Journal of Theoretical Biology 257 (2009) 345–358

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Informed Generation: Physical origin and biological evolution of genetic codescript interpreters

Peter R. Wills *,[1]

37 Fairlands Avenue, Waterview, Auckland 1026, Aotearoa-New Zealand

ARTICLE INFO

Article history:
Received 10 January 2008
Received in revised form
18 November 2008
Accepted 17 December 2008
Available online 4 January 2009

Keywords:
Molecular sequence information
Genetic codescript
Coding self-organization
GRT system
Genotype–phenotype mapping

ABSTRACT

The information stored in the genome of an organism has long been thought of in computational terms as a kind of codescript for the construction, operation and control of the system in which it is found. However, genomic sequence information can be interpreted as biological instructions and executed as a genetic codescript only by a suitably prepared cell with which the program is in proper registration. We enquire into the character of the evolutionary process that generates physical systems capable of interpreting, in increasingly elaborate ways, the genetic information they contain. The principle of Informed Generation specifies the need for the spontaneous emergence and evolutionary development of self-organizing processes that generate phenotypes from genotypes. The principle of Informed Generation describes a ubiquitous feature of biological systems: without the prior existence of certain components or functionalities, which are required for the production of themselves and others, no configuration of genetic information that accumulated through Natural Selection could ever serve as a codescript for an organism. The operation of Informed Generation is demonstrated in the stepwise evolution of genetic coding and the general distinction between Natural Selection and Informed Generation is illustrated through consideration of gene-replicase-translatase (GRT) system. It is proposed that Informed Generation represents a quite general process of evolutionary self-organization in biological systems whereby essentially irreversible transitions in the systems' dynamics take them to historically contingent, isolated states whose characteristics are determinants of biological specificity. The operation of Informed Generation may have left detectable traces in topographical features of complex intracellular and ecological networks.

1. Introduction

Darwin's principle of Natural Selection offers a partial explanation of biological evolution. It accounts for the accumulation of genetic information as a result of its continual copying and variation. However, competition for survival cannot explain the most obvious feature of evolution: functionally organized, low-entropic physical systems which autonomously maintain control by processing molecular information emerged from a state of high-temperature thermal disorder on the surface of the planet some 3.5 billion years ago, and some such systems have since displayed sporadic progress toward increased organizational complexity. Darwin's principle does not attempt to account for the existence of physico-chemical systems that are able to process information, but major transitions in the way that genetic information is interpreted and replicated have been significant causes of biological evolution, sometimes constituting a complete revision of the domain of structures and processes on which Natural Selection acts (Maynard Smith and Szathmáry, 1995). Hodgson and Knudsen (2008) have recently discussed the need to describe a general mechanism of "generative replication", as have Edelmann and Denton (2007) the failure of the Darwinian paradigm to account for "creative agency" in evolution.

It is the purpose of this paper to enunciate a principle, which we call *Informed Generation*, to describe how genetic information is used to generate the progressively complex physico-chemical systems upon which Natural Selection acts in the struggle for survival. The name Informed Generation has been chosen because the hallmark of biological systems is their capacity to generate ordered structures out of disordered material resources and because that generative capacity is, in two senses, "informed". In the first sense, each generation of an organism utilizes genetic *information*, a copy of a stable, variegated molecular configuration, which is inherited from its parent(s) in the form of nucleic acid sequences. In the second sense, the specific, detailed non-equilibrium physical state needed to generate an organism resides in the parental state from which it is derived. It is already *in-formed*, not derived or arising from some set of environmental

* Tel: +64 9 373 7599x88889; fax: +64 9 373 7445.
E-mail address: p.wills@auckland.ac.nz
[1] Department of Physics, The University of Auckland, Private Bag 92019, Auckland, New Zealand.

0022-5193/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2008.12.030

boundary conditions or events outside of internal cellular processes. All such processes, wherever they occur, have a continuity that is traceable to the origin of life. Indeed, the operation of Informed Generation is seen most readily in the molecular processes responsible for the coevolution of genetic information and coding, that is, the first informationally determined genotype–phenotype relationships. We can expect that the explicit role of Informed Generation in evolution will be demonstrated through bioinformatic analysis of the sequences of the proteins responsible for genetic coding.

Organisms are distinguished from non-living natural systems in that the thermodynamic constraints on their innate biochemical processes are continually controlled by reference, either direct or indirect, to repositories of molecular sequence information. Schrödinger (1944) envisaged that an organism's chromosomes "contain in some kind of codescript the entire pattern of the individual's future development and its functioning." Von Neumann (1949) was the first formally to use the theory of machine computation (Turing, 1936) to describe how organisms use information to function and reproduce. He demonstrated that it is logically possible to construct a closed cycle of primitive automatic processes for the controlled production and location of all of a system's component parts as well as the replication of information needed for their specification. More importantly, because they possess a universal constructor, von Neumann automata are capable of undergoing "open-ended evolution" in which the organizational complexity of the systems generated is, in principle, unbounded (McMullin, 2001). The capacity for open-ended evolution is regarded as a definitive feature of living systems that distinguishes them from all others (Bedau, 1996; Ruiz-Mirazo et al., 2008). On the other hand, von Neumann automata are not explicitly subject to the constraints that the laws of physics and the disordering effects of thermal motion impose on material processes, the problems with which Schrödinger was concerned. Nor does von Neumann's theory give any indication as to whether we should expect to find actual material instantiations of his automata in the natural world. That is why it is necessary to explain the spontaneous occurrence of ordered physical systems that are able to generate themselves by interpreting genetic codescripts. The lack of naturalistic explanations, explicitly for the appearance of von Neumann-type autonomous systems in the universe, seems to have lent succor to the idea that organisms must be the product of some external "intelligent design" and that biology is still subject to intentional, supernatural intervention (Behe, 2007; Dembski and Ruse, 2004).

At the dawn of molecular biology Schrödinger (1944) identified two order-generating processes that characterize the physico-chemistry of living systems. The first of those order-generating processes, "order from order", is the storage of information in what he called an "aperiodic crystal" whose exact form can somehow be copied and expressed. Schrödinger's specification of the carrier of genetic information, which he borrowed from Delbrück (Timoféeff-Ressovsky et al., 1935), was vaguely prescient of the linear array of nucleotides in nucleic acids and the consequences of complementary base-pairing (Watson and Crick, 1953). For more than half a century, molecular biology has been dominated by elaborations of Schrödinger's "order from order" principle. Following the discovery of the base-paired structure of DNA by Watson and Crick (1953), information stored in DNA sequences quickly came to be viewed as the central agent of biological control. This idea was formalized by Crick (1958, 1970) when he enunciated the Sequence Hypothesis and Central Dogma of molecular biology. Taken together, the ideas of the genetic control of biological specificity and the one-way transmission of genetic information continue to lend credibility to an exclusively selectionist interpretation of all evolutionary phenomena (Dawkins, 1976, 1986).

