The GS (Genetic Selection) Principle

David L. Abel

TABLE OF CONTENTS

1. ABSTRACT:

The GS (Genetic Selection) Principle states that biological selection must occur at the nucleotide-sequencing molecular-genetic level of 3'5' phosphodiester bond formation. After-the-fact differential survival and reproduction of already-living phenotypic organisms (ordinary natural selection) does not explain polynucleotide prescription and coding. All life depends upon literal genetic algorithms. Even epigenetic and "genomic" factors such as regulation by DNA methylation, histone proteins and microRNAs are ultimately instructed by prior linear digital programming. Biological control requires selection of particular configurable switch-settings to achieve potential function. This occurs largely at the level of nucleotide selection, prior to the realization of any isolated or integrated biofunction. Each selection of a nucleotide corresponds to pushing a quaternary (four-way) switch knob in one of four possible directions. Formal logic gates must be set that will only later determine folding and binding function through minimum-free-energy sinks. These sinks are determined by the primary structure of both the protein itself and the independently prescribed sequencing of chaperones. Living organisms arise only from computational halting. Fittest living organisms cannot be favored until they are first computed. The GS Principle distinguishes selection *of existing* function (natural selection) from selection *for potential* function (formal selection at decision nodes, logic gates and configurable switch-settings).

2. INTRODUCTION

All known organisms are prescribed and largely controlled by information (1-22). Most biological prescriptive information presents as linear digital programming (23-26). Living organisms arise only from computational halting. Von Neumann, Turing and Wiener all got their computer design and engineering ideas from the linear digital genetic programming employed by life itself (27-32). All known life is cybernetic (33-35). Regulatory proteins, microRNAs and most epigenetic factors are digitally prescribed (3). MicroRNAs can serve as master regulators of gene expression (36-38). One microRNA can control multiple genes. One gene can be controlled by multiple microRNAs.

Nucleotides function as physical symbol vehicles in a material symbol system (39-41). Each selection of a nucleotide corresponds to pushing a quaternary (four-way) switch knob in one of four possible directions. The most perplexing problem for evolutionary biology is to provide a natural mechanism for setting functional

configurable switch-settings at the genetic level. These logic gates must be locked in open or closed positions with strong covalent bonds prior to folding of biopolymers. At the point of polymerization of informational positive single strands, no selectable three-dimensional shape exists for the environment to favor. In addition, the environment does not select for isolated function. The environment only selects for fittest already-living organisms.

The challenge of finding a natural mechanism for linear digital programming extends from primordial genetics into the much larger realm of semantics and semiotics in general. Says Barham: "The main challenge for information science is to naturalize the semantic content of information. This can only be achieved in the context of a naturalized teleology (by 'teleology' is meant the coherence and the coordination of the physical forces which constitute the living state)."(42) The alternative term "teleonomy" has been used to attribute to natural process "the appearance of teleology" (43-45). Either way, the bottom line of such phenomena is *selection* for higher function *at the logic gate programming level.*

3. WHERE SELECTION MUST OCCUR

Linear digital prescription requires selection of monomers *at the point of polymerization* of the initial positive informational strand. Primary structure (sequencing) instructs secondary and tertiary structure (three-dimensional shape). While chaperones and other factors affect folding, the sequencing of the polyamino acid itself is by far the biggest determinant of shape, electrostatic charge, grooves, knobs, tunnels, hydrophobicities, and lock-and-key binding of the globular protein molecular machine or enzyme to its substrate. Folding proceeds according to minimum Gibbs free-energy sinks (46-54). But rigidly-bounded monomeric sequencing largely determines what these thermodynamic and kinetic tendencies will be.

It is not sufficient for the environment to select the fittest living organisms. Organisms do not exist until after cooperative computational haltings occur on many different levels. None of these computational haltings will occur without selection of appropriate symbols so as to generate formally efficacious programming sequences. In addition, all these programming sequences must be integrated into a holistic operating system in order to organize even the simplest protometabolism. Organization, too, is mediated using multiple layers of material symbol systems. Physical symbol vehicles (nucleotide "tokens") are used to represent formal quaternary (four-way) switch-setting "choices." These in turn determine higher order levels of transcriptional regulatory networks, multilayer hierarchical structures, transcript turnover regulation, and three-dimensional information retention in genomes (55-57).

Self-replication tends to "get all the press" in life-origin literature. But the real issue of life origin lies in answering how the initial single positive strands of RNA instructions got sequenced so as to prescribe microRNA regulation, amino acid sequencing and eventual folding function. No new information is generated in base-

pairing replications. Base-pairing has nothing to do with the generation of genetic information or coding. Base-pairing is purely physicodynamic, and quite secondary to the already-programmed, formal, linear digital instructions of the single positive strand.

4. THE REQUIREMENTS OF SELECTION

Selection first requires categorization. A clear differentiation of real options must exist. Second, one category must be preferential or superior in its functionality compared to the others for selection to be worthwhile. Third, a means of selection has to exist. Last, impetus is needed to drive the selection process.

Does natural selection meet all of these criteria? The fittest organisms are clearly categorized from less fit organisms independent of human knowledge and description. Differential survival and reproduction were quite real ontologically before Homo sapiens ever appeared on the scene to ponder them. Superior fitness also readily meets the criterion of functional superiority. What about the third requirement: *means*? The means of selection is provided by gradual extinction of less successful competing organisms. Finally, the impetus to select is automatic. Differential survival and reproduction indirectly drives the selection process to its endpoint of maximum utility through time with no requirement of any external unnatural or supernatural force. No non-physical formal component is required for environmental selection of superior organisms to occur.

Problems arise, however, in explaining how any organism, fit or unfit, came into existence in the first place.

