

Mathematical (Quantitative) and Cell Linguistic (Qualitative) Evidence for Hypermetabolic Pathways as Potential Drug Targets

Ji S*

Department of Pharmacology and Toxicology, Ernest Mario School of Pharmacy, Rutgers University, Piscataway, New Jersey, USA

Abstract

Objectives: According to the cell language theory first proposed in 1997, living cells use a molecular language whose structure is similar to (or isomorphic) with the structures of the human language with respect to the 10 out of the 13 design features established by linguists. One of the predictions of the cell language theory is that there should exist in the living cell what is referred to as 'hypermetabolic pathways' that correspond to texts in human language deemed essential for reasoning and computing. A mathematical method known as the Planck-Shannon plot is described that can be employed to identify the predicted hypermetabolic pathways that underlie human breast cancer and hence can serve as potential anti-cancer drug targets.

Data and analytic method: The gene expression profile data measured with microarrays were provided by Perez-Ortin's group in Valencia, Spain and Perou and his coworkers at Stanford University. The mRNA data were transformed into histograms which were then fitted to the Planck Distribution Equation (PDE $y = \left(\frac{A}{(x+B)^5} \right) / (e^{C/(x+B)} - 1)$), to generate the numerical values for the parameters, A, B and C, that quantitatively characterize the shape of each histogram and hence the information contained in the original mRNA data set. The fitting of mRNA data to PDE was performed by the Sovler program available in Excel.

Results: The hypermetabolic pathways, both intra-organismic, and inter-organismic, that are predicted by the cell language theory can be identified with the PDE-based analysis of mRNA data. The intra-organismic hypermetabolic pathway identified with PDE consists of 3 or more traditional metabolic pathways, while the inter-organismic hypermetabolic pathway consists of one traditional metabolic pathway whose activity is correlated among 3 or more organisms exhibiting a common phenotype, e.g., breast cancer.

Conclusion: Ribonoscropy, defined as the genome-wide study of mRNA levels within an organism or between different organisms, when combined with the quantitative method of analysis afforded by the Planck Distribution Equation (PDE), can identify a novel class of metabolic structures referred to as "intra-organismic hypermetabolic pathways" and "inter-organismic hypermetabolic pathways" that can serve as potential targets of cancer drug therapy.

Keywords: Gene expression profiles; Cell-linguistic analysis of gene expression profiles; Hypermetabolic pathways; Planckian distribution equation; Planckian information; Shannon entropy; Planck-Shannon plot

Introduction

Ilya Prigogine (1917-2003) divides all structures in the Universe into two classes – the equilibrium and dissipative structures. The former is exemplified by rocks, chairs, nucleotide sequences of DNA in a test tube, and 3-dimensional X-ray structures of enzymes and DNA molecules, etc. that can exist without using up energy, while the latter is exemplified by clouds, tornados, candle flames, action potentials across cell membranes, calcium ion gradients in the cytosol, life itself, etc. that cannot exist without dissipating free energy into heat [1,2]. For convenience, equilibrium and dissipative structures are referred to as 'equilibrons' and 'dissipatons', respectively in [3].

Based on the concepts of equilibrium and dissipative structures, it is possible to define three distinct approaches to drug discovery researches as summarized in Table 1. The Top-Down approach of herbal medicine, for example, considers as drug targets equilibrium structures such as macroscopically visible disease symptoms of various kinds, with an estimated success rate of discovering a drug of about 1 in a million. The Bottom-Up approach of molecular pharmacology targets 3-dimensional structures such as enzyme active sites determined by X-ray crystallography with the well-known success rate of 1 out of 100,000 [4]. The third approach referred to as "the hybrid" or "the complementary medical" approach combines both the

macroscopic top-down and the microscopic bottom-up approaches, with the anticipated success rate of about 1 in 1000 [5]. These ideas are summarized in Table 1.

The 16th century Swiss physician, alchemist and astrologer, Paracelsus (1493/4-1541) is famous for having stated that

"The dose makes the medicine." (1)

The dose, when interpreted as the blood level of a substance, is an example of the dissipative structure of Prigogine, since the movement of a substance, say, X, into and out of the blood compartment entails dissipating free energy into heat (Figure 1). This means that [X], the blood concentration of substance, X, at time t, is a dissipative structure or a dissipaton [3].

It is evident that there are at least two kinds of dissipative structures in blood:

***Corresponding author:** Dr. Sungchul Ji, Department of Pharmacology and Toxicology, Ernest Mario School of Pharmacy, Rutgers University, Piscataway, New Jersey-08854, USA, Tel: +1 848 445 2675; E-mail: sji.conformon@gmail.com

Received March 05, 2018; Accepted March 29 2018; Published April 02, 2018

Citation: Ji S (2018) Mathematical (Quantitative) and Cell Linguistic (Qualitative) Evidence for Hypermetabolic Pathways as Potential Drug Targets. J Mol Genet Med 12: 343 doi:10.4172/1747-0862.1000343

Copyright: © 2018 Ji S. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

The “right” dissipative structures that promote health (e.g., normal blood glucose levels) and the “wrong” dissipative structures (e.g., excessive blood glucose levels) that are harmful to health. (2)

Thus, we can re-express Paracelsus’ dictum as follows:

“The wrong dissipative structure makes the poison.” (3)

Since Statement (3) is true, by definition (and experience), its opposite statement, (4), must also be true:

“The right dissipative structure makes the medicine.” (4)

This statement is theoretically isomorphic with the following dictum stated by L. Pauling (1901-1994) that underlies orthomolecular medicine:

The functioning of the brain is affected by the molecular concentrations of many substances that are normally present in the brain. The optimum concentrations of these substances for a person may differ greatly from the concentrations provided by his normal diet and genetic machinery. Biochemical and genetic arguments support the idea that orthomolecular therapy, the provision for the individual person of the optimum concentrations of important normal constituents of the brain, may be the preferred treatment for many mentally ill patients (5).

It is reasonable to assume that the term “brain” in Statement (5) can be replaced by the more general term, “body”, without losing the general validity.

Approach (examples; estimated success rates)	Drug Target	
	Equilibrium Structures (also called <i>Equilibrons</i> [3]); e.g. macroscopic disease symptoms, microscopic 3-dimensional structures of receptors, enzymes, DNA, etc.	Dissipative Structures (also called Dissipatons [3]); e.g., action potentials, metabolite gradients in the cytosol, life, etc.
1. Top-down (e.g., herbal medicine; 1 out of 10 ⁶ ?)	-	+
2. Bottom-up (e.g., molecular pharmacology, receptor pharmacology; 1 out of 10 ⁵ [5])	+	-
3. Hybrid (or <i>Complementary medical</i>) (e.g., PDE-based ribonoscopy [4, 5]; 1 out of 10 ² ?)	+	+

Table 1: The three main approaches to drug discovery [5].