The second type of order-generating process identified by Schrödinger, "order from disorder", comprises the maintenance of out-of-equilibrium macroscopic structures which "feed off negentropy", converting unstructured matter and free energy into specific functional forms. The potential relevance for theoretical biology of general developments in the area of irreversible thermodynamics has long been recognized (Glansdorff and Prigogine, 1971; Morowitz, 1968; Prigogine and Defay, 1954; Prigogine and Nicolis, 1971) and over a period of decades biologists themselves have elaborated quite general theories of biology based on thermodynamic principles (Wiley and Brooks, 1982; Toussaint and Schneider, 1998; Schneider and Kay, 1994; Smith, 2008a, b, c; Weber et al., 1989). These theories have contributed an understanding of how ordered structures can be generated in dissipative systems, but they lack the aspect of cybernetic control, the integrated information-based modulation of physico-chemical processes, that is the *lingua franca* of molecular biological descriptions of intra-cellular events. Although the thermodynamic description allows us to attribute the intricate dynamic structure of cells to the complex network of couplings among dissipation-driven reaction and diffusion processes that continually transform and move, in an apparently orderly fashion, the multitude of distinct molecular species found inside cells, it doesn't explain the importance of genetic information. Thus, Kauffman (1993) has assembled evidence and analysis covering virtually every aspect of the operation of biological systems to show how order can be generated in complex dissipative systems and he goes on to suggest that the dynamics of such complex systems are typically poised at a point of instability, an idea that Bak (1997) formalised in the theory of self-organized criticality. However, neither Kauffman (1993) nor Bak (1997) attempts to relate how the execution of a genetic codescript might be required for the emergence, maintenance and control of complex biological systems.

According to Kauffman (1995) genetic information storage is not a precondition for life. Rather, he sees generalized forms of autocatalysis and their coupling to cyclical processes that do thermodynamic work as constituting "autonomous agents" (Daley et al., 2002; Kauffman, 2000) which are the forerunners of functional organisms, irrespective of whether they carry or utilize genetic information. However, by not incorporating an obligatory role for stored information in proto-biological systems, Kauffman's analysis lacks an explanation of the connection, which molecular biologists see as so definitive, between functional specificity, the details of physical structures and their genetic encoding. Likewise Bak (1997) seeks the cause of the dynamic order, observed in phenomena ranging from the sub-microscopic to the cosmological, in the instability that characterizes the transition between the "frozen-in" structure that strong interactions impose on a population of entities and the "boiled up" chaos that they display when they operate independently of one another. Although Eigen's criterion for the accumulation of information due to Natural Selection applies to systems with these dynamics (Wills et al., 2004), information *per se* plays no essential role in Bak's theory of self-organized criticality.

We now proceed with a discussion of how genetic information functions as a codescript and the relationship between the ideas of molecular biology and the theory of computation. This leads us to the enunciation of the principle of Informed Generation as a general description of the inherent capacity of physical systems to produce self-organized interpretations of information. We then describe the operation of Informed Generation in a simple model system that illustrates how spontaneously generated genetic coding of increasing complexity can occur when specified molecular components have certain catalytic properties. It is postulated that Informed Generation drives biological systems,

including evolving organisms, to historically contingent dynamic states in which genetic information is exquisitely matched with a particular interpretation, a question that should be empirically testable. The different roles played by Natural Selection and Informed Generation in evolution are discussed, particularly as they relate to our understanding of the origin and character of living systems.

2. Genomic sequence as codescript

Consider the process employed in cloning experiments when a denucleated zygote is impregnated with the nucleus of an adult cell. An extant physical structure is supplied with the DNA sequence information requisite for the construction of an organism with already determined individual traits. The specification of these traits is carried to an observable degree in the sequence of the DNA that is transferred. In fact, much more than just DNA sequence information is supplied in such experiments. The subtle epigenetic effects of the state of methylation and other features of the newly supplied DNA (Bird, 2007; Johannes et al., 2008) show already that a genome sequence is not in itself an autonomous codescript for the generation of any organism. The genomic sequence must be in a suitable state of "registration" with the molecular equipment that interprets it. However, the very fact that an individual animal can be cloned through chromosome transfer confirms the role of nucleic acid sequence information as a precise determinant of biological specificity when it is supplied to a matching physical structure in an appropriate state. The transfer of complex genetic traits between diverse taxa by using well-established techniques of genetic engineering is further proof of the role of nucleic acids as carriers of detailed biological specificity encoded in the form of molecular sequence information according to the principles first enunciated by Crick (1958).

There is obviously a minimal degree of matching that is needed between DNA sequence information, the physical structure that surrounds it and the state of registration between the two, for the joint system to be viable. Why is it not possible to place the genome of an *E. coli* cell into a denucleated human cell eventually to produce further ordinary *E. coli* cells, or *vice versa*? The simple answer is that the interpretation of genetic information is, to a high degree, specific to the detailed physical structure of the environment in which it occurs. A human cell in an appropriate state is needed to interpret the human genome as a codescript for a human organism and it can do so with an extraordinary degree of individual specificity, but the process has inherent limits. Species divisions, or possibly coarser divisions between taxa, likely give a rough guide to the limits of variation tolerated in the state and structure of the physical system minimally needed to generate a specified organism from a given genomic codescript (Lartigue et al., 2007). Controversy surrounds proposals to investigate these limits through experimentation with cells of human origin (Hopkin, 2007).

Every organism inherits from its parent(s) not only a repository of genetic information but also a complex physical structure which interprets that information to generate the organism. The question of how the processes that generate organisms have arisen historically and become organized into specific constellations associated with diverse taxa is a central problem for theoretical biology.

When we use the parlance of molecular biology there is a well-defined separation between the molecular sequence codescript found in an organism's genome and the molecular machinery that uses that information to "compute" a new physical state of the system, including the generation of an organism from a zygote, in the manner of a von Neumann automaton. Eigen (1971) takes the separation between information and function to be fundamental to biology but it is quite arbitrary from the perspective of the theory of computation, so the proper connection between computational and physical theories of biological processes needs to be carefully established. The operation of a self-reproducing automaton can be computed in any number of ways on any Universal Turing Machine, each version of the computation having been constructed with a different specification of the nominal separation between machine states and stored information. Thus, the evident separation between genetic information and its phenotypic functional expression is, although conceptually important from a biological perspective, of no significance when cells are represented as classical Turing-type machines (von Neumann, 1949) with multiple states and corresponding, albeit stochastic, transition rules. In principle, we could, as allowed by Turing's theory of computation, represent the reproduction of a cell deterministically in such a way that we assigned some suitably chosen feature of its complete, spatially-extended molecular configuration to be its genetic codescript. In that case the DNA of the cell would be one necessary part, among many others, of a machine that performed operations and went through multiple changes of state, but the DNA would no longer serve the role of the machine tape that instructed the construction and operation of the machine. Gatenby and Frieden (2007) have made suggestions down these lines.

Therefore, from a computational perspective it becomes necessary to ask why is it that the characteristics of cells can be changed so simply by altering their DNA sequences, but there are only a few known examples, prion effects being the most demonstrable, of heritable changes that can be brought about through alteration in some other detail of molecular configuration. The molecular biological discussion about DNA as a program *cum* codescript (Schrödinger, 1944), instructions (Eigen, 1971), algorithm (Dawkins, 1986), blueprint (Kornberg, 1987) or Master Molecule (Kornberg, 1983) demonstrates the widespread belief that organisms indeed carry out computationally definable operations using molecular sequence information.