Let us now subject genetic programming at the molecular level to the same four essential criteria of selection referred to above. All four nucleotides polymerize with equal difficulty. Can single-stranded polynucleotide primary structures be distinguished and categorized by physicodynamics alone, prior to folding? Probably. Physicochemical, steric, electrostatic, and structural differences do exist between various stochastic ensembles of polynucleotides even before folding. But do any physicodynamic differences relating to nucleotide sequence *matter* to nature at the point of primary structure formation? In a prebiotic environment, would the environment prefer one stochastic ensemble over another? Although ribozymes and DNA enzymes exist, they contribute needed function only in a holistic metabolic context. No reason exists for nature to prefer a catalytic DNA over a non catalytic one. Nature has no goals, preferences or motives, evolution included. At the programming level of gene formation, function in an integrated metabolic scheme does not yet exist. No living phenotypic superiority exists for the environment to favor. Apart from a polynucleotide's participation in the instructional symbol system of an already living organism, any one single strand of RNA or DNA is just as good as any other. A self-replicating RNA could theoretically form spontaneously. But as we will discuss later, it is not at all clear what such a self-replicating RNA would contribute to any potential *metabolic* scheme. It would also consume so many resources few would be left for thousands of other needed metabolites to form out of a severly depleted sequence space. So *functional superiority*

of one sequence option (primary structure) over another is completely lacking in an abiotic environment. Natural selection fails the first two essential criteria of selection at the genetic programming level. The environment cannot adequately categorize options. In addition, functional superiority of some options over others does not yet exist for the environment to prefer.

What about the 3rd criterion of selection: *means*? With natural selection, the means is differential survival and reproduction of already living organisms. But at the genetic programming level in a prebiotic world, no life or differential survival exist yet. *Means* is totally lacking for evolution to occur at the programming level.

What about the fourth essential criterion of selection: *impetus*? In environmental selection, differential survival and reproduction of small populations drives the selection process. The impetus is automatic. But at the positive strand formation level, what is the natural-process impetus for selection of one nucleotide selection or one sequence over another? Phenotypic fitness does not yet exist. Life does not yet exist. Differential survivability cannot be a factor to drive the selection process. While a self-replicating polynucleotide might differentially "survive," a sequence optimized for self-replication would not have the ideal sequencing for almost any other metabolic function. The self-replicating strand would merely consume all the resources mass-producing itself. But with respect to prescribing all of the metabolic functions needed for life, the mass-produced strand would be gibberish.

Environmental selection can play no role whatever in the selection of nucleotides or codons. Yet these selections constitute the setting of critical logic gates. Nucleotide selections clearly constitute the programming of configurable switches. If the switches are not set properly, no life will come into existence to be favored. Yet at the point of polymerization of any certain sequence, no physicochemical superiority exists for the environment to favor.

So a purely physicalistic nucleotide polymerization of single positive strands in solution fails to manifest all four of the essential criteria of any selection process. No natural-process basis exists for programming the covalently-bound strand to be formally computational. Under these conditions, linear digital programming by environmental selection is impossible. Natural selection cannot occur at the programming level of configurable switch-setting (the choice of which nucleotide to polymerize next). Thus environmental selection cannot program computational linear digital programs. Yet environmental selection is the only kind of selection known to natural process.

Programming selections at successive decision nodes requires anticipation of what selections and what sequences *would be* functional. Selection must be *for potential* function. Nature cannot anticipate, let alone plan or pursue formal function. Natural selection can only preserve the fittest already-existing holistic life.

5. THE INSTANTIATION OF FORMALISM INTO PHYSICALITY

Non physical formalism can be instantiated into a material symbol system. Physical symbol vehicles (nucleotide tokens) can be used to record and transmit prescriptions of biochemical function. But both linear digital prescription and coding bijection (translation) are fundamentally formal, not physical. The field of applied genetic algorithms was modeled after the genetic control of living organisms. The applied model begins with a pool of potential formal solutions. This pool of potential formal solutions is presupposed, not explained materialistically. No such pool of potential formal solutions exists in the inanimate physical world.

Genetics is a representational Material Symbol System (MSS) (39-41, 58, 59). Instructions are instantiated into *the selection* of each of four possible alternative nucleotides, and into a particular sequence of those alternative nucleotides. From physicochemical standpoint, any of the four nucleotides can be polymerized next onto a single positive informational string with equal difficulty. Polymerization of the informational positive strand is therefore considered to be dynamically inert (physicochemically decoupled or incoherent) (41, 60). This detachment from physicodynamic determinism is essential for instantiation of prescription information into a physical matrix. Dynamic inertness is also necessary for random mutation and evolutionary transition (61-64). Without a freely resortable material symbol system, evolution is impossible (41).

As we saw above, environmental selection flunks all four essential criteria of selection *at the genetic level*. Natural selection is utterly blind to specific nucleotide needs at each locus in a forming single-stranded polynucleotide strand. Primordial soup models of life-origin are limited to either true stochastic ensembles, or to very low informational redundant sequences forced by law (e.g., adsorption onto clay surfaces). Of course purely physicodynamic base-pairing to already-existent highly informational strands can occur. But where did the already-existing highly-informational strands come from? If they are stochastic ensembles, by what mechanism did they acquire their linear digital prescription of eventual folding, lock-and-key binding, highly sophisticated formal biofunction, and holistic organization of metabolism? Coin flips cannot program linear digital cybernetics and symbol system coding. Random variation does not program sophisticated formal integration. No basis for formal selection of functional nucleotide sequencing exists in inanimate nature.

6. THE GS (GENETIC SELECTION) PRINCIPLE

The GS Principle states that *biological selection must occur at the nucleotide-sequencing molecular-genetic level of 3'5' phosphodiester bond formation.* After-the-fact differential survival and reproduction of already-living phenotypic organisms (ordinary natural selection) does not explain polynucleotide prescription, noise-reducing “block coding” using 3 to 1 symbol bijection, and life itself. Each nucleotide must be selected at

the point of polymerization with strong covalent bonds. Other non controlled constraints (e.g., environmental stresses), semi-controlled constraints (e.g., the corrupted information of prion misfoldings), and controlled constraints (e.g., programmed chaperone proteins) also affect protein and nucleic acid functional conformation. But by far the main determinant of conformational structure and function is the primary structure of polyamino acid chain itself. The determinant of amino acid sequencing is in turn post-edited ribonucleotide and polycodon sequencing.