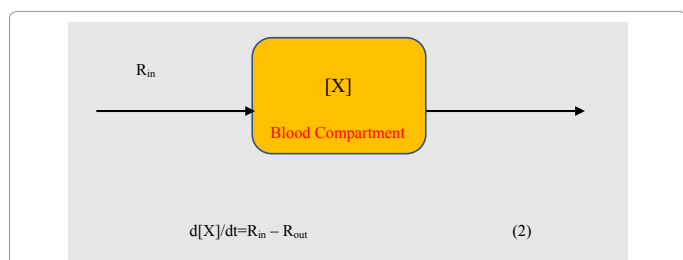


Figure 1: A simplified diagram (upper panel) is representing the mathematical definition (lower panel) of the dose as the balance between the rates of input into and output from the blood compartment of a substance, X. R=the rate of the movement of substance X in or out of the blood compartment of the human body; i.e., the amount of X moved during a given time period, $R = d[X]/dt$, where d[X] and dt can be infinitesimally small quantities.

Variables	Paracelsus (1493/4-1541)	Pauling (1901-1994)	Prigogine (1917-2003)
Time	16 th century	20 th century	20 th century
Original dictum	The dose makes the poison.	Optimum molecular concentrations of substances normally present in the body. [43].	Life is a dissipative structure.
Dictum re-expressed in the language of irreversible thermodynamics	The wrong dissipative structure makes the poison	The right dissipative structure makes the medicine.	The right dissipative structure makes the medicine.
Paracelsus-Pauling-Prigogine Paradigm of Drug Discovery (or the P ³ paradigm)	Dissipative structures can be drug targets.		

Table 2: The derivation of the Paracelsus-Pauling-Prigogine (P³) Paradigm of Drug Discovery (or the P³ paradigm).

Statement (6) below made by Prigogine when I visited him at the University of Texas Austin in or round 1984, I think, is also true and consistent with both Statements (3), (4) and (5):

“Life is a dissipative structure.” (6)

It appears to me that Statements (3) through (6) are consistent with and support the following generalization:

“Dissipative structures can be the targets of drug actions.” (7)

A corollary of Statement (7) would be:

“There exist drugs that target dissipative structures.” (8)

Statement (8) is supported by the finding that doxorubicin can target hypermetabolic pathways. These loosely related general statements are collected in Table 2 to reveal the hidden common threads.

The main objectives of this paper are three-fold –

(i) To provide indirect evidence based on the analogy between cell and human languages that there exists a whole new class of structures in biology that underlie biological functions and hence can serve as drug targets [6-9].

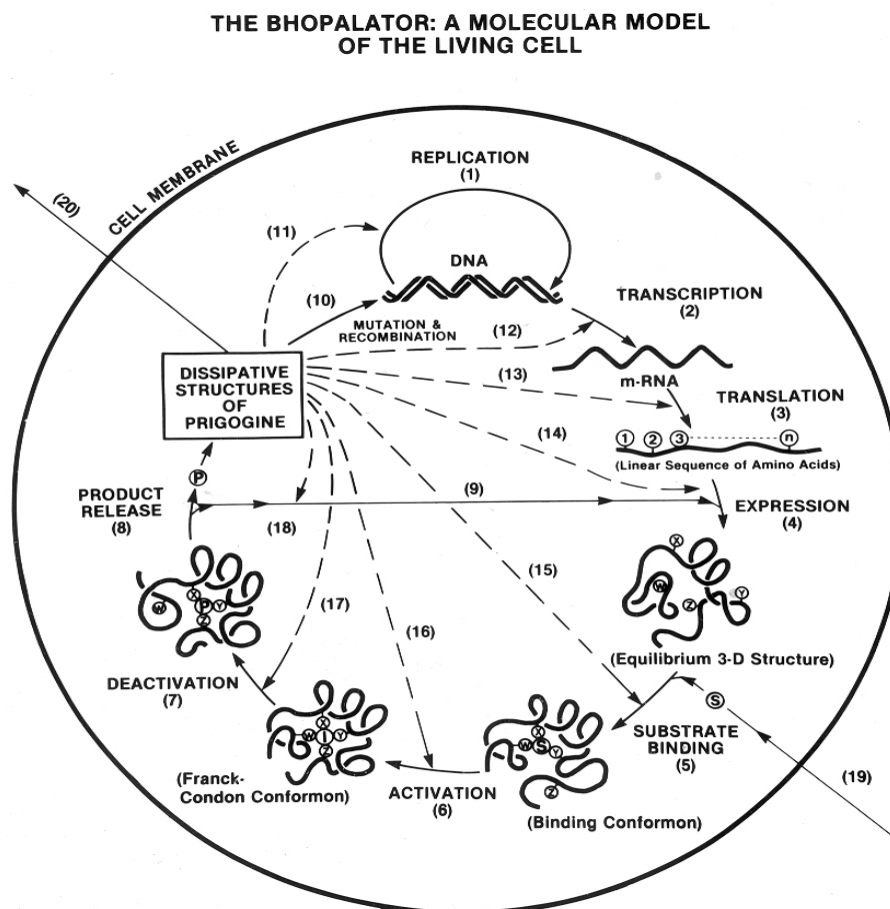
(ii) To introduce a new mathematical equation, the Planckian Distribution Equation, that can identify certain dissipative structures that are related to what V. Norris called “hyperstructures” in 1999 [10-13].

(iii) To present the microarray evidence that there are sets of metabolic pathways referred to as ‘hypermetabolic pathways’ that are associated with drug-responsive breast cancer.

(iv) The bhopalator model of the living cell and the cell language.

Pharmaceutical scientists who are trying to design a drug without a theoretical model of the living cell, I suggest, is akin to atomic physicists who are trying to explain atomic spectra without a theoretical model of the atom such as Bohr’s atom or its more modern versions [14].