Consider the sequence of the human genome as it is recorded on a common compact disk that can hold of the order of 10^9 bytes of information. It is unproblematic to describe the construction of a machine that plays some chosen recording of a piece of music, or carries out any other computation requiring a Turing machine input tape of about the same length, when a compact disk record of the human genome sequence is inserted into it. Such a machine could come into existence through a series of improbable, chance, accidental events rather than through the exercise of any conscious intent to create a device that decodes the sequence of the human genome as a recording of an orchestral performance. The blind creation of read-only memory (ROM) holding what a cryptographer would call the necessary "one time pad", followed by its functional incorporation into an audio CD player, is all that would be required. An engineer intending some completely different outcome from the creation of a one-time-pad ROM chip could, through a simple error of mislabelling, produce the machine in question. Thus, under accidental circumstances that could plausibly occur in the physical universe, the sequence of the human genome could be the codescript for a recording of an orchestral performance, rather than the construction of a living human being. The possible interpretations of any codescript depend only on the availability of suitable physical systems that produce different particular outputs when the codescript is executed. The human genome is a codescript for a human being only because human beings already exist. It could not act as a codescript for the construction of a human in a universe in which no humans, or their close relatives, existed.

Our theoretical task is to circumscribe the features of the physical world and the natural processes which are prerequisites for the sequence of a genome to constitute a heritable codescript for the generation of a complete organism rather than some other physical occurrence. One solution proposed to this problem is to say that the evolution of anything resembling life is not an inherently characteristic outcome of natural processes. Then, the only proper scientific explanation of the link between genomic sequences and organisms would be a list of the undirected unique events whereby slowly accumulating molecular information and its expression in the form of increasingly complex phenotypes have been linked together in an unbroken series of events spanning the entire process of biological evolution. This is like reconstructing the accidental recording of each organism's one-time-pad genetic codescript. The phylogenetic trees that practitioners of bioinformatics commonly build by using DNA and protein sequence analysis address the theoretical problem in such terms. The physical systems (organisms) in which the genetic sequences have occurred appear in the analysis only as labels. Anything to do with organisms' forms, how they were generated or how they are maintained, is irrelevant to the very successful enterprise of building phylogenetic trees from genetic sequence information based on assumptions about mutation rates. On the other hand, from the perspective of systems biology, all of the interesting and important information is missing in simple sequence analysis.

Systems biology seeks an understanding of organisms and their evolution that goes far beyond the changes that have progressively occurred in DNA-sequence codescripts (Boogerd et al., 2007). It seeks to describe the structure and operation of the functional machinery which autonomously executes algorithms to generate organisms from codescripts. Ultimately, the curious systems biologist will want to know how self-constructing machinery that uses internal symbolic records first appeared out of disordered matter and why such machinery has been elaborated in increasingly complex forms through the aeons of biological evolution.

3. Principle of Informed Generation

Informed Generation is the coordinated process whereby certain traits or structural components of biological systems produce other such entities and participate in their own production. The operation of this coordinated process has an obligatory requirement for some fixed configuration within the system, usually the sequence of nucleic acids, in which genetic information is represented. Taken together, the entities that participate in the process constitute a self-generating interpreter of genetic information. Without the existence of such a self-generating interpreter, the genetic information found in any biological system would be of no more systemic significance or functional relevance than, for example, the random pattern of atomic defects found in any crystal.

The information-based generation of specific biological structures requires the presence of extant physical structures and components whose mode of action is highly specific. Informed Generation occurs as a result of multiple specific components acting in a concerted, often sequential fashion to produce new or replacement structures and components, most of which themselves play some role in determining the phenotypic specificity of genetic information which they together interpret. The constellations of coordinated processes that determine phenotypic specificity have developed, diversified, specialized and undergone steep transitions in complexity during different phases of evolution, thereby molding their representation in, and their interpretation of, the genetic information with which they are associated both spatially and temporally. The processes that support the Informed Generation of biological structures have a historical continuity that reaches back to the origin of life, described by West-Eberhard (2003) as "the continuity of the phenotype". It has so far proved impossible to sustain life outside of this continuity, for example, by synthesizing, *de novo*, an intracellular milieu which can be used to interpret a genome properly. Because of the fundamentally interactive, cooperative mode of operation of the processes through which phenotypes define themselves, this continuity cannot always be traced accurately through tree-like genetic lineages and its evolutionary development cannot be attributed alone to corresponding changes in the genetic representation with which it is associated.

It is useful at this stage to allude to the analogy of Dawkins (1986, p. 111) who discusses the dissemination of seeds from a willow tree as a shower of tree-generating programmes. What is lacking in Dawkins' description of evolution is a discussion of how self-organized computers which are capable of executing any tree-generating programmes first arise and then continue to operate. The existence of such machines is taken as a simple fact. They somehow cobble themselves together. However, Dawkins' analogy might be framed more accurately by describing the dissemination of seeds as a shower of robotic machines programmed to transform themselves into trees using environmental resources. Such a reframing of the analogy detracts from the simplicity of the Neo-Darwinian emphasis on the necessary role of genetic inheritance in Natural Selection, but it points to a much more difficult problem in evolutionary theory. Darwin himself left the origin of living organisms as an unanswered question and von Neumann (1949) likewise gave no indication as to how a self-reproducing automaton could arise spontaneously in any physico-chemical system. How then does functional information-processing arise in natural systems such that genetic information can act as a codescript for an organism? As a result of what mechanisms does Informed Generation occur?

Informed Generation is an inherent possibility in any dynamic physical system fulfilling two criteria. First, the system processes must be autocatalytic, and second, the operation of the autocatalytic cycle must depend on the stable configuration of some generic component of the system. The first criterion specifies the generative capacity of the system and the second criterion specifies the requirement for information. The emergence of genetic coding at the origin of life demonstrates Informed Generation as a result of a symmetry-breaking non-equilibrium phase transition (see below). Beyond the transition are to be found populations of proteins with a limited range of highly specific structures and catalytic specificity adequate for the maintenance of their integrity, in spite of individual turnover, as long as the required nucleic acid information remains intact. More generally, Informed Generation could occur as a result of self-organization within and among complex dynamic networks at all levels in the biological hierarchy: metabolism, protein interactions, genetic expression and control, intra- and inter-cellular signaling, biomass transfer and ecology.

Biological self-organization has been characterized in terms of diverse, sometimes disparate, concepts derived from the theories of Natural Selection (Dawkins, 1976, 1986; Eigen, 1971), thermodynamics (Prigogine and Nicolis, 1971), network connectivity (Kauffman, 1993) and criticality (Bak, 1997). The principle of Informed Generation is distinguished from all of these in that it characterizes biological self-organization as the automatic production of a self-generating interpreter of genetic information. Such a characterization of biological self-organization requires a detailed account of the alternative algorithmic interpretations of extant genetic information that are possible at any stage of evolution, an explanation of why certain interpretations end up

dominating over others and a description of how dominant interpretations determine, sometimes solely, the selective fitness of the genotypes with which they are associated. In what follows we seek to take some modest first steps in this direction by considering the emergence of von Neumann-type "general construction" processes in a model physico-chemical system that is subsequently capable of undergoing open-ended evolution.

We consider a system in which bytes of information are recognized in the general manner of the three-nucleotide codons whose sequences dictate the process of protein synthesis in molecular biological systems. An alternative might be to choose much larger bytes, whole genes for example, and consider their expression in the manner of a genetic regulatory network of the type that controls morphogenesis, or in the manner of the traits that govern the interactions between species in an ecosystem. In these more complicated examples, the mathematical construction and results would be of a different form, but the general conclusion to be reached concerning Informed Generation in biological systems would be unaltered.

