformations capable of complementary intermolecular interactions essential for the survival of organisms.

The interaction between variation and selection that goes on inside the cell on the molecular level may be related to the variation–selection interactions that proceed on the level of biological evolution (Kauffman, 1993; Williams, 1966). If this is true, there is the interesting possibility that the
principle of variation and selection manifested on
the cellular and molecular levels are applicable
universally throughout the organizational hier-
archy in biology, leading to what may be called
the hierarchy of Darwinsisms—namely, the molecu-
lar, cellular, individual, specific, and suprasepecific
Darwinsisms. The term ‘Darwinism’ is here used as
a convenient symbol for designating the physical,
chemical and biological mechanisms of interac-
tions between variation and selection processes at
all levels of biological organization.

For designing molecular computers, the poten-
tial practical relevance of the mapping between
the sequence and shape spaces cannot be overem-
phasized. It should be pointed out that the
stereoselective conformational interactions intrin-
sic to the shape space corresponds, according to
the cell language theory, to the first articulation in
human language, while the covalent structurations
characteristic of the sequence space correspond to
the second articulation (Table 3).

8. Molecular mechanisms of gene expression

Both cell-linguistic sentences and sentences in
spoken language are 4-dimensional, since they
possess not only spatial (amplitude) but also tem-
poral characteristics (frequencies). In this sense,
they can be identified as examples of self-organi-
zation and dissipative structures of Prigogine
(Babloyantz, 1986). In contrast, nucleotide se-
quences in DNA and sequences of words in writ-
ten language are at most 2-dimensional, since they
can be completely described on a 2-dimensional
surface (e.g. on a piece of paper). Just as it is
impossible to reconstruct the spoken version of an
unknown language based solely on its written
form (I know this from my own experience, I did
not know how to correctly pronounce ‘the’ from
studying English grammar books; I learned the
sound value of ‘the’ only after I actually heard it
pronounced by American G.I’s during the Korean
War!), so I assert that it is impossible to construct
cell-linguistic sentences in action (e.g. mechanisms
of gene expression) based solely on nucleotide
sequences of DNA. This may be ultimately be-
cause it is impossible to reconstruct n-dimensional
entities from (n – 1)-dimensional ones, if the n-to-
(n – 1)-dimensional contraction entails an irre-
versible loss of information, as is usually the case,
except perhaps in holography.

One of the testable predictions of the cell lan-
guage theory is that DNA contains two classes of
genes—the trans-acting structural (or lexical)
genes and the cis-acting spatiotemporal (or seman-
tic and syntactic) genes (Ji, 1991, 1999a). Cell-lin-
guistic sentences are generated when a set of
structural genes are expressed coordinately at right times and places under the control of syntactic and semantic genes (resulting in the production of IDS’s)—all driven by the free energy stored in DNA and DNA-binding proteins as conformational strains, namely conformons. The physico-chemical properties of individual nucleotides and their polymeric derivatives provide the cell-linguistic grammar, distributed over the whole DNA molecule and controlling the 4-dimensional (i.e. space- and time-dependent) folding patterns of both structural and spatiotemporal genes in response to intracellular chemical and mechanical milieu (Ji, 1999a).

9. Conclusion: Towards the Laws of Molecular Semiotics or Microsemiotics

Semiotics is the study of signs that has its roots in linguistics, logic, and mathematics (Sebeok, 1976; Eco, 1979; Deely, 1994). It is generally accepted that the American scientist–logician–philosopher C.S. Peirce (1839–1914) (Hausman, 1997) played a pioneering role in formalizing semiotics in such a way that it can be applied to all sign processes in nature, including those in human societies (Eco, 1979; Deely, 1994) and biology (Sebeok, 1976). When Peirce was active in the second half of the last century, semiotics was concerned primarily with communications among human beings using linguistic signs on the macroscopic level. With the advent of molecular biology in the middle of this century, a new branch of semiotics, which may logically be called ‘molecular semiotics’, ‘microsemiotics’, ‘cellular semiotics’, or ‘cytosemiotics’, can be said to have been born—distinct from the traditional semiotics (Eco, 1979; Deely, 1994; Hausman, 1997), which may now be referred to as ‘macrosemiotics’. In ‘microsemiotics’, the main focus is on sign- or signal-dependent measurements (e.g. receptor–ligand interactions), computations (e.g. signal transductions), and communications (e.g. heredity), taking place on the molecular and cellular levels in animals (including humans), cells, and—thanks to Adleman—DNA molecules in test tubes.

In the interest of stimulating discussions and productive debates among emerging ‘molecular semioticians’, the following set of generalizations is offered here for the first time in the form of ‘laws’—laws in the tentative sense that these generalizations do possess the potential for becoming laws if they pass critical experimental testings.

10. The (Putative) laws of molecular semiotics

The First Law (the definition of cell language)

“It is impossible for cells to communicate without a language.”

The Second Law (isomorphism between cell and human languages)

“Cell language is isomorphic with human language.”

The Third Law (the gnergy principle; see also the ‘von Neumann–Pattee principle’ (Ji, 1999a))

“It is impossible for cells to communicate without information and free energy.”

“It is impossible for cells to communicate without information-energy.”

1 The notion of matter–symbol complementarity formulated by Pattee (1995) based on the earlier work of von Neumann on self-reproducing automata (hence called the von Neumann–Pattee principle for self-reproduction, for convenience) applies to all self-reproducing systems, both microscopic and macroscopic. The von Neumann–Pattee principle is deemed necessary but not sufficient for biological self-reproduction. It is postulated here that only the gnergy principle (i.e. the Third Law of Microsemiotics) provides both the necessary and sufficient conditions for biological self-reproduction. The ‘matter-symbol complementarity’ may be more generally called the ‘matter-sign complementarity,’ since, according to the American chemist–logician–philosopher C.S. Peirce (1839–1914), sign includes symbol as well as icon and index.
“It is impossible for cells to communicate without energy.”

The Fourth Law (the enzyme principle)

“It is impossible for cells to communicate without enzymic catalysis.”

The Fifth Law (the cell principle)

“Cells are the smallest material entities that can self-reproduce.”

“Cells are the smallest material entities that can communicate and compute.”

“Cells are the smallest material entities that can develop (as individuals) or evolve (as groups).”

Acknowledgements

I thank J. Lee, J.L. Colaizzi, R. Snyder, F. Kaufman, B. Goldstein, and many other colleagues, post docs, and graduate students at Rutgers, UMDNJ/Robert Wood Johnson Medical School for their encouragement, stimulating discussions, and critical comments. My particular thanks to Matt Hoffmann for reading the manuscript and suggesting improvements.

References


The abstract at http://www.cifn.unam.mx/Computational_Biology/w98/program.html