Although it had been known since the mid-19th century that the cell is the smallest unit of the structure and function of all living systems [15], it was apparently not until 1985 that the first comprehensive theoretical model of the cell was proposed [16-18]. In that year, a theoretical model of the living cell called the Bhopalator (Figure 2 and Table 3) appeared in which both the energetic and informational aspects of the living cell were integrated on an equal footing, based on the supposition that life is driven by gnergy, the complementary union



Note: (Upper panel) The Bhopalator - A molecular model of the living cell. Adopted from [16-18]. The cell is viewed as the physical system wherein micro-meso correlations occur under a wide variety of environmental conditions supported by free energy utilizing enzymes acting as molecular machines [19]. The Bhopalator consists of a total of 20 major steps: 1= DNA replication; 2=transcription; 3= translation; 4=protein folding; 5=substrate binding; 6=activation of the enzyme-substrate complex; 7=equilibration between the substrate and the product at the transition state; 8=product release contributing to the formation of the intracellular dissipative structures (IDSs); 9=recycling of the enzyme; 10=IDS-induced changes in DNA structure; 11 through 18= feedback interactions mediated by IDSs; 19=input of substrate into the cell; and 20=the output of the cell effected by IDSs, which makes cell function and IDSs synonymous.

(Lower panel) Isomorphism between cell and human languages [7-9].

¹Just as verbal sentences (as written) are strings of words arranged linearly in the Euclidean space, so the cell-linguistic (or molecular) sentences are visualized as series of gene expression events arranged in space and time leading to dissipative structures or dissipatons [3].

²Of all the folds of DNA and polypeptides allowed for by the laws of physics and chemistry, only small subsets have been selected by evolution (thereby giving rise to biological information) to constitute the genome of a cell.

³Sequence-specific conformational strains that carry both free energy (to do work) and genetic information (to control work) [20,21]. Conformons are akin to molecular batteries that provide immediate driving force (or serve as the force generators) for all molecular machines catalyzing non-random molecular processes inside the cell. Experimental evidence for conformons [3].

⁴Space- and time-specific intracellular gradients of ions, biochemicals, and mechanical stresses (e.g., of the cytoskeletal system) that serve as the immediate driving forces for all cell functions on the microscopic level [3].

⁵Also called "conformational" interactions which involve neither breaking nor forming covalent bonds and depend only on the rotation around, or bending of, covalent bonds. Non-covalent interactions implicate smaller (free) energy changes (typically around 1 to 3 Kcal/mole) than covalent interactions which entail (free) energy changes in the range of 30-100 Kcal/mole.

⁶Molecular interactions that involve changes in covalent bonds, i.e., changes in valence electronic configurations around nuclei of atoms within a molecule.

⁷This row is added to the original table published in [7,8]. The third articulation [22] is a generalization and an extension of second articulation. Intercellular communication through chemical concentration gradients is well established in microbiology in the phenomenon of quorum sensing [3,23,24], whereby bacteria express a set of genes only if there are enough of them around so that they can combine and coordinate their efforts to accomplish a common task which is beyond the capability of individual bacteria. This phenomenon can be viewed as a form of reasoning and computing on the molecular level and the cell therefore can be viewed as the smallest DNA-based computational unit [25], which may be referred to as the computon.

Figure 2: The Bhopalator model of the living cell and its molecular language.

Variables	Human Language (Humanese)	Cell Language (Cellese)
1. Alphabet (L)	Letters	4 Nucleotides (or 20 Amino acids)
2. Lexicon (W)	Words	Genes (or Polypeptides)
3. Sentences (S)	Strings of words	Sets of genes (or polypeptides) expressed (or synthesized) coordinately in space and time dictated by DNA folds ¹ (cell states).
4. Grammar (G)	Rules of sentence formation	The <i>physical laws</i> and <i>biological rules</i> mapping DNA sequences to folding patterns of DNA (polypeptides) under biological conditions ² .
5. Phonetics (P)	Physiological structures and processes underlying phonation, audition, and interpretation, etc.	Concentration and mechanical waves responsible for information and energy transfer and transduction driven by <i>conformons</i> ³ and <i>intracellular dissipative structures</i> (IDSs) ⁴ .
6. Semantics (M)	Meaning of words and sentences	<i>Codes</i> mapping molecular signs to gene-directed cell processes
7. First Articulation	Formation of sentences from words	Organization of gene expression events in space and time through <i>non-covalent interactions</i> ⁵ between DNA and proteins (or Space- and time-dependent non-covalent interactions among proteins, DNA, and RNA molecules). Thus, macromolecular complexes can be viewed as molecular analogs of sentences.
8. Second Articulation	Formation of words from letters	Organization of nucleotides (or amino acids) into genes (or polypeptides) through covalent interactions ⁶ .
9. Third Articulation	Formation of texts from sentences	Organization of chemical concentration gradients in space and time called dissipative structures [3,27,28] or dissipatons in order to 'reason' and 'compute' ⁷ .

Table 3: A formal comparison between human and cell languages [7,8,9].

Formal Aspect (Function)	Material Aspect (Function)	DNA Language (DNese; Information transmission in time)	RNA Language (RNese; Information transmission in space, from DNA to proteins)	Protein Language (Proteinese; Energy transduction from chemical to mechanical; i.e., conformon production)	Chemical Language (Chemicelese; Source of free energy)
Letters (Basic building blocks)		4 ⁿ nucleotides n = 1, 2, 3, 4, ... Exons (?)		Protein domains	Partial chemical reactions
Words (To denote)		Genes		Proteins	Full chemical reactions
Sentences (To decide)		cis-Genes (?)*		Metabolic pathways	Chemical gradients
Texts (To reason)		trans-Genes (?)*		'Hypermetabolic pathways'	Chemical waves (?)

Note: *cis-Genes are here defined as those genes covalently linked to each other and hence being in the same chromosome, whereas trans-genes are defined as those genes that are located in different chromosomes and yet can interact with one another through non-covalent interactions.

Table 4: The formal and material aspects of the cell language (Cellese).

of information and energy [3,18,19]. The name Bhopalator reflects the fact that the cell model was born as a result of the lectures that I presented at the international conference entitled The Seminar on the Living State, held in Bhopal, India in 1983, ably organized by Prof. R. K. Mishra of the All India Institute of Medical Science, New Delhi. The suffix, "-ator" indicates that the model assumes that the cell is a self-organizing chemical reaction-diffusion system (i.e., a dissipative structure or a dissipaton) [3].

The Bhopalator model of the cell consists of a set of arrows (i.e., directed edges) and nodes enclosed within a 3-dimensional volume delimited by the cell membrane (Figure 2). The system is thermodynamically open so that it can exchange matter and energy with its environment [19,20]. The arrows indicate the directional flows of information driven by free energy dissipation. The solid arrows indicate the flow of information from DNA to the final form of gene expression postulated to be the dissipative structures theoretically investigated by Prigogine and his schools [21-27].