4. Informed Generation in a simple system

We consider a physico-chemical system in which the generative processes that produce new components of the system are carried out by the components themselves. We assume that there are μ different operations which components can potentially perform, and that a series of ν operations is required to synthesize a new component. Given a situation in which a unique series of operations produces a unique structure, the number of different possible components that could ever be found in a large ensemble of such systems is μ^ν, which is likely to be a hyper-astronomically large number.

We now assume that the system contains a repository of information and that the operations needed to generate components of the system are performed only through the sequential recognition of bytes of information. If there are λ unique forms of a byte, then there will be $\lambda \times \mu$ composite, recognition-dependent "read then do" operations. If such byte-recognition-dependent operations are all performed with equal probability then the structure of the new component produced through execution of ν operations will be one selected randomly from the possible μ^ν, no matter what information is present in the system. Specificity of both structure and function can only be attained if there is some statistical bias in the association of some of the μ operations with some of the λ bytes. For the sake of simplicity we shall assume that the system contains genetic sequences which each consist of ν bytes that inform the step-by-step generation of new components through series of recognition-dependent operations. Every operation that is carried out in the system will be assumed to be executed by an independently acting component of the system.

The dynamic behavior of a system such as this depends primarily on the structure-function relationship that defines the specificity of action of different components. We consider "unbiased" examples of possible structure-function relationships in which there is no preferential association, in terms of the structure of all μ^ν possible components, between any byte and the operation that may be carried out when it is recognized. That is to say, in the array of all possible component structures, there could be found as many that recognize any particular byte and perform a particular operation as there are for any other byte-operation pair: from an *a priori* structural point of view, any byte of information can potentially be interpreted in every possible way. In order to make the analysis more tractable, we also suppose that any given component performs at most one recognition-dependent operation and that a small fraction f of all possible component structures, randomly distributed with respect to the sequence of operations needed to produce them, can perform any composite recognition-dependent operation. On the basis of these assumptions we can calculate the time evolution for the probability p that the operations performed in the system belong to a selected subset **S** of all those possible. The subset **S** must ordinarily contain at least one operation to be performed when any one of the λ byte types is recognized. In the case of the standard genetic code, there are $\mu = 61$ codons and $\lambda = 20$ amino acids. Our assumption corresponds to the reasonable proposition that, from a structural perspective alone, the construction of a specific amino acyl tRNA-synthetase is equally feasible for any chosen amino acid to codon assignment, whether or not it exists in an extant biological system.

For ν sufficiently large, it is readily shown (Wills, 1993, 1994) that when the information in the system permits the self-construction of components that perform operations in **S**, and the probability of performance of any operation depends solely on the availability of components capable of performing that operation anywhere within the system, the time evolution of p is given by

$$\frac{dp}{dt} = \alpha - (\alpha + \beta)p + (\gamma - \alpha)p^\nu - (\gamma + \delta - \alpha - \beta)p^{\nu+1} \quad (1)$$

where α, β, γ and δ are constants that depend on purely formal properties of the information-carrying polymer sequences in the system. The coefficient α represents the proportion of ν-long sequences of informed operations, not all belonging to **S**, that produce components that perform operations belonging to **S**; and β represents the proportion of ν-long sequences of informed operations, not all belonging to **S**, that produce components that perform operations not belonging to **S**. The coefficients γ and δ have definitions complementary to those of α and β, respectively, for sequences of operations all belonging to **S**. Synthesis of system components can be specified as follows: the rate of production of components, having population $x_{\mathbf{S}}$ in the system, that perform operations belonging to **S** is given by

$$\frac{dx_{\mathbf{S}}}{dt} = w_0[\alpha(1 - p^\nu) + \gamma p^\nu] \quad (2)$$

and the corresponding rate of production of components that perform operations not belonging to **S** is given by

$$\frac{dx_{\bar{\mathbf{S}}}}{dt} = w_0[\beta(1 - p^\nu) + \delta p^\nu] \quad (3)$$

Here w_0 represents the overall rate at which components are produced in the system. Under the assumptions of this model we obtain the simplification $p = x_{\mathbf{S}}/(x_{\mathbf{S}} + x_{\bar{\mathbf{S}}})$, whence Eq. (1) is readily derived from Eqs. (2) and (3).

We can assign the convenient values of $\gamma = 1$ and $\delta = 0$ to the constants in Eq. (1) without significant influence on the conclusions to be drawn. In the simple case that $\lambda = \mu$, the remaining constants have approximate values $\alpha \approx \lambda f$ and $\beta \approx \lambda(1-\lambda)f$ and, in the physically meaningful range of the parameter space, there are two stable stationary solutions of Eq. (1) separated by a transcritical bifurcation (Wills, 1993, 1994). The first stable state occurs at $p = 1/\lambda$ and describes the random performance of operations and the generation of components with random structures. The second occurs at $p = 1$ and describes the synthesis of only those components that perform operations belonging to **S**. Both of these stationary states are stable for f in the range with the approximate limits $\nu(\lambda - 1)/\lambda^{\nu+2} < f < [\nu\lambda(\lambda - 1)]^{-1}$. The domain of attraction for the $p = 1$ state expands and that of the $p = 1/\lambda$ state contracts, with decreasing f, along the transcritical boundary.

In this example, the system could begin with a random selection of components that collectively performed all operations

at approximately equal, very low rates. In this initial state, $p = 1/\lambda$. Then, as a result of a fluctuation in $x_{\mathbf{S}}$ that took p just above the point of instability marking the boundary between the domains of attraction of the two stable stationary solutions of Eq. (1), the system would be driven to a state with $p = 1$ and the only active components in the system would be those that perform operations belonging to the subset **S**. It should be noted that Eqs. (1)–(3) describe a range of dynamic possibilities, depending on the values of the constants α, β, γ and δ. These constants derive their values from the formal structure of the mapping through which subsets of the μ^{ν} possible components are assigned the capability of carrying out the different μ operations. Thus, the dynamic behavior of systems described by Eq. (1), including their capability for self-organization, depends on the "embedding" of operations in the space of possible components, what biologists would call the "structure-function relationship" of the active components (Nieselt-Struwe and Wills, 1997; Wills, 2001).

Although Eq. (1) represents but one, quite specific example of an information-directed self-constructing system, it serves as an illustration of the general principle of Informed Generation. In the self-organized state beyond the dynamic instability, the joint operation of the functional components in the system described by Eq. (1) serves as an interpreter of the information available to the system. The most basic requirements for this possibility to be realised are (i) that there exists some mechanism for sequential interaction between the information configuration and the active components, and (ii) that the information be "reflexive" (Wills, 2001) in the sense that the action of operations belonging to the specified class **S** produce, by reference to that information, components that perform the operations in question (i.e., $\gamma \neq 0$). Although the transition from an initial random state to a fully ordered, functional, dynamic state of Eq. (1) is a rather simple and biologically atypical example of self-organization in a genetic information-processing system, features of this transition are typical of the mechanisms through which progressively precise general constructors, comprised of essentially independent functional components, can emerge from an disordered physico-chemical system, as occurred in the case of genetic coding. Weak coupling produces cooperativity of the sort defined in the theory of phase transitions, not the complex functional cooperativity found in biological interaction networks. However, in the same way as the idealized theory of quasi-species (Eigen, 1971; Eigen and Schuster, 1979) provides a sound basis for understanding the basic process of Natural Selection, even though it does not describe the richness and complexity of evolution in the real world of ecological networks, so too is Eq. (1) a useful starting point for understanding the main feature of Informed Generation, namely, the spontaneous emergence and establishment of a previously non-existent phenotypic meaning for some fixed information. We will now discuss the application of Eq. (1) to the specific case of genetic coding.