The specific selection of one ribonucleotide from among four real options functions as a quaternary (four-way) configurable switch-setting. Configurable switches *control and prescribe*, not merely describe, translated metabolic utility. Covalently-bound selection commitments are rigid by comparison to weaker H-bonded secondary folding. Primary structure is the primary determinant of what H-bonding and van der Waals forces can accomplish. Configurable switch-settings in the form of specific nucleotide selections constrain minimum-free-energy folding space. Both ribonucleotide polymerization reactions and folding are subject to the laws of motion and to dynamic constraints. But the cybernetic function of the genetic material symbol system *controls* those constraints. The sequencing of nucleotides as physical symbol vehicles is not determined by physicodynamics. It is *dynamically inert* (physicodynamically incoherent; decoupled from physical determinism) (41, 60). Once instantiated into a Material Sign System, however, this dynamically-inert programming physically constrains dynamic folding space. The GS Principle attributes the main *control* of folding constraints to *nucleotide sequencing selections.*

All life depends upon literal genetic algorithms, including most epigenetic factors. Fittest organisms cannot be favored until they are first computed. The GS Principle elucidates the source of messenger molecules' representational prescription of biofunction, a phenomenon unique in nature to life (26). The GS Principle distinguishes selection *of existing* function (natural selection) from selection *for potential* function (formal selection at decision nodes, logic gates and configurable switch-settings).

Symbol systems employ alphabetical characters, signs, and physical symbol vehicles (tokens such as nucleotide options) to represent meaning or function. Selections must be made from an option space of real, uncoerced alternatives. Each selection represents the setting of a logic gate or configurable switch that computes or integrates a circuit. The purposeful selection of letters alone constructs words, sentences and paragraphs. Any denial of "choice with intent" reduces language to gibberish. Any attempt to replace arbitrary rules of convention with fixed law destroys information potential. Rules are voluntarily followed, not forced. Only when the destination voluntarily applies the same free and arbitrary rules of the source can the destination successfully interpret the source's intended meaning. By arbitrary, we do not mean random. We mean, "could have been otherwise" within the constraints of nature. But abiding by arbitrary rules also embodies selection in accordance with those rules at the decision-node level. The programming of polynucleotide prescription of biofunction cannot be reduced to mere linguistic and computer science metaphor (20). Indeed, if any

analogy exists, it is in the reverse direction. Life's linear digital cybernetics predates humans, their languages and their computers.

7. THE CAPABILITIES OF NATURAL SELECTION

Only existing genetic algorithms can be optimized. Prior to an algorithm having computational function, no basis exists for selection in nature. So the question becomes, "How did *any* computational program arise in nature? Computation is formal, not physical. Natural selection cannot generate formalisms. It can only prefer *the results* of formal computations—already living organisms. What would be the basis of natural selection for a half-written program that does not yet compute? Even if a formal computational program were to somehow spontaneously arise, why would an inanimate environment value and preserve it? The only basis for natural selection from the start was survival of the fittest already-living organisms. But no organism exists without hundreds of cooperating formal algorithms all organized into one holistic scheme. The more computational steps that are required to achieve integrated halting, the harder it becomes for an inanimate environment to explain optimization. And the more algorithms that must be simultaneously optimized and integrated, the harder it is to explain metabolism.

Natural selection resembles public consumption of the best available software. The programming details and methodology of production are of no interest to the purchasers of software. Pre-programmed, bug-free, superior utility is the only criterion of public selection. The consumer plays no role whatever in the writing or refinement of the program's computational efficiency. The finished product with the best reputation, availability, and lowest cost becomes "the fittest species." Just as consumers are oblivious to how the best software was produced, natural selection is oblivious to how the fittest species was produced. Natural selection offers no explanation whatever for programming at the genetic level. Similarly, natural selection does not explain the derivation of the many cooperative computational processes leading up to the origin of life.

Stunningly, information has been shown *not* to increase in the coding regions of DNA with evolution. Mutations do not produce increased information. Mira et al (65) showed that the amount of coding in DNA actually decreases with evolution of bacterial genomes, not increases. This paper parallels Petrov's papers starting with (66) showing a net DNA loss with Drosophila evolution (67). Konopka (68) found strong evidence against the contention of Subba Rao et al (69, 70) that information increases with mutations. The information content of the coding regions in DNA does not tend to increase with evolution as hypothesized. Konopka also found Shannon complexity not to be a suitable indicator of evolutionary progress over a wide range of evolving genes. Konopka's work applies Shannon theory to known functional text.
Kok et al. (71) also found that information does not increase in DNA with evolution. As with Konopka, this finding is in the context of the change in mere Shannon uncertainty. The latter is a far more forgiving definition of information than that required for

prescriptive information (PI) (21, 22, 33, 72). It is all the more significant that mutations do not program increased PI. Prescriptive information either instructs or directly produces formal function. No increase in Shannon or Prescriptive information occurs in duplication. What the above papers show is that not even *variation* of the duplication produces new information, not even Shannon "information."

All of the above work correlates well with Weiss et al (73) finding only 1% deviation from randomness in coding regions. One cannot increase "information" (really "uncertainty") very much when starting from only 1% deviation from randomness in the coding regions. Only 1% deviation from randomness is already nearly maxed out in uncertainty. How did a text that deviates only slightly from seeming randomness get *so* instructional and biofunctional? Clearly, mere combinatorial uncertainty is not going to explain the phenomenon of cybernetic genetic prescription.
No empirical evidence exists of mere variation ever having generated sophisticated PI, computational halting, or cybernetic integration of large numbers of pathways and cycles, or the achievement of metabolic goals.

8. THE LIMITS OF PHYSICS AND CHEMISTRY

No physicochemical factors determine monomeric sequencing. Physicodynamic determinism would severely reduce the uncertainty required for information retention in any physical matrix. Sequencing is physicodynamically inert (dynamically incoherent) (41, 60). Sequencing is decoupled from physicodynamic causation. It is independent of cause-and-effect physical determinism. This freedom is the very key to carbon chemistry being ideal for instantiation of large amounts of genetic prescriptive information into a physical matrix.

Attempts to explain the origin of formal programming from chance and necessity, thermodynamics and kinetics (74-84), have been unconvincing. Brillouin and others have attempted to equate Shannon's "informational" uncertainty with Maxwell-Boltzmann-Gibbs entropy. Notions of negentropy abound despite Boltzmann's prohibition of a negative constant in his famous equation: $S = k \log W$ (85). Neither the number of microstates (W) nor Boltzmann's constant ($k = 1.38065 \times 10^{-23}$ joule/Kelvin) can be negative. In addition, every probability distribution is unique. Yockey showed that the probability distribution phase space of Boltzmann's physical entropy (S) cannot be equated or synthesized with Shannon's probability distribution of "informational" uncertainty (H) despite seemingly identical S and H equations (apart from the disallowed sign reversal of S) (23).