It is noteworthy that nothing is new in the Bhopalator model of the cell, except the concept of the intracellular "dissipative structures" (IDSs) of Prigogine (Figure 2), which can be viewed as including the "hyerstructures" of Norris et al. [13]. Until now, there has been no mathematical method to characterize IDSs, and it is one of the main objectives of this paper to present one that is based on the Planckian Distribution Equation (IDS) derived from the blackbody radiation equation discovered by M. Planck in 1900 that revolutionized physics in the 20th century [11].

The Cell Language Theory (CLT)

The lower panel of Figure 2 compares the cell and human languages at 9 different levels of organizations [7-9]. In contrast, Table 3 analyzes the cell language based on the principle of matter-form complementarity, the top row representing the material aspect and the left-most column representing the formal aspect. Probably the most significant features of Table 3 are:

(i) The cell language consists of 4 sub-languages (called DNese, RNese, proteinese, and chemicelese), and

(ii) These sub-languages have distinct functions (as indicated in the first row), all of which are essential for living cells to communicate with one another in space and time, driven by the free energy supplied by the chemical reactions that enzymes catalyze.

(iii) To the best of my knowledge, Table 4 provides for the first time the principle-based rationale for the existence of the 4 molecular sub-languages in the living cell that are mediated by DNA, RNA, proteins, and biochemicals, the four material components that constitute the Bhopalator.

(iv) There are many empty boxes in Table 4 yet to be filled. The most well-established sub-language may be the protein language consisting of well-established concepts of "domains" which are thought to correspond to letters, "proteins" to words, "metabolic pathways" to sentences, and finally "hypermetabolic pathways" to texts, for the last of which experimental evidence will be presented below. Thus, the protein language may provide a valuable guide for inferring the content of the empty or uncertain boxes in other sub-languages.

Human Language	Cell Language		
	Structure of molecular alphabets	Function	
Alphabets	1 st -order	4 ¹ = 4 singlets (A, C, G, T)	Encoding 1-nucleotide frame shift?
	2 nd -order	4 ² = 16 doublets (AC, AG, AT, CA, CG, CT, GA, etc.)	Encoding 2-nucleotide frame shift? DNA shape code [31], chromatin code [31]
	3 rd -order	4 ³ = 64 triplets (AAA, AAC, AAG, AAT, ACA, etc.)	Encoding amino acids, stop, and start codons
	4 th -order	4 ⁴ = 256 tetrads (AAAA, AAAC, AAAAG, AAAAT, etc.)	Translation frame code? [31]?
	5 th -order	4 ⁵ = 1024 pentads (AAAAA, AAAAC, AAAAG, AAAAT, etc.)	Translation frame code? [31]?
Words	Genes		Encoding the primary structure of proteins (e.g., insulin)
Sentences	Gene systems		Encoding systems of enzymes catalyzing metabolic pathways (e.g., glycolysis)
Texts	Systems of gene systems		Encoding systems of metabolic pathways working as functional units (e.g., chemotaxis)

Table 5: The multiple genetic alphabet (MGA) hypothesis. The structure and function of the cell language inferred on the basis of the postulated isomorphism between human and cell languages and the role of vibrational resonances in genetic structures [7-9,29,30].

(v) It is generally assumed in the current biochemistry and molecular biology text books that there is only one genetic alphabet consisting of 4 nucleotides whose bases are adenine (A), cytosine (C), guanine (G) and thymine in DNA (or uracil in RNA). In stark contrast, Table 4 assumes that there is n (with $n=1 \sim 10^3$) genetic alphabets (named the n^{th} -order alphabet), each containing 4^n letters and each letter in turn consisting of n nucleotides (Table 5). In this view, the 64 codons are the so-called 3rd-order letters, not words as widely assumed. There are evidences that each of the multiple genetic alphabets postulated here may have distinct biological functions, some of which have been discovered by Trifonov [28-31]. If this interpretation is correct, what Trifonov refers to as “multiple codes” would be related to what is here called the “multiple genetic alphabets”.

What is the Planckian Distribution Equation (PDE)?

The Planckian Distribution Equation (PDE) can be viewed as a generalization of the blackbody radiation equation discovered by Planck (1858-1947) in 1900 that led to the development of quantum mechanics around 1925, revolutionizing physics in the 20th century.

Blackbody radiation refers to the emission of photons by material objects that completely absorb photons impinging on them. An example of the radiation from a heated object is given in Figure 3(a) which shows emission of different color (i.e., different wavelength light) as a function of temperature which varies on the surface of the lava. When the light intensity of a blackbody is measured at a fixed temperature, the so-called “blackbody radiation spectrum” is obtained as shown in Figure 3(b). Max Planck (1858-1947) succeeded in deriving the mathematical equation given in Figure 3(c) Equations (8.1) that quantitatively accounted for the blackbody radiation spectra [32-34]. The key to his success in deriving the so-called Planck Radiation Equation (PRE) was his assumption that light is emitted or absorbed by matter in discrete quantities called “quanta of action”. When Planck discovered PRE-in 1900, he probably could not have imagined that his equation one day might be extended beyond physics to biology and related fields implicating temperatures far lower than those required for blackbody radiation. However, since 2008 (reviewed in [3]), Planck’s radiation equation, Eq. (8.1) in Figure 3, when generalized in the form of what has been variously referred to as the blackbody radiation-like equation (BRE), the generalized Planck equation (GPE), or the Planckian Distribution Equation (PDE) (see Eqs. (8.2) and (8.3) in Figure 3), has been found to fit not only the long-tailed histograms generated from atomic physics (i.e., the blackbody radiation spectra) but also those generated from or associated with (i) protein folding, (ii) single-molecule enzyme catalysis, (iii) genome-wide RNA levels measured in yeast, (iv) genome-wide RNA levels measured in human breast tissues, (v) human T-cell receptor gene sequence diversity, (vi)

7-mer frequency distribution in *Pyrococcus abyssi*; (vii) the codon profile in the human genome; (viii) protein length frequency distribution in *Haemophilus influenzae*; (ix) brain neuroarchitectural changes induced by stress in rats; (x) electrocorticographic responses of the olfactory cortex to impulses; (xi) functional magnetic resonance imaging (fMRI) signals from the human brain before and after the infusion of the hallucinogen, psilocybin; (xii) sentence-length frequency distribution in private letters; (xiii) word-length frequency distribution in English texts; (xiv) word-length frequency distribution in John Kerry’s speech in 2004; (xv) The F0 histogram of the reading sound of a book; (xvi) the decision-time histogram; (xvii) the 1996 US annual income distribution; (xviii) the 2013 US annual income distribution; and the polarized cosmic microwave background radiation (Eq. 8.6).