5. Molecular biological coding

Protein synthesis in cells is a concrete example of Informed Generation because the structures of the protein components needed to maintain the algorithmic integrity of the process of translation are encoded in genes and are themselves products of their collective operations. Such information-based protein production could have first occurred only when there was some mechanism for the synthesis of specific amino acid sequences that were collinear with molecular sequence information which was stored in nucleic acids. One can envisage a primitive system in which tRNA-like molecules, charged with amino acids, were lined up, base-pair conjugated to a genetic sequence through their anticodons, so that some catalytic process, probably initially very inefficient, could concatenate the amino acid residues to form a peptide. In modern cells the process of protein synthesis is highly refined and is carried out by ribosomes and their numerous cofactors. However, whatever the mechanism of protein synthesis happens to be, there could be no protein-based genetic coding, as there exists in modern cells, and therefore no specificity in protein production, without the operation of a population of enzymes capable of differentially charging tRNA species with cognate amino acids. In terms of the theory of computation there is a bootstrap problem. How could the enzymatic specificity that effects codon to amino acid assignments be generated unless it already existed? The answer is that self-organization in appropriate physical systems includes the possibility of the emergence of self-defining algorithms that map genotypes to phenotypes through a form of computation in which the separate generic elements of a stable molecular configuration serve as bytes of information.

Let us take the bytes of genetic information to be codons, the operations to be the charging of tRNAs and the components in question to be amino acyl-tRNA synthetases (AARS enzymes) and let us choose $\mu = \lambda = 2$ so that only two classes of both codons and amino acids are distinguishable in the system. Eq. (1) now describes the emergence of a binary genetic code from a system in which RNA-dependent peptide synthesis begins without any preferential assignment of codons to amino acids, that is, a situation in which the extant nucleic acid sequence information has no meaning.

The ready emergence of a binary code explains the existence of two distinct classes of modern assignment catalysts, the Class I and Class II AARS enzymes which, with minor exceptions, still divide amino acids into roughly equal disjoint sets in all known organisms. Each AARS class has a distinctive catalytic core, whose basic three-dimensional structure is common to all enzymes of that class across all taxa, The catalytic cores of both Class I and II AARSs have been preserved since the origin of specific protein synthesis in prebiotic systems (O'Donoghue and Luthey-Schulten, 2003). One could speculate that the essential protein structures were originally generated from the information in two easily replicated, short, complementary nucleic acid sequences (Rodin and Rodin, 2006). Whatever the case, it seems clear that a very early bifurcation in AARS functionality was an integral step in the evolution of information-based molecular biological specificity and that the results of the dynamic transition involved in the process have been preserved and maintained ever since the first proteins were generated through a process akin to molecular biological translation. It has often been asked why AARS functionality should have evolved in two quite distinct ways when later adaptations of one version would apparently have sufficed to produce the diversity of function needed for the modern code. In terms of our current discussion, the existence of the two classes of AARS enzymes is a palimpsest of Informed Generation in the prebiotic environment on the surface of the earth some 3.5×10^9 years ago, not the result of unnecessary functional redundancy that has somehow survived subsequent rounds of mutation and selection in every branch of the tree of life.

Let us now enquire a little further into the character of the information-processing that this first coding transition produced. When only two classes of amino acids can be distinguished, the population of proteins produced through translation of any genetic information is indeed "statistical", exactly in the manner described by Woese (1965), comprising, in all probability, mostly non-functional or inactive species, but with overall catalytic activity for the "correct" assignments dominating over "incorrect" assignments. Results of an illustrative simulation of the transition

to this state [details are available in Wills (2004)] are shown in Fig. 1(a). For these simulations it has actually been assumed that four extant amino acids, {*a*, *b*, *c*, *d*}, are available, a situation thought to have existed in the primitive prebiotic world (Eigen and Winkler-Oswatitsch, 1981). The codons have been partitioned into four corresponding classes, {*A*, *B*, *C*, *D*}, assumed to be disjoint and labeled so that they corresponded to the amino acids to which they finally become assigned. The binary subsets of amino acids, $k \equiv \{a, b\}$ and $l \equiv \{c, d\}$, and codons, $K \equiv \{A, B\}$ and $L \equiv \{C, D\}$, refer to Fig. 1(a). The simulation was started with an initial protein population which was chosen at random and catalysed all 16 assignments $\{A, B, C, D\} \equiv X \rightarrow y \equiv \{a, b, c, d\}$ at approximately equal, almost infinitesimally small rates and all 1.7×10^7 possible polypeptide sequences of length 12 were produced with equal probability. Following the first transition [Fig. 1(a)] the system produced a subset of polypeptide sequences, comprising just 4.1×10^3 of the 1.7×10^7 possible. This population of "statistical proteins" (Woese, 1965) had sequences that were disproportionately represented by those which catalyse the binary coding assignments $[K \rightarrow k,\ L \rightarrow l]$ but most of which were practically inactive. A second transition evident in Fig. 1(b) takes the system from the binary code $[K \rightarrow k;\ L \rightarrow l]$, metastable in this system, to a fully fledged quarternary code $[A \rightarrow a;\ B \rightarrow b;\ C \rightarrow c;\ D \rightarrow d]$. After the second transition the system produces, by and large, only four different polypeptide sequences, each of which corresponds to a species capable of catalysing one of the four assignments of the full quarternary code. As an alternative measure of the increasing complexity of information processing in the system, it could be said that before the first transition all potential phenotypes were indistinguishable whereas after the first transition there were 4.1×10^3 distinguishable phenotypes; and after the second transition the system was capable of producing 1.7×10^7 distinguishable protein phenotypes, given genetic information appropriate to any one of them. The evolution of the AARS enzymes is characterized by successive bifurcations in the identity of the recognizable classes of the amino acids (O'Donoghue and Luthey-Schulten, 2003; Delarue, 2006), akin to the transitions that occur in the these simulations, culminating in the set of the twenty canonical amino acids found in the proteins of modern organisms.