Schneider (86) and Adami (87, 88) are correct that uncertainty is not information. But mere subtractions of "after uncertainty" from "before uncertainty" do not measure up to the formal cybernetic proficiency of genetic control. Even archaeal genomes positively program thousands of integrated computations. These cybernetic processes are not just decreasing measurements of combinatorial probabilism. Even if they were, inanimate nature cannot measure (89-92). Human knowledge and measurement of the

change in uncertainty inserts a human mental factor into the definition of even Shannon "information" that did not exist for most of the presumed 3.8 billion years of genetic prescription. Genetic instructions stand alone in their proficiency at making life happen.

Adami is also correct that bona fide information must be *about* something (87, 88). But no source of formal "aboutness" has ever been provided from physics, chemistry and the physical environment alone. Yes, the environment provides a context. But the environment does not program the meaningful (cybernetically functional) configurable switch-settings that enable living organisms to overcome environmental insults and challenges that will occur only in the future. Aboutness will not be found in the environment itself. Aboutness requires choice contingency. Aboutness is generated by the particular settings of configurable switches that program computational halting and that organize integrated circuits to meet future challenges. Meaning is formal, not physical. The environment cannot exercise choice contingency at genetic logic gates. The environment cannot program configurable switches. It can only favor the best already-computed living organisms.

A long string of Nobel laureates including Niels Bohr (93) and Jacques Monod (94) have argued that chance and necessity cannot generate non trivial linear digital biological instructions. Bohr argued that "Life is consistent with, but undecidable from physics and chemistry." (93). Many additional first-rate biologists such as Ernst Mayr (95, 96) and Bernd-Olaf Küppers (97, pg 166) have argued that physics and chemistry do not explain life. Says Hubert Yockey, "More than any other characteristic, computational linear digital algorithms distinguish life from non life" (98).

Clay surface adsorption is also purely physicodynamic with highly redundant order. Nucleotide adsorption results in homopolymers such as poly[A] and poly[G] onto montmorillonite and kaolinite clay surfaces (99-102). Such homopolymerizations do not even contain Shannon information, let alone semantic information. Such homopolymers offer no potential for prescription of formal function. Inflexible laws cannot select nucleotides and codon "block codes" (many-to-one symbol assignments used to reduce noise pollution in the Shannon channel) to achieve formal programming function. Life is the most highly informational phenomenon known to humans. Physicochemical propensities only preclude the generation of new information. New information generation requires uncertainty, not forced physical causation according to fixed laws. Laws are compression algorithms of reams of data made possible because the information content of all that data is so reducible. Functional sequencing, therefore, will never be explained by law. This includes hoped-for, as-of-yet undiscovered imaginary laws. Laws are too low-informational. They describe redundant, monotonous, highly ordered behavior with a probability approaching 1.0. Such behavior has almost no uncertainty. Law-like behavior precludes highly informational uncertainty required for linear digital prescription of formal function.

Another major problem for life-origin science is that untemplated single-stranded oligoribonucleotides do not exceed 8 to 10 mers in length in aqueous solution. Biological prescription requires extremely long positive prescriptive strands. Base-pairing or

templating of some kind (as on montmorillonite) is required to expedite polymerization. Base-pairing, of course, merely duplicates the strand's sequence in reverse direction. Templating on clay typically produces polyadenosines of no more than 50 mers (100). To achieve that length requires the intervention of highly intelligent chemists. Such highly ordered, short sequences contain little or no information. The high probability of adenosine's occurrence at each "decision node" in clay adsorption strings approaches 1.0. The uncertainty of such an RNA strand is close to the summation of 0 bits at each locus in the string times the number of ribonucleotides ($- \log_2 p = - \log_2 1.0 = O$ bits of uncertainty times n loci = 0 additive bits). Physicodynamics cannot generate nontrivial Prescriptive Information (PI). The generation of programming instructions requires freedom of selection at each logic gate. More importantly, it requires selection *for potential function* at each decision node. This is the very reason that sequencing is able to become a control function, not just a physical constraint. And the sequencing is not stochastic either. The control function manifests too much computational utility. No empirical evidence, predictive success, or rationality exists in the history of human experience to justify believing that unaided Markov processes can generate sophisticated algorithmic programming apart from hidden investigator involvement. Although the experiment is begun with a random pool, selections of particular iterations are usually employed behind the scenes for the desired potential function.

The pre RNA and RNA World models of life origin provide simplification and ideal reductionism for the study of the birth of biocybernetics and biosemiotics. Small RNA's provide a multitude of controlling (regulatory) functions even in current life. But spontaneous RNA generation is a biochemical vertical cliff (103-107). In addition, ribose and RNA are too unstable for the long highly informational strands needed for life to have slowly developed by small increments (104, 108). As a result, many life-origin specialists have been forced to return to Peptide First and Metabolism First models advocated by Gánti (109), Shapiro (110, 111), Dyson (112), Kauffman(113), Wachtershauser (114), Morowitz (115), Deamer (116), Lindhal (117), Russell (118), and many others.

Yet a Metabolism First origin of life is far from a foregone conclusion (119, 120). Few life-origin scientists have been more respected than Leslie Orgel. Wrote Orgel, "In my opinion, there is no basis in known chemistry for the belief that long sequences of reactions can organize spontaneously---and every reason to believe that they cannot." (121) Indeed, organization should never be confused with mere self-ordering phenomena in nature (22, 33, 35).

9. THE GENETIC CODE IS CONCEPTUALLY IDEAL

The source of genetic programming lies in the selections of nucleotides, and in the sequencing of those particular nucleotide selections. Says Fontana and Schuster, "Understanding which phenotypes are accessible from which genotypes is fundamental for understanding the evolutionary process." (50) The sequencing of DNA nucleotides

has no meaning or function independent of an overarching formal system of arbitrary (could have been otherwise) symbol assignments to each amino acid.

A representational symbol system is clearly employed in the triplet codon table of amino acid prescription. Codons are a form of Hamming "block code" wherein consistent groups of three symbols are used to represent each single amino acid prescription. Block coding is a form of redundancy coding used to reduce noise pollution in the transmission channel. These arbitrary assignments have been shown to be conceptually ideal (122) (123). Despite wobbles and point mutations, codons are often still able to prescribe the correct amino acid because of this extraordinary redundancy coding.