It is suggested that the Planckian distribution equation (PDE), either the 4- or 3-parameter version, i.e., Eqs. (8.2) or (8.3) in Figure 3(d), respectively, is a new distribution law, comparable to the Gaussian distribution equation (GDE), that applies to a wide range of experimental data as does GDE. One plausible explanation for this seeming universality of PDE may be that, underlying all the so-called Planckian processes (defined as the physicochemical processes generating data that fit PDE [37]), there are common physical processes mediated by ‘standing waves’ (electromagnetic, gravitational, mechanical, and concentration) as represented by the first term in the Planckian distribution law (Figure 3(e)). The number of standing waves present within a system is determined by the volume and topology of the system being heated, as schematically represented in Figure 4.

The Wave-Particle Duality in Biology and Medicine

Since the blackbody radiation equation (BRE) consists of two components – the first term related to the number of standing waves and the second term related to the average energy of the standing waves [33] and since the Planck Distribution Equation (PDE) has the same mathematical form as BRE, it is assumed that the same interpretation of the two terms of BRE applies to those of PDE, as indicated in Figure 3(d). If this postulate is valid, it may be inferred that the wave aspect (which is related to the global information of the system under consideration) of the wave-particle duality would play a role as important in biomedical sciences as the particle aspect which is related to local energy production from individual enzymes inside living cells.

All dissipative structures may be viewed as “wave packets”, involving (i) electromagnetic waves, (ii) mechanical waves (e.g., sounds, conformational waves in DNA, RNA, and proteins), (iii) chemical waves (e.g., calcium waves in muscle cells, action potentials), and/or gravitational waves. Since the frequency and the shape of standing waves are well known to be determined by the mass and

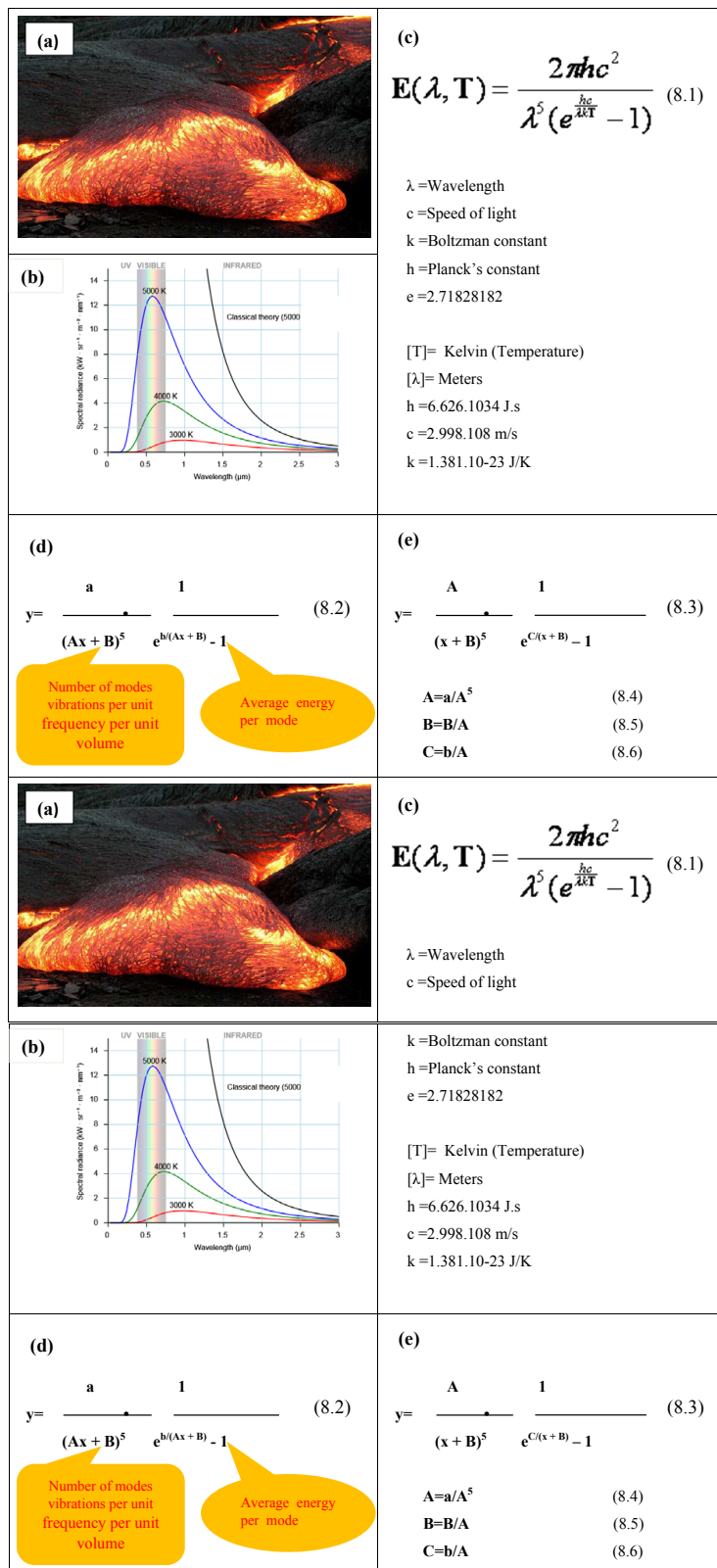


Figure 3: Reproduced from [9]. (a) Blackbody radiation. (b) The blackbody radiation spectra. (c) The Planck radiation equation. Reproduced from [33]. (d) The blackbody radiation-like equation or BRE [3], also called the generalized Planck equation (GPE) or the Planckian Distribution equation (PDE). The interpretation of the two terms were reproduced from <http://hyperphysics.phy-astr.gsu.edu/hbase/mod6.html>. (e) The 3-parameter (colored red) version of BRE/GPE/PDE. The relations between the 4- and 3-parameter versions of BRE/GPE/PDE are given in Figure 3 Eqs. (8.4), (8.5), and (8.6).