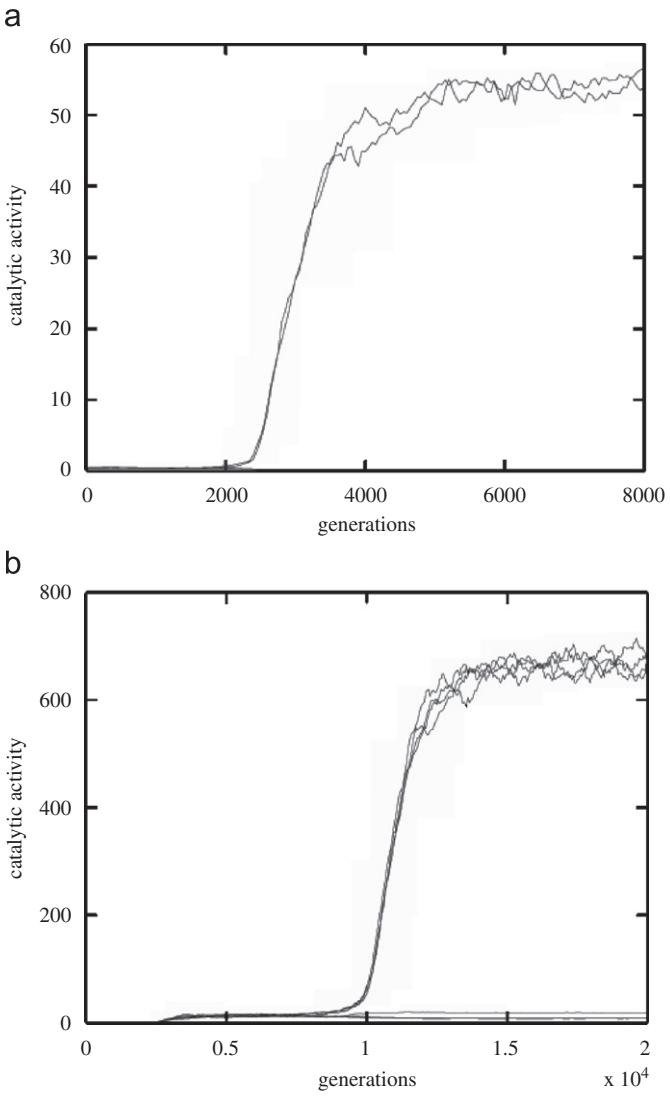

Fig. 1. Simulation of coding self-organization. Each of the time traces represents the net rate at which one codon-to-amino-acid assignment $X \rightarrow y$ is catalyzed in the system: (a) time traces are shown for each of the four possible binary assignments $\{K, L\} \rightarrow \{k, l\}$. The two assignment functions comprising the binary code $[K \rightarrow k, L \rightarrow l]$ are selected during a transition that occurs at about 3500 generations and (b) time traces are shown for each of the 16 possible binary assignments $\{A, B, C, D\} \rightarrow \{a, b, c, d\}$. After the first transition to a binary code $[\{A, B\} \equiv K \rightarrow k \equiv \{a, b\}, \{C, D\} \equiv L \rightarrow l \equiv \{c, d\}]$ the four assignment functions corresponding to the quarternary code $[A \rightarrow a, B \rightarrow b, C \rightarrow c, D \rightarrow d]$ are selected during a transition that occurs at about 1.05×10^4 generations.

Unlike the complex system of ribosomal protein synthesis, actual AARS enzymatic functionality has not become subject to extensive protein network interactions. This independence of functionality of individual coding assignments is believed to be an essential feature of genetic coding in present-day cells (Woese et al., 2000) and it also allows us to use Eq. (1) to describe, in terms of a series of simple dynamic transitions, how the successive operation of an elementary form of Informed Generation has produced the coding components of the "general constructor" that every organism needs in order to synthesize the proteins it requires to maintain and reproduce itself by using genetic information.

It is anticipated that it will be possible to test empirically whether the catalytic cores of the AARS enzymes originally evolved through Informed Generation. Using protein sequence data from sufficiently diverse taxa and covering all AARS types, it may be possible to show that the pattern of amino acid placements in specific sequence positions is, after more than 3.5 billion years of subsequent evolution, still biased toward the pattern of bifurcations in amino acid recognition through which the AARS enzymes first developed and eventually produced the species, each of which recognizes one of the canonical amino acids. Such work is in progress. It must be emphasized that this bioinformatic undertaking is fundamentally different from the standard phylogenetic analyses reported by either Woese et al. (2000), based on the primary structures of the complete enzymes, or by O'Donoghue and Luthey-Schulten (2003), whose work included considerations of three dimensional molecular structures. In the novel analysis underway, early branchpoints in the phylogenetic trees of the AARS enzymes are presumed to represent times when entire extant populations of the tRNA synthetase "statistical proteins", rather than specifically selected species with well-defined sequences, self-organized so that the emergent population was able to discriminate new amino acids (or classes thereof). After each transition the extant classes of amino acids were better differentiated, leading eventually to the set of species-specific AARS enzymes with separate coding specificities for the 20 canonical amino acids found in modern day cells.

6. Generative and replicative processes

We now turn our attention to the coupling between generative and replicative processes in biological evolution. As we saw in our discussion of genomes as codescripts, any generative

genotype–phenotype mapping depends on the existence of a suitable physical system for effectively constructing phenotypes, given the necessary genetic information. Informed Generation can occur only when the information available to the system is "reflexive": its interpretation by the constructive components of the system must result in the production of those particular components, at least (Wills, 2001). For example, the system whose simulation is illustrated in Fig. 1 can make the transition to the state of ordered information processing only when the relevant assignment catalysts, once present in the system, interpret the extant information so that they concertedly and cooperatively produce themselves rather than other polypeptide sequences. Where could the information that reflexively encodes the sequences of a coding set of assignment catalysts have come from? Even if such information were present in a system to start with, as soon as it was replicated through any error-prone mechanism, the specific information conferring on the system the property of reflexive production of assignment catalysts would quickly be lost. An illustration of the rapidity with which the capacity for dynamic self-organization is lost from systems of assignment catalysts as a result of information decay has been provided by Wills (1994). The problem we must now pose is exactly the opposite: how, against all odds, could a specialized genetic sequence of the sort required for the emergence of genetic coding through Informed Generation first arise in biological systems and, even if it did arise spontaneously, why should it have been preserved in the presence of the disordering effects of mutation? We wish to investigate answers to these questions that do not resort to the tautology that genetic information can only be preserved in systems that survive, through Natural Selection, the deleterious effects of mutation.

In order to be quite concrete, let us reconsider our simulations of prebiotic coding, but with an extension. Let us assume that, in addition to the complete range of assignment catalysts, there are, among the μ^{ν} possible polypeptide species, those which serve as general nucleic acid replicases and which, whenever present in the system, indiscriminately make copies of whatever genetic sequences are present. The general scheme of such a gene-replicase-translatase (GRT) system is shown in Fig. 2. We shall consider the operation of a system like this in a reactor vessel and assume that genetic sequences, as well as protein species, are lost from the system in an indiscriminate fashion. The lack of discrimination is imposed on the mode of action of the putative replicases, as well as loss processes, so that there is no intrinsic bias conferring either a reproductive advantage or disadvantage on the genetic sequences needed to support the occurrence of Informed Generation. Clearly, Natural Selection will not occur among the genetic sequences present in the system as we have described it so far, because all of the rates of replication (species fitnesses) will be equal and the rates of mutation from one sequence to another will also be equal. Accordingly, the system will be incapable of maintaining the information necessary for any ordered function (Eigen, 1971). The information necessary to support Informed Generation can be maintained, or accumulate in the first place, only when the replication of an appropriate reflexively-encoding sequence is somehow favored over other sequences, conferring on it a selective advantage.

A fairly general characteristic of biological systems is that the physical structures needed for reproduction are generated as a result of the expression of genetic information contained within the system. The role of behavioral characteristics in the reproduction and survival of complex organisms is similarly subject to modulation through genetic effects, often emanating from other species living in a particular organism's environment. In the very simple system we are considering, it has been assumed that the nucleic acid replicase is a polypeptide, itself a product of the expression of genetic information. That being the case, it is possible for the replication of genetic information to become associated, temporally and spatially, with regions of high chemical activity. Such an association between the capability to generate system components T and R from genetic information G and to replicate the genetic information can occur in our GRT system as a result of an undirected physico-chemical process: chemical reaction-diffusion coupling. In far-from-equilibrium systems local changes in the relative population numbers of different molecular species due to the simultaneously acting processes of chemical reaction and diffusive movement can, as a result of the coupling of those processes, generate mesoscopic or macroscopic spatially differentiated patterns, as first described by Turing (1952) in relation to the spontaneous appearance of spatial patterning during the morphogenesis of multicellular organisms.