Life-origin models cannot reduce these phenomena to human epistemology. They are objective phenomena, not merely heuristic tools of human mental construction. Biosemiosis and biocybernetic management was integrating and engineering life's processes long before *Homo sapiens* appeared on the scene to ascribe their linguistic and cybernetic analogies to molecular biology. How would chance and necessity have conceived such an effective, formal, noise-reducing scheme?

Additional layers of coding sophistication also exist. Independent coding overlays the genetic code in DNA (124). A separate set of rules controls the binding of transcription factors and histone proteins to DNA. These additional rules control messenger RNA splicing and folding. The later contribute to regulating protein manufacture. The two coding systems are independent, but they are also coordinated. The two codes jointly control metabolism (124). The genomic code is far more vast than the genetic code, as if we weren't already burdened trying to explain the genetic code alone through natural process. The genomic code includes the three-dimensional structure of DNA and many additional overlaid codings in molecular biology (123).

All of these formally integrated systems require selection contingency, not chance contingency or fixed law, to organize (125). Selection must take place at the genetic level of nucleotide selection for any phenotype to come into existence, let alone the fittest phenotype. This fact of reality constitutes the GS Principle.

10. PERSPECTIVE

The GS Principle states that selection must occur at the molecular/genetic level, not just at the fittest phenotypic/organismic level, to explain the generation of polynucleotide and polycodon linear digital prescription. Organismic/phenotypic selection (natural selection) cannot prescribe the linear digital programming of coded genetic instructions. Environmental selection cannot set configurable switches so as to achieve potential integrated circuits. Selection pressure is after the fact of computational halting. No fittest organisms exist for the environment to favor without prior computational haltings on many cooperative levels. Nucleotides must be selected and rigidly bound in a certain sequence to prescribe and integrate metabolism. Sequencing

(primary structure) is the major determinant of three-dimensional molecular-machine shape (tertiary structure) and biofunction. Nucleotide sequencing is covalently (rigidly) bound into a linear digital string long before that string can fold into a ribozyme or can digitally prescribe polyamino acid sequencing. Metabolism depends on holistic integration of thousands of individual protein prescriptions, including epigenetic regulatory proteins and microRNAs. Ultimately even most epigenetic factors such as methylations are prescribed and tightly controlled by liner digital genetic instructions. Such symbol systems are fundamentally formal, not physical.

Ribonucleotides and oligoribonucleotides are physical. Their polymerization reactions are fully subject to the laws of motion and to dynamic constraints. But their specific sequencing is dynamically inert. Their cybernetic function consists of instantiations of formal selections for potential function, not already-existing physicodynamic necessity. Selection pressure plays no role in determining *potential* function at the formal programming level of configurable switch-setting (polymerization of each particular nucleotide).

Genes are linear digital programs. Even their editing is ultimately controlled by other linear digital programs. Genes only function exists in the context of a formal representational material symbol system using physical symbol vehicles. Their configurable switch-settings are *dynamically inert.* Highly informational metabolic instructions cannot be generated by low-informational laws. The GS Principle defines the kind of selection that is required to set physical configurable switches so as to compute metabolic integration. Environmental selection has never been observed to generate the simplest example of formal computational halting. Worse yet, the latter is a logical impossibility. Physicodynamics is limited to chance and necessity. Formal computation is abstract, conceptual and non physical. Formal computation cannot be generated apart from selection contingency at true decision nodes. Metabolic organization and life have never been observed to exist independent of formally integrated computational haltings.

The specific selection of each nucleotide from among four real options *controls,* not merely constrains biofunction. The selection of each nucleotide also *prescribes*, not merely describes, its metabolic contribution. The GS Principle states that natural selection of already optimized genetic algorithms (the fittest already-computed living phenotypes) is inadequate to explain the derivation of a single-stranded polynucleotide's digital programming and computational prowess. Selection for function must occur at each decision node—each logic gate—each configurable switch-setting—each nucleotide polymerization onto the programming string.

11. REFERENCES

1. Stegmann, U. E.: Genetic Information as Instructional Content. *Phil of Sci,* 72, 425-443(2005)
2. Jacob, Francois: The Logic of Living Systems---a History of Heredity. Allen Lane, London (1974)
3. Alberts, B. & et al: Molecular Biology of the Cell. Garland Science, New York (2002)
4. Davidson, E. H., J. P. Rast, P. Oliveri, A. Ransick, C. Calestani, C. H. Yuh, T. Minokawa, G. Amore, V. Hinman, C. Arenas-Mena, O. Otim, C. T. Brown, C. B. Livi, P. Y. Lee, R. Revilla, A. G. Rust, Z. Pan, M. J. Schilstra, P. J. Clarke, M. I. Arnone, L. Rowen, R. A. Cameron, D. R. McClay, L. Hood & H. Bolouri: A genomic regulatory network for development. *Science,* 295, 1669-78(2002)
5. Dose, K.: On the origin of biological information. *Journal of Biological Physics,* 20, 181-192(1994)
6. Griffiths, P. E. & K. Sterelny: Sex and Death: An Introduction to Philosophy of Biology. Univ. of Chicago Press, Chicago (1999)
7. Wolpert, L.: Principles of Development. Oxford University Press, Oxford (2002)
8. Sterelny, K., K. Smith & M. Dickison: The extended replicator. *Biology and Philosophy,* 11, 377-403(1996)
9. Hoffmeyer, J. & C. Emmeche: "Code-Duality and the Semiotics of Nature," (Forward to and reprinting of, with new footnotes). *Journal of Biosemiotics,* 1, 37-91(2005)
10. Barbieri, M.: Biology with information and meaning. *History & Philosophy of the Life Sciences*, 25, 243-254(2004)
11. Barbieri, M.: Introduction to Biosemiotics: The New Biological Synthesis. Springer-Verlag New York, Inc (2006).
12. Barbieri, M.: Is the cell a semiotics system? In: Introduction to Biosemiotics: The New Biological Synthesis. Ed: M. Barbieri. Springer, Dordrecht, The Netherlands (2007)
13. Barbieri, M.: BioSemiotic Research Trends Nova Science Publishers, Inc.(2007).
14. Barbieri, M.: The Codes of Life: The Rules of Macroevolution (Biosemiotics). Springer(2007).
15. Hoffmeyer, J.: The swarming cyberspace of the body. *Cybernetics and Human Knowing*, 3, 1-10(1995)
16. Hoffmeyer, J.: Biosemiotics: Towards a new synthesis in biology. *European Journal for Semiotic Studies,* 9, 355-376(1997)
17. Hoffmeyer, J.: Code-duality and the epistemic cut. *Ann N Y Acad Sci*, 901, 175-86(2000)
18. Hoffmeyer, J.: Semiotic scaffolding of living systems. In: Introduction to Biosemiotics: The New Biological Synthesis. Ed: M. Barbieri. Springer-Verlag New York, Inc. , Dordrecht, The Netherlands; Secaucus, NJ, USA (2006)
19. Abel, D. L.: Is Life Reducible to Complexity? In: Fundamentals of Life. Eds: G. Palyi, C. Zucchi&L. Caglioti. Elsevier, Paris (2002)
20. Abel, D. L. & J. T. Trevors: More than metaphor: Genomes are objective sign systems. *Journal of BioSemiotics,* 1, 253-267(2006)