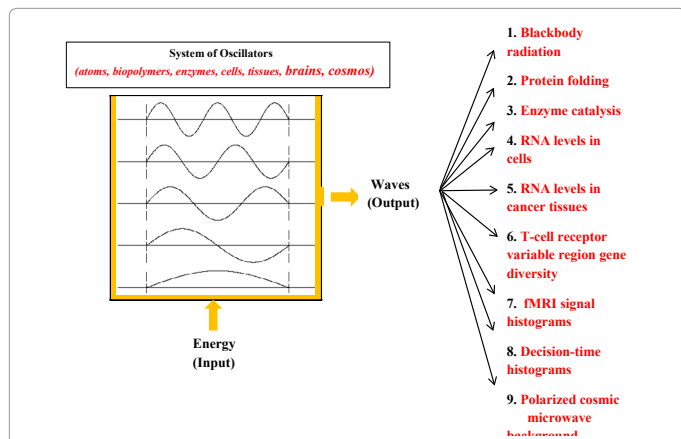


Figure 4: The possible common mechanism underlying the Planckian processes, i.e., those processes that generate numerical data that fit PDE. The material system embodying the Planckian processes is represented as a system of oscillators (i.e., atoms, biopolymers, enzymes, cells, tissues, brains, cosmos) that generate standing waves powered by the input energy. Depending on shapes of the standing waves and their average energies, different observables are thought to be outputted: 1=Blackbody radiation; 2=Protein folding; 3=Enzyme catalysis; 4=RNA levels in cells; 5=RNA levels in cancer tissues; 6=T-cell receptor variable region gene diversity; 7=fMRI signal histograms; 8=Decision-time histograms; 9=Polarized cosmic microwave background. Adopted from [9,32].

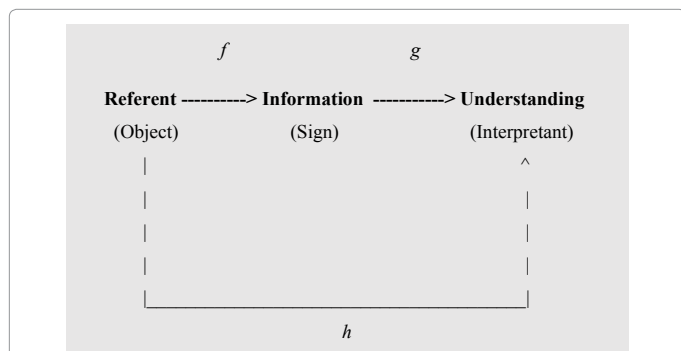


Figure 5: The arbitrariness of the term *information*. That is, the term 'information' (viewed as a Peircean sign) can mean (i.e., can refer to) anything as long as it is understood by the interpreter of the sign. f=sign production; g=sign interpretations; h=correlation, grounding, or information flow [9] (Figure 6.1). The terms in blue letters are those of semiotics (the study of signs) developed by the American chemist, logician and philosopher, Charles Sanders Peirce (1839-1914).

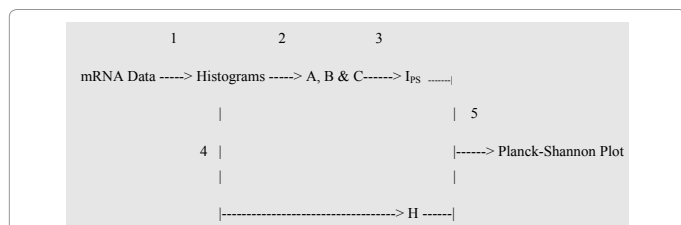


Figure 6: The 5-step analysis of mRNA data based on PDE. 1=Histogram software in Excel; 2=Fitting of mRNA data to PDE (Planckian Distribution Equation), (8.3) in Figure 3, implemented by the Solver program in Excel; 3=computed based on Eq. (12); 4=Computed based on Eqs. (9) and (10); 5=Scatter plot in Excel. A, B and C=the parameters of PDE; I_{PS} =Planck information of the second kind, Eq. (11); H=Shannon entropy, Eqs. (11) and (12).

geometry of the oscillator [35,36], many of the numerical regularities revealed by the nucleotide sequence structures and the atomic numbers of DNA (viewed as an organized system of oscillators obeying the Fourier theorem) that Petoukhov [29] and others have uncovered may find natural explanations in the language of the wave-particle duality

embodied in PDE (Figure 3). Thus, one possible way to account for the universality of the Planckian distribution equation (PDE) in nature is to postulate that the wave-particle duality first discovered in atomic physics operates at all scales of material systems, from atoms to the Universe [32] as schematically depicted in Figure 4.

Planckian Information of the Second Kind (I_{PS})

One mechanism of generating PDE from Gaussian distribution is what I call the "Rutgers University Admissions Mechanism" (RUAM). If RAUM does not take into account the students' heights in their admissions process, the height distribution of the RU students would be most likely Gaussian. However, if RUAM favors short students over tall ones, the RU students' height distribution will be skewed from the normal curve thus producing a long-tailed histogram that will most likely fits PDE. The degree of skewness of PDE from its Gaussian counterpart (with an equal area under the curve) can be used as a measure of the information used by RAUM in selecting RU students. The information derived from PDE based on its skewness will be referred to as the Planckian information of the second kind, I_{PS} , defined by (Eq. 9) to be distinguished from the Planckian information defined previously (Eqn 10), i.e., Planckian information of the first kind, I_{PF} .

$$I_{PS} = -\log_2((\mu - mode)/\sigma) \quad (9)$$

where μ and σ are the mean and the standard deviation of the long-tailed histogram under consideration.

We have found that some experimental data (e.g., digitized water wave patterns produced by the sonified Raman spectral bands measured from single cells) that fit PDE are better modeled with I_{PF} and some others (e.g., the mRNA levels measured from yeast cell ensembles) are better modeled with I_{PS} .

These observations indicate that:

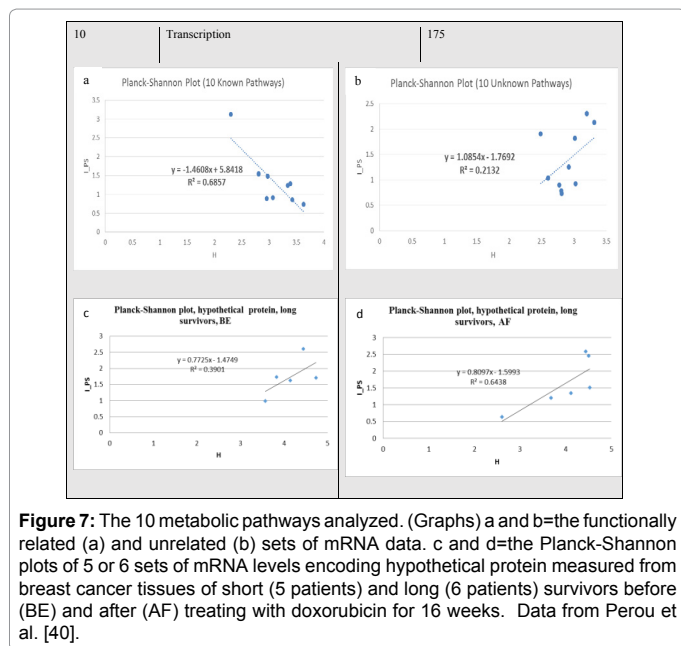
(a) There can be more than one kind of information that can be defined based on the same empirically derived mathematical equation, probably depending on underlying physical mechanisms.