It has been demonstrated by Füchslin and McCaskill (2001) and Markowitz et al. (2006) that reaction-diffusion coupling in GRT systems can result in the generation and maintenance of spatially differentiated "spots" of encoded protein production and preferential gene replication. The most important feature of GRT dynamics is the co-dependent bootstrapping, first of Informed Generation and then of Natural Selection, from an initial situation in which the primary structures of the macromolecules in the system, and those being produced, both polypeptide and nucleic acid, are random. What is demonstrably impossible when the system is well-mixed (Füchslin and McCaskill, 2001; Wills, 1994) occurs spontaneously, as a result of reaction-diffusion coupling, when rapid mixing does not exclude the possibility of sustained

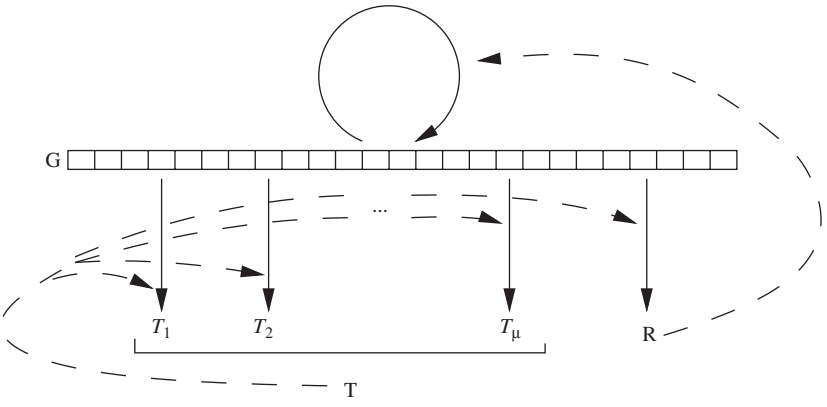

Fig. 2. Gene-replicase–translatase system. Solid lines represent synthetic processes, dashed line represent catalysis. The information in the gene operon G is used to generate the two classes of proteins, the replicase R and the assignment catalysts (translatases) in the class $T \equiv \{T_1, T_2, \ldots, T_\mu\}$. The replicase catalyses the replication of the gene G. The set T of translatases is required to catalyse synthesis of any protein using the information in G.

spatial differentiation. The spatially resolved patches of gene replication and translation that can be seen emerging from initial disorder can be regarded as stochastically structured, thermodynamically driven, self-reproducing fluid automata that use and incorporate information in the manner of an elementary Turing machine. However, these GRT systems are able to maintain their structure and synthesize all of their various macromolecular components without need of the deterministic, system-wide control, which von Neumann (1949) required in his theory of self-reproducing automata.

The GRT system has much in common with the chemoton model of Gánti (2003) which consists of three stoichiometrically coupled autocatalytic cycles of metabolism, membrane production and template replication. The basic idea of the chemoton was first developed in 1971 (Gánti, 2003), but is only now receiving the detailed attention it deserves (Munteanu and Solé, 2006; Munteanu et al., 2007). The chemoton is a fluid automaton that was designed to represent the elementary "unit of life" and the motivation for its construction had much in common with the motivation behind this paper. The chemoton differs from the GRT system in that the coupling of synthetic processes is preserved by confining them all within a membrane whose production and maintainence is internally controlled. In the GRT system, the relative stoichiometry of template replication and replicase production is an emergent rather than a "built-in" property of the model system. The same is true of the spatial association between protein production and template replication. In the chemoton the information template is confined, together with its "products", inside a membrane, whereas in the GRT system the spatial association between an information template and its interpretation has to be maintained through Turing-type reaction-diffusion coupling.

It is clear that the control of membrane production was a very important step in the evolution of cells, which we can take to be modern "units of life" in the sense of Gánti (2003), but the GRT model shows that coding could potentially have emerged before there was controlled compartmentalization of protocells as envisaged in the chemoton model. The GRT model lacks the controlled coupling of template replication and membrane production with basic metabolism (Szathmáry et al., 2005; Szathmáry, 2006), assuming rather that energy-rich monomers are freely available for nucleic acid and protein production. However, these small building-block molecules could only have been available once coupled metabolic cycles of the type that Gánti (2003) incorporated into the chemoton were already operating and such cycles are likely to have become established before the reactions came under the control of genetic information-processing (De Duve, 2005; Morowitz, 1992). Therefore, it seems likely that if GRT-like self-organization was a crucial factor in the origin of life it probably occurred concomitantly with its coupling to the additional metabolic and compartmentalization processes of the chemoton. Such chemoton-like coupling of GRT processes with their essential sources of metabolic energy, as well as the local confinement of the information carrier and functionally catalytic species within a membrane barrier, would have a stabilizing effect on the operation of all the relevant processes and provide a more robust platform, a cell-like structure, for open-ended evolution.

While comparison of the alternative chemoton and GRT fluid automaton models is likely to increase our general understanding of life's origins and requirements, one difference stands out clearly. In the GRT system, information templates function explicitly as symbolic codescripts that are used in computational operations to produce components with the specific structures that are needed for specific functionalities, like controlled metabolism and membrane-building. The chemoton does not have an equivalent system for interpreting templates directly as codescripts and does not feature a "general constructor" of the von Neumann-type. By way of contrast, a functioning GRT system has the capacity to produce potentially novel functional components from any new template sequences it encounters (Füchslin and McCaskill, 2001) and so it could potentially serve as a platform for open-ended evolution.

Fernando and Rowe (2007, 2008) have recently demonstrated Natural Selection in systems that have many of the main features of chemotons. They show that Natural Selection at the system level stabilizes the dynamic production of lower-level molecular replicator units (Fernando and Rowe, 2007) and this can produce quite elaborate networks of reactions inside the system (Fernando and Rowe, 2008). However, the production of lower level entities does not depend on the existence of a fixed source of directly interpretable information, so the internal self-organization observed in these systems does not satisfy the criteria for Informed Generation. As indicated above, a more realistic model of a primitive living system capable of open-ended evolution might be achieved by forging a synthesis of chemoton-type and GRT-type models. In that case the unit of replication would become the whole system, as in the chemoton model, rather than the information template, as in the GRT model, and the protocell system would potentially be able to evolve through both Natural Selection and Informed Generation, the former occurring when a mutant template caused an increase in the rate of reproduction of the system and the latter occurring when the system underwent an essentially irreversible transition in its internal dynamics. Such a transition could occur as a result of new species or chemical cycles becoming incorporated into the coupled, genetically controlled, operation of the system.