21. Abel, D. L. & J. T. Trevors: Three subsets of sequence complexity and their relevance to biopolymeric information. *Theoretical Biology and Medical Modeling*, 2, Open access at http://www.tbiomed.com/content/2/1/29(2005)
22. Abel, D. L.: The BioSemiosis of Prescriptive Information. *Semiotica In Press*(2008)
23. Yockey, H. P.: Information Theory and Molecular Biology. Cambridge University Press, Cambridge (1992)
24. Yockey, H. P.: Information in bits and bytes. *BioEssays*, 17, 85-88(1995)
25. Yockey, H. P.: Information theory, evolution and the origin of life. *Information Sciences*, 141, 219-225(2002)
26. Yockey, H. P.: Information Theory, Evolution, and the Origin of Life. Cambridge University Press, Cambridge (2005)
27. Quastler, H.: A Primer on Information Theory in Symposium on Information Theory in Biology. *The Gatlinburg Symposium*, H. P. Yockey, R. P. Platzman&H. Quastler, Eds., Pergamon Press, pages 37 and 360 (1958).
28. von Neumann, J. & A. W. Burks: Theory of Self-Reproducing Automata. University of Illinois Press, Urbana, (1966)
29. Turing, A. M.: On computable numbers, with an application to the *entscheidungs problem*. *Proc. Roy. Soc. London Mathematical Society*, 42, 230-265 [correction in 43, 544-546](1936)
30. von Neumann, J.: Collected works. Pergamon Press, New York (1961)
31. Wiener, N.: Cybernetics. Wiley, New York (1948)
32. Wiener, N.: Cybernetics, its Control and Communication in the Animal and the Machine. New York (1948)
33. Abel, D. L. & J. T. Trevors: Self-Organization vs. Self-Ordering events in life-origin models. *Physics of Life Reviews*, 3, 211-228(2006)
34. Abel, D. L.: Life origin: the role of Complexity at the Edge of Chaos. *Washington Science 2006*, J. Chandler&P. Kay, Eds.(2006).
35. Abel, D. L.: Complexity, Self-organization, and Emergence at the Edge of Chaos in Life-Origin Models. *Journal of the Washington Academy of Sciences*, 93, 1-20(2007)
36. Ma, L. & R. A. Weinberg: MicroRNAs in malignant progression. *Cell Cycle*, 7, (2007)
37. Morin, R. D., M. D. O'Connor, M. Griffith, F. Kuchenbauer, A. Delaney, A. L. Prabhu, Y. Zhao, H. McDonald, T. Zeng, M. Hirst, C. J. Eaves & M. A. Marra: Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res*(2008)
38. Royo, H. & J. Cavaille: Non-coding RNAs in imprinted gene clusters. *Biol Cell*, 100, 149-66(2008)
39. Rocha, L. M.: Evidence Sets and Contextual Genetic Algorithms: Exploring uncertainty, context, and embodiment in cognitive and biological systems. . *Systems Science*, *http://informatics.indiana.edu/rocha/dissert.html (last accessed 9/6/05)*, State University of New York(1997).
40. Rocha, L. M.: Selected self-organization and the semiotics of evolutionary systems. In: Evolutionary Systems: Biological and Epistemological Perspectives on Selection and Self-Organization. Eds: S. Salthe, G. van de Vijver&M. Delpos. Kluwer, The Netherlands (1998)
41. Rocha, L. M.: Evolution with material symbol systems. *Biosystems*, 60, 95-121(2001)

42. Barham, J.: A dynamical model of the meaning of information. *Biosystems*, 38, 235-41(1996)
43. Pross, A.: On the chemical nature and origin of teleonomy. *Origins of life and evolution of the biosphere*, 35, 383-94(2005)
44. Lifson, S.: Chemical selection, diversity, teleonomy and the second law of thermodynamics. Reflections on Eigen's theory of self-organization of matter. *Biophys Chem*, 26, 303-11.(1987)
45. Pittendrigh, C. S.: Teleonomy. In: Behavior and Evolution. Eds: A. Roe&G. G. Simpson. Yale University Press, New Haven, CN (1958)
46. Flamm, C., W. Fontana, I. L. Hofacker & P. Schuster: RNA folding at elementary step resolution. *Rna*, 6, 325-38(2000)
47. Woodson, S. A.: Folding mechanisms of group I ribozymes: role of stability and contact order. *Biochem Soc Trans*, 30, 1166-9(2001)
48. Martin, A. & F. X. Schmid: The Folding Mechanism of a Two-domain Protein: Folding Kinetics and Domain Docking of the Gene-3 Protein of Phage fd. *J Mol Biol*, 329, 599-610(2003)
49. Munoz, V.: Conformational Dynamics and Ensembles in Protein Folding. *Annual Review of Biophysics and Biomolecular Structure*, 36, 395-412(2007)
50. Fontana, W. & P. Schuster: Shaping space: the possible and the attainable in RNA genotype-phenotype mapping. *J Theor Biol*, 194, 491-515(1998)
51. Schuster, P., P. F. Stadler & A. Renner: RNA structures and folding: from conventional to new issues in structure predictions. *Curr Opin Struct Biol*, 7, 229-35(1997)
52. Schuster, P. & P. F. Stadler: Landscapes: complex optimization problems and biopolymer structures. *Comput Chem*, 18, 295-324(1994)
53. Schuster, P., W. Fontana, P. F. Stadler & I. L. Hofacker: From sequences to shapes and back: a case study in RNA secondary structures. *Proc R Soc Lond B Biol Sci*, 255, 279-84(1994)
54. Martinez, H. M.: An RNA folding rule. *Nucleic Acids Res*, 12, 323-34(1984)
55. Cao, G. S., A. L. Liu & N. Li: [Exploration of the hidden layers of genome.]. *Yi Chuan*, 26, 714-20(2004)
56. Ma, H. W., B. Kumar, U. Ditges, F. Gunzer, J. Buer & A. P. Zeng: An extended transcriptional regulatory network of Escherichia coli and analysis of its hierarchical structure and network motifs. *Nucleic Acids Res*, 32, 6643-9(2004)
57. Mata, J., S. Marguerat & J. Bahler: Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem Sci*, 30, 506-14(2005)
58. Rocha, L. M.: Syntactic autonomy: or why there is no autonomy without symbols and how self-organizing systems might evolve them. In: Closure: Emergent Organizations and Their Dynamics. Annals of the New York Academy of Sciences. Eds: J. L. R. Chandler & G. van deVijver. Vol 901, 207-223 (2000)
59. Rocha, L. M. & W. Hordijk: Representations and emergent symbol Systems. *Cognitive Science*(2000)
60. Rocha, L. M. & W. Hordijk: Material representations: from the genetic code to the evolution of cellular automata. *Artif Life*, 11, 189-214(2005)
61. Pattee, H. H.: Clues from molecular symbol systems. In: Dahlem Workshop Reports: Life Sciences Research Report 19: Signed and Spoken Language: Biological Constraints