(b) The reasoning in (a) suggests that the mathematically defined "information" is arbitrary in the sense of Saussure's arbitrariness of signs [37-39].

(c) The mathematically defined "information" can be viewed as a sign in the Peircean sense [40-44] and hence is irreducibly triadic as depicted in Figure 5.

Data and Analysis

The experimental data analyzed in this paper were published by M. Perou and his coworkers who measured variation in gene expression patterns in a set of 65 surgical specimens of human breast tumors from 42 different individuals, using complementary DNA microarrays representing 8,102 human genes [40]. Twenty of the tumors were sampled twice, before and after 16-week doxorubicin chemotherapy. We analyzed the latter data by first transforming them into histograms (using Excel program) which were then fitted into PDE utilizing the Solver program available in Excel. The key steps involved in analyzing mRNA data based on PDE are summarized in Figure 6. Once a set of mRNA data is transformed into a long-tailed histogram that can be fitted into PDE (Figure 7), two numbers can be obtained – (i) Planckian information of the second kind, I_{PS} (see Step 3), and (ii) Shannon entropy, H (see Step 4). These numbers can be represented as a point



Pathway #	Biological Process	Number of Open Reading Frames
1	Cell cycle	72
2	Cell wall biogenesis	53
3	Chromatic structure	44
4	Cytoskeleton	71
5	DNA repair	32
6	rRNA processing	37
7	Nuclear protein targeting	43
8	Protein synthesis	156
9	Transport	129
10	Transcription	175

Table 6: Biological process with number of open reading frames.

in the so-called Planck-Shannon plot (Figure 6 and Table 2) [38]. In short, the Planck-Shannon plot reduces a set of 40~50 numbers (each representing a mRNA level) or a given metabolic pathway to a point in the Planck-Shannon graph or space.

Several sets of about 10 metabolic pathways selected randomly were analyzed following the scheme in Figure 6, each pathway having 40 or more ORFs (Open Reading Frames). In most cases, the mRNA levels of each of these pathways produced a distinct long-tailed histogram whose shape fitted PDE thus generating three numbers corresponding to the three parameters of PDE, Eq. (8.3) in Figure 3.

The Planck-Shannon Space as the Semantic or Functional Space of Cell Language

As already indicated above, once a long-tailed histogram is fitted into PDE, two numbers can be calculated – (i) the Planckian information of the second kind, I_{ps} , Eq. (11), and (ii) the Shannon entropy (H) that can be calculated based on Eqs. (12) and (13):

$$H = -\sum p_i \log_2 p_i \quad (10)$$

where p_i is the probability of observing the i^{th} event or entity calculated as

$$p_i = \frac{y_i}{\sum y_i} \quad (11)$$

where y_i is the frequency of the i^{th} event of entity and the index I runs from 1 to n, the total number of events or entities. Thus, the information encoded in a long-tailed histogram can be visualized as a point in the Planck-Shannon space (Figure 7).

When a group of 10 metabolic pathways (each having a varying number of open reading frames as shown in the Figure 7) is chosen from the budding yeast transcriptome (measured over the 850 minute period of glucose-galactose shift experiments) [40,41], their mRNA levels were transformed into histograms, and the histograms fitted into PDE, 10 pairs of numbers can be generated, each pair corresponding to the Shannon entropy and Planckian information of the second kind as discussed above. When these 10 pairs of numbers are plotted on the Planckian-Shannon space, a reasonably good linear correlation was found (Figure 7(a)). However, when a similar set of 10 groups of mRNA levels are those that have no known metabolic functions, although each of them too generated a long-tailed histogram that fitted PDE, the resulting 10 pairs of numbers did not produce any correlation when plotted in the Planck-Shannon space (Figure 7(b) and Table 6).

Therefore, the linear correlation among the 10 points seems to occur only when they are functionally related. We tested this hypothesis with 8 other pairs of 10 groups of mRNA levels of known and unknown metabolic functions and found that 6 out of the 8 set of metabolic pathways showed linear correlations with correlation coefficients ranging from 0.60 to 0.75 while the 8 sets of mRNA levels with no known function showed no correlations, their correlation coefficients being less than 0.2. These observations support the following hypothesis:

“Each point on the Planck-Shannon space represents a metabolic pathway and a linear correlation among 3 or more such points represent what is here referred to as the ‘hypermetabolic pathway’ that may underlie a cell function” (12)

The ten points forming a correlated line indicates that the Planck-Shannon space can recognize the third level of metabolic organizations as predicted in the last row of Tables 3 and 4 and Row 9 in Figure 2. In other words, the Planck-Shannon space can distinguish molecular sentences (or metabolic pathways) as individual points, regardless of whether correlated with one another or not, and molecular texts as linearly correlated points three or more in number.

Applications of Ribnoscopy, the Cell Language Theory, and Planck-Shannon plots in Drug Discovery Research

Just as the study of electrons in atoms (i.e., atomic spectroscopy) revolutionized physics and information technology in the last century, so it may be predicted that the study of RNA molecules in living cells (called ‘ribnoscopy’, from ‘looking at RNA molecules’ using the microarray technique and its equivalent [3, Chapter 18 and 19) may revolutionize biology and medicine in the 21st century. Quantum mechanics that developed in physics between 1900 and 1925 provided the theoretical foundation for the study of electrons in atoms. Similarly, I suggest that the cell language theory whose beginning may be traced at least to Chargaff’s discovery of his parity rules in the middle of the 20th century [42-44] may provide the theoretical foundation for the study of RNA molecules in living cells.

The practical applications of the cell language theory, especially the concept of molecular text, as implemented by the PDE-based analysis of microarray data are illustrated in Figures 7(c) through f. If we define molecular texts as a functionally related set of 3 or more metabolic pathways within a patient (to be referred to as the intra-

organismic hypermetabolic pathways) or an identical pathway, e.g., the hypothetical protein pathway in Figure 7(c) through f, distributed among 3 or more patients (to be referred to as the inter-organismic hypermetabolic pathways), we can study the effect of drugs on the latter kind of molecular texts using the Planck-Shannon plots, just as we can study the effect of drugs on ligand-receptor interactions using the Scatchard plot in biochemistry and pharmacology.