on Linguistic Form. Eds: U. Bellugi&M. Studdert-Kennedy. Verlag Chemie GmbH, Dahlem Konferenzen, Weinheim (1980)
62. Pattee, H. H.: Cell psychology: an evolutionary approach to the symbol-matter problem. *Cognition and Brain Theory*, 5, 325-341(1982)
63. Turvey, M. T. & P. N. Kugler: A comment on equating information with symbol strings. *Am J Physiol*, 246, R925-7(1984)
64. Pattee, H. H.: Universal principles of measurement and language functions in evolving systems. In: Complexity, Language, and Life: Mathematical Approaches. Eds: J. L. Casti & A. Karlqvist. Springer-Verlag, Berlin (1986)
65. Mira, A., H. Ochman & N. A. Moran: Deletional bias and the evolution of bacterial genomes. *Trends Genet*, 17, 589-96(2001)
66. Petrov, D. A., Y. C. Chao, E. C. Stephenson & D. L. Hartl: Pseudogene evolution in Drosophila suggests a high rate of DNA loss. *Mol Biol Evol*, 15, 1562-7(1998)
67. Petrov, D. A. & D. L. Hartl: High rate of DNA loss in the Drosophila melanogaster and Drosophila virilis species groups. *Mol Biol Evol*, 15, 293-302(1998)
68. Konopka, A. K.: Is the information content of DNA evolutionarily significant? *J Theor Biol*, 107, 697-704.(1984)
69. Subba Rao, G., Z. Hamid & J. Subba Rao: The information content of DNA and evolution *J. Theoretical Biology*, 81, 803(1979)
70. Subba Rao, J., C. P. Geevan & G. Subba Rao: Significance of the information content of DNA in mutations and evolution. *J Theor Biol*, 96, 571-7(1982)
71. Kok, R. A., J. A. Taylor & W. L. Bradley: A statistical examination of self-ordering of amino acids in proteins. *Origins of life and evolution of the biosphere*, 18, 135-42(1988)
72. Durston, K. K., D. K. Chiu, D. L. Abel & J. T. Trevors: Measuring the functional sequence complexity of proteins. *Theor Biol Med Model*, 4, 47(2007)
73. Weiss, O., M. A. Jimenez-Montano & H. Herzel: Information content of protein sequences. *J Theor Biol*, 206, 379-86(2000)
74. Rojdestvenski, I. & M. G. Cottam: Mapping of statistical physics to information theory with application to biological systems. *J Theor Biol*, 202, 43-54.(2000)
75. Bennett, C. H.: The thermodyamics of computation - a review. *Int. J. Theor. Phys.*, 21, 905-940(1982)
76. Schneider, T. D.: Evolution of biological information. *Nucleic Acids Res*, 28, 2794-9(2000)
77. Wallace, R. & R. G. Wallace: Information theory, scaling laws and the thermodynamics of evolution. *J Theor Biol*, 192, 545-559.(1998)
78. Weinberger, E. D.: A theory of pragmatic information and its application to the quasi-species model of biological evolution. *Biosystems*, 66, 105-19(2002)
79. Yokomizo, H., J. Yamashita & Y. Iwasa: Optimal Conservation Effort for a Population in a Stochastic Environment. *J Theor Biol*, 220, 215-231(2003)
80. Klir & Wierman: Uncertainty-Based Information. Physica-Verlag, (1998)
81. Ricard, J.: What do we mean by biological complexity? *C R Biol.*, Feb;326, 133-40(2003)
82. Strippoli, P., S. Canaider, F. Noferini, P. D'Addabbo, L. Vitale, F. Facchin, L. Lenzi, R. Casadei, P. Carinci, M. Zannotti & F. Frabetti: Uncertainty principle of genetic information in a living cell. *Theor Biol Med Model*, 2, 40(2005)