Since doxorubicin treatment induces the correlation of the hypothetical protein pathway among 5 long survivors (Figures 7(c) and 7(d)), it seems logical to conclude that the activation of this metabolic pathway (i.e., inter-organismic hypermetabolic pathway) is beneficial for breast cancer tissues and hence the hypothetical protein pathway can serve as a biomarker for anti-breast cancer drug discovery. That is, whenever a drug candidate induces the activation of this particular metabolic pathway in long surviving breast cancer patients (or in their breast cancer cell cultures), that drug candidate can be identified as an anti-breast cancer drug.

If one examines other metabolic pathways in the human genome using the Planck-Shannon plots, it may be possible to discover breast-cancer biomarkers other than the hypothetical protein pathways. To discover potential anti-breast cancer drugs, it would be necessary to test them on the transcriptional profiles of the cultured cells biopsied from short and long survived breast cancer patients.

Conclusion

The microarray technique or its equivalent, when used in combination with Planckian Distribution Equation, will enable biomedical scientists to discover a novel class of metabolic structures here called “hypermetabolic pathways” that can serve as biomarkers for anti-cancer drug development without knowing detailed underlying molecular mechanisms.

Acknowledgements

I thank many students who performed the computational analysis reported here as part of the research in pharmacology elective course that I taught at the Ernest Mario School of Pharmacy during the academic years 2014-2017. Special thanks go to Vinay Vadali and Beum Jun Park.

References

1. Prigogine I (1977) Dissipative structures and biological order. *Adv Biol Med Phys* 16: 99-113.
2. Prigogine I (1980) From being to become: Time and complexity in physical sciences. W. H. Freeman and Company, San Francisco, USA.
3. Ji S (2012) Molecular theory of the living cell: Concepts, molecular mechanisms, and biomedical applications. Springer, New York, USA.
4. <http://www.conformon.net>
5. Ji S (2014) Mathematical models of RNA expression profiles: Potential applications to drug discovery research and personalized medicine. *J Bioequiv Availab* 6: 80-81.
6. https://en.wikipedia.org/wiki/The_dose_makes_the_poison
7. Ji S (1997) Isomorphism between cell and human languages: Molecular biological, bioinformatics and linguistic implications. *BioSystems* 44: 17-39.
8. Ji S (1999) The linguistics of DNA: Words, sentences, grammar, phonetics and semantics. *Ann NY Acad Sci* 870: 411-417.
9. Ji S (2018) The cell language theory: Connecting mind and matter. World Scientific Publishing, New Jersey, USA.
10. Ji S (2008) Modeling the single-molecule enzyme kinetics of cholesterol oxidase based on Planck's radiation formula and the principle of enthalpy-entropy compensation. In: Short Talk Abstracts, The 100th Statistical Mechanics Conference, December 13-16, Rutgers University, Piscataway, New Jersey, USA.
11. Ji S (2012) Molecular theory of the living cell: Concepts, molecular mechanisms, and biomedical applications. Springer: New York, USA.
12. Ji S (2018) The universality of the Planckian distribution equation. In: Cell language theory.
13. Norris V (1999) Hypothesis: Hyperstructures regulate bacterial structure and the cell cycle. *Biochimie* 81: 915-920.
14. Ji S (2012) The atom-cell isomorphism postulate. Sections 10.5 – 10.7.
15. Swanson CP (1964) The cell, (2nd edn), Prentice-Hall, Inc, Englewood Cliffs, New Jersey, USA.
16. 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.
17. 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.
18. Ji S (2002) The Bhopalator: An information/energy dual model of the living cell (II). *Fundamenta Informaticae* 49: 147-165.
19. Alberts B (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92: 291-294.
20. Ji S (1974) Energy and negentropy in enzymic catalysis. *Ann N Y Acad Sci* 227: 419-437.
21. Ji S (2000) Free energy and information contents of conformons in proteins and DNA. *BioSystems* 54: 107-130.
22. Ji S (2005) First, second and third articulations in molecular computing in the cell. In: Abstracts, 2005 World DNA and Genome Day, Dalian, China. pp: 25-29.
23. Stock AM, Robinson VL, Goudreau PN (2000) Two-component signal transduction. *Ann Rev Biochem* 69: 183-215.
24. Watters JW, Roberts CJ (2006) Developing gene expression signatures of pathway deregulation in tumors. *Molecular Cancer Therapeutics* 5: 2444-2449.
25. Ji S (1999) The cell as the smallest DNA-based molecular computer. *BioSystem* 52: 123-133.
26. Babloyantz A (1986) Molecules, dynamics and life: An introduction to self-organization of matter. Wiley-Interscience, New York, USA.
27. Kondepudi D, Prigogine I (1998) Modern thermodynamics: From heat engine to dissipative structures. John Wiley and Sons, Inc, Chichester.
28. Trifonov EN (1989) The multiple codes of nucleotide sequences. *Bull Math Biol* 51: 417-432.
29. Petoukhov SV (2016) The system-resonance approach in modeling genetic structures. *BioSystems* 139: 1-11.
30. Petoukhov SV (2017) The rules of long DNA-sequences and tetra-groups of oligonucleotides. Cornell University Library.
31. http://www.worldscientific.com/doi/abs/10.1142/9789814503655_0009
32. Ji S (2015) Planckian distributions in molecular machines, living cells, and brains: The wave-particle duality in biomedical sciences. In: Proceedings of the International Conference on Biology and Biomedical Engineering, Vienna.
33. <http://hyperphysics.phy-astr.gsu.edu/hbase/mod6.html>
34. <http://hyperphysics.phy-astr.gsu.edu/hbase/mod6.html/>
35. https://en.wikipedia.org/wiki/Harmonic_oscillator.
36. <http://hyperphysics.phy-astr.gsu.edu/hbase/waves/standw.html>
37. http://www.conformon.net/wp-content/uploads/2016/09/PDE_Vigier9.pdf.
38. Ji S (2018) RASER model of single-molecule enzyme catalysis and its application to the ribosome structure and function. *Arch Mol Med Genetics* 1: 31-39.
39. Culler J (1991) Ferdinand de Saussure, Revised edition. Cornell University Press, Ithaca.
40. Perou M, Surlie T, Eisen MB (2000) Molecular portraits of human breast tumours. *Nature* 406: 747-752.
41. Garcia-Martinez J, Aranda A, Perez-Ortin JE (2004) Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. *Mol Cell* 15: 303-313.

42. Chargaff E (1971) Preface to a grammar of biology: A hundred years of nucleic acid research. Science 172: 637-642.
43. Pauling L (1968) Orthomolecular psychiatry. Varying the concentrations of substances normally present in the human body may control mental disease. Science 160: 265-271.
44. <https://plato.stanford.edu/entries/peirce-semiotics/>