83. Balanovski, E. & P. Beaconsfield: Order and disorder in biophysical systems: a study of the correlation between structure and function of DNA. *J Theor Biol*, 114, 21-33(1985)
84. Theise, N. D. & R. Harris: Postmodern biology: (adult) (stem) cells are plastic, stochastic, complex, and uncertain. *Handbook of experimental pharmacology*389-408(2006)
85. Boltzmann, L.: Weitere Studien über das Wärmegleichgewicht unter Gasmolekulen. . *Königliche Academie der Wisschenshaft (Wien), Sitzungsberichte II Abteilung*, 66, 275(1877)
86. Schneider, T. D.: Information Is Not Entropy, Information Is Not Uncertainty. (2000).
87. Adami, C.: Introduction to Artificial Life. Springer/Telos, New York (1998)
88. Adami, C. & N. J. Cerf: Physical complexity of symbolic sequences. *Physica D*, 137, 62-69(2000)
89. Pattee, H. H.: The measurement problem in artificial world models. *Biosystems*, 23, 281-9; discussion 290(1989)
90. Pattee, H. H.: Evolving Self-Reference: Matter, Symbols, and Semantic Closure. *Communication and Cognition-Artificial Intelligence*, 12, 9-28(1995)
91. Pattee, H. H.: Artificial Life Needs a Real Epistemology. In: Advances in Artificial Life Ed: F. Moran. Springer, Berlin (1995)
92. Pattee, H. H.: The physics of symbols: bridging the epistemic cut. *Biosystems*, 60, 5-21(2001)
93. Bohr, N.: Light and life. *Nature*, 131, 421(1933)
94. Monod, J.: Chance and Necessity. Knopf, New York (1972)
95. Mayr, E.: Introduction, pp 1-7; Is biology an autonomous science? pp 8-23. In: Toward a New Philosophy of Biology, Part 1. Ed: E. Mayr. Harvard University Press, Cambridge, MA (1988)
96. Mayr, E.: The place of biology in the sciences and its conceptional structure. In: The Growth of Biological Thought: Diversity, Evolution, and Inheritance Ed: E. Mayr. Harvard University Press, Cambridge, MA (1982)
97. Küppers, B.-O.: Information and the Origin of Life. MIT Press, Cambridge, MA (1990)
98. Yockey, H. P.: Origin of life on earth and Shannon's theory of communication. *Comput Chem*, 24, 105-123.(2000)
99. Franchi, M. & E. Gallori: A surface-mediated origin of the RNA world: biogenic activities of clay-adsorbed RNA molecules. *Gene*, 346, 205-14(2005)
100. Ferris, J. P.: Montmorillonite catalysis of 30-50 mer oligonucleotides: laboratory demonstration of potential steps in the origin of the RNA world. *Origins of life and evolution of the biosphere*, 32, 311-32(2002)
101. Sowerby, S. J., C. A. Cohn, W. M. Heckl & N. G. Holm: Differential adsorption of nucleic acid bases: Relevance to the origin of life. *Proc Natl Acad Sci U S A*, 98, 820-2.(2001)
102. Ertem, G., R. M. Hazen & J. P. Dworkin: Sequence Analysis of Trimer Isomers Formed by Montmorillonite Catalysis in the Reaction of Binary Monomer Mixtures. *Astrobiology*, 7, 715-722(2007)
103. Shapiro, R.: The improbability of prebiotic nucleic acid synthesis. *Orig Life*, 14, 565-570(1984)

104. Shapiro, R.: Prebiotic ribose synthesis: a critical analysis. *Orig Life Evol Biosph*, 18, 71-85(1988)
105. Shapiro, R.: Prebiotic cytosine synthesis: a critical analysis and implications for the origin of life. *Proc Natl Acad Sci U S A*, 96, 4396-4401.(1999)
106. Shapiro, R.: A replicator was not involved in the origin of life. *IUBMB Life*, 49, 173-176.(2000)
107. Shapiro, R.: Comments on `Concentration by Evaporation and the Prebiotic Synthesis of Cytosine'. *Origins Life Evol Biosph*, 32, 275-278(2002)
108. Shapiro, R.: Origins : A Skeptic's Guide to the Creation of Life on Earth. Bantam, New York (1987)
109. Gánti, T.: The Principles of Life. Oxford University Press, Oxford, UK (2003)
110. Shapiro, R.: Small molecule interactions were central to the origin of life. *Quarterly Review of Biology*, 81, 105-125(2006)
111. Shapiro, R.: A simpler origin of life. *Scientific American*, Feb 12, (2007)
112. Dyson, F. J.: Origins of Life. Cambridge University Press, Cambridge (1998)
113. Kauffman, S. A.: Investigations. Oxford University Press, New York (2000)
114. Wächtershäuser, G.: On the Chemistry and Evolution of the Pioneer Organism. *Chemistry & Biodiversity*, 4, 584-602(2007)
115. Morowitz, H. J., J. D. Kostelnik, J. Yang & G. D. Cody: From the cover: the origin of intermediary metabolism [see comments]. *Proc Natl Acad Sci U S A*, 97, 7704-7708(2000)
116. Deamer, D. W.: The first living systems: a bioenergetic perspective. *Microbiol Mol Biol Rev*, 61, 239-61.(1997)
117. Lindahl, P. A.: Stepwise evolution of nonliving to living chemical systems. *OLEB*, 34, 371-389(2004)
118. Russell, M. J. & A. J. Hall: The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *Journal Geological Society London,*, 154, 377-402(1997)
119. Orgel, L. E.: Self-organizing biochemical cycles. *Proc Natl Acad Sci U S A*, 97, 12503-7(2000)
120. Anet, F. A.: The place of metabolism in the origin of life. *Curr Opin Chem Biol*, 8, 654-9(2004)
121. Orgel, L. E.: The origin of life--a review of facts and speculations. *Trends Biochem Sci*, 23, 491-5(1998)
122. Bradley, D.: Informatics. The genome chose its alphabet with care. *Science*, 297, 1789-1791(2002)
123. Itzkovitz, S. & U. Alon: The genetic code is nearly optimal for allowing additional information within protein-coding sequences. *Genome Res*, 17, 405-12(2007)
124. Segal, E., Y. Fondufe-Mittendorf, L. Chen, A. Thastrom, Y. Field, I. K. Moore, J. P. Wang & J. Widom: A genomic code for nucleosome positioning. *Nature*, 442, 772-8(2006)
125. Trevors, J. T. & D. L. Abel: Chance and necessity do not explain the origin of life. *Cell Biology International*, 28, 729-739(2004)

David L. Abel
Director, The Gene Emergence Project
The Origin-of-Life Foundation, Inc.
113 Hedgewood Dr. Greenbelt, MD 20770-1610 USA
life@us.net
USA Tel: 301-441-2923 Fax 301-441-8135

Keywords: Bifurcation points; Biosemiotics; Biosemiosis; Biocybernetics; Complexity; Configurable switches; Decision nodes; Dynamics; Epigenetic; Formalism; Genetic code origin; Logic gates; Natural selection; Physicodynamics; Selection pressure; Sign Systems; Symbol Systems; Self-replication; Self-organization; Self-assembly.

Abbreviations: GS: Genetic Selection; MSS: Material Sign Systems; OEE: Open-Ended Evolution.

Running Title: The GS Principle: Abel

Biographical note: Dr. David L. Abel is a theoretical biologist focusing on primordial biocybernetics. He is the Program Director of The Gene Emergence Project, an international consortium of scientists pursuing the derivation of initial biocybernetic/biosemiotic function.