The emergence of genetic coding in GRT-type systems, whether through reaction-diffusion coupling in a membrane-free environment or in some form of compartments, represents one of the major transitions in the origin of life (Maynard Smith and Szathmáry, 1995), and arguably the first which allowed open-ended evolution to ensue (Bedau, 1996; Ruiz-Mirazo et al., 2008). A system with genetic coding can act as a general constructor, especially through its ability to synthesize novel enzymatic catalysts, and can therefore provide a platform for the transition from a "proto-metabolism" that is driven by chemical constraints (De Duve, 2005; Morowitz, 1992), to a fully fledged "biochemical metabolism" that is subject to genetic control and can evolve in an open-ended fashion through the addition and elimination of specific anabolic and catabolic pathways. In fact, it is difficult to imagine that increases in the specificity of coding and the specificity of metabolic control, especially of amino acid synthesis, did not evolve in a coupled fashion in which self-organization through Informed Generation and the accumulation of useful information through Natural Selection both played essential roles.

An interesting question arises in respect of the putative RNA World stage of life's origin (Gesteland et al., 2006). During this period the same generic species, RNA, is supposed to have served the roles of both information carrier and functional catalyst, probably in a catalytic network whose size was somewhere between the large, densely-coupled systems described by Kauffman (1986) and the small, generically functional systems of Wills and Henderson (2000). Whatever the case, base-complementary template reproduction would have played the major role in RNA synthesis and the accurate reproduction of RNA molecules with high catalytic specificity for important reactions in the underlying metabolic network would have been essential (Copley et al., 2007; Orgel, 2004). At first sight it would seem adequate to conceive of self-organization in the RNA World in terms of Natural Selection between competing, replicating RNA variants. Adaptive changes in multiple species can be understood in such terms, even when they

happen in several catalytic species simultaneously. However, it is evident that a more sophisticated description is required to explain how multiple species can coexist so that their functions cooperatively contribute to the survival of the entire system. In that regard the principle of Informed Generation is helpful and formalizes the description of processes which are germaine to evolution in the RNA World but which do not come under the rubric of Natural Selection.

When a catalytic RNA (−)-sequence is synthesized from its complementary (+)-strand, the process can be characterized, in computational or automaton-theoretical terms, as an interpretation of the information in the (+)-strand. That being the case, the catalytic network comprising the RNA World can be viewed as an interpreter of the information in the relevant (+)-strand population. This population need not be absolutely disjoint in relation to the catalytic (−)-strand population, which may itself include species whose sequences are derived in part through transcription of more than one (+)-strand. The evolution of such a system, whether it be found in an "open soup" or encapsulated within a protocellular membrane, includes changes that occur as a result of dynamic instabilities rather than any change in the sequences of the species comprising the information-carrying (+)-strand population. This is especially likely to be true of any part of the reaction network where the dynamics are "on the edge of chaos" (Bak, 1997; Kauffman, 1993); where small changes in population numbers can be amplified to have global effects and what would otherwise be short-lived peculiarities can propagate as new features of the system. Such changes in the RNA population would be the result of cooperative rather than competitive processes and a modified system could be said to have generated itself on the basis of the informational possibilities inherent in the extant (+)-strand population.

Thus, although the emergence of genetic coding is the point in prebiotic evolution at which the separate operation of Informed Generation and Natural Selection becomes clearly evident, the two processes can be differentiated in the RNA World to the extent that RNA sequences can be considered to contain information which has a functional interpretation.

7. Relationship between Informed Generation and Natural Selection

The illustration of the mutual operation of Informed Generation and Natural Selection in GRT systems illuminates the distinction and relationship between these two evolutionary processes, which are clearly separated in Fig. 2. Informed Generation occurs within functionally autocatalytic sets T of translatase components whose synthesis requires information stored in G. Natural Selection occurs at the most rudimentary level among individual variants of the genetic component G as they are produced during imperfect replication by the enzyme R. In the GRT system, Informed Generation and Natural Selection occur on comparable timescales and they are tightly coupled, but their effects are clearly separate and if one of the processes were artificially halted the other would continue. Natural selection results in an irreversible change in the population of genetic variants present in the system. Informed Generation results in an irreversible change in the dynamic state of the system which uses the genetic information to generate itself. In neither case is the irreversibility absolute, but we distinguish Natural Selection from essentially reversible genetic drift whereby different neutral variants temporarily dominate the population of information carriers and we distinguish Informed Generation from essentially reversible relaxation processes in which the system fluctuates or moves between readily accessible dynamic states.

The timescale for evolutionary transitions due to Informed Generation is generally much faster than that for Natural Selection. A new genotype takes many generations to become dominant a population, whereas changes in the cooperative dynamics of biological systems occur on the shorter timescales that characterize interactions between individual components: the timescale of well-catalysed reactions in the case of biochemical systems, the timescale of gene expression for molecular biological processes, and less than an individual organism's lifetime for ontogenic processes. Only in the case of ecological systems is the timescale for transitions in the underlying population dynamics often comparable with that for Natural Selection, as in the GRT system. Therefore, it is not surprising that the translatases in the GRT system comprise a highly symbiotic, cross-catalytic set T of individuals T_1, T_2,…, all of which are required for the synthesis of any one of them. The GRT system can support other symbionts, which catalyse, for example, reactions for the production of polymer monomers, and it is also susceptible to invasion by destructive parasites that catalyse any of the $\lambda(\lambda-1)$ assignment reactions not belonging to the coding set.

Informed Generation requires a system to have different dynamic states separated by some sort of transition barrier. In the case of the system described by Eq. (1), the barrier occurs along the transcritical boundary. In more complex biological systems the situation is unlikely ever to be so simple. However, the existence of multiple dynamic states in networks of biological processes has long been a subject of enquiry (Kauffman, 1993). The question of interest to us in this context is whether complex biological processes can have significantly different dynamic states that are separated by some sort of transition barrier such that on the timescale of mutation-selection events the space of possible dynamic states is only sparsely sampled. If that were the case, then the dynamic state of a biological system, the way in which its inherent genetic information is interpreted, could serve as an epigenetic memory of events in its evolutionary history.

This line of reasoning applies to the evolutionary adaptation of individual species, the classical domain of Natural Selection. Consider, for example, the lineages of *E. coli* that have been studied in detail by Lenski et al. (1991) and Lenski and Travisano (1994) over something like 4×10^4 generations. It is reasonable to hypothesize that the "historical contingency" apparent in the most recent findings reported by this group (Blount et al., 2008) is due to Informed Generation. This could be determined by using the strategy described in Fig. 3. The step $(G_1\ P_1)\rightarrow(G_1\ P_1)$ represents an epigenetic transition in the internal dynamics of cells, a spontaneous change in the interpretation of the genetic information they contain, that confers on them the capability of subsequently adapting to their environment through Natural Selection. Whether or not Informed Generation is observable in these experiments with *E. coli*, the process may be demonstrable in populations of yeast that display persistence of epigenetic states across more than one generation (Kauffman et al., 2007). Prion-bearing yeast undergo epigenetic transitions that show the whole gamete of the phenomena, strain mutation and selection, *etc.* (Tessier and Linsquist, 2007), usually associated exclusively with genetic inheritance. Recent experimentation with *C. elegans* (Jordan et al., 2008) suggests that higher organisms may provide similar opportunities for observing Informed Generation.

Let us return briefly to our original analysis of genomes as codescripts. A denucleated zygote can act as a general constructor given a variety of input codescripts. The range of input codescripts that will give rise to new organisms, when supplied to such a denucleated cell, is limited. Appropriate cells can act as general constructors but not universal constructors. It is not known whether an appropriately prepared denucleated human cell can correctly interpret a chimpanzee genome, or *vice versa*, and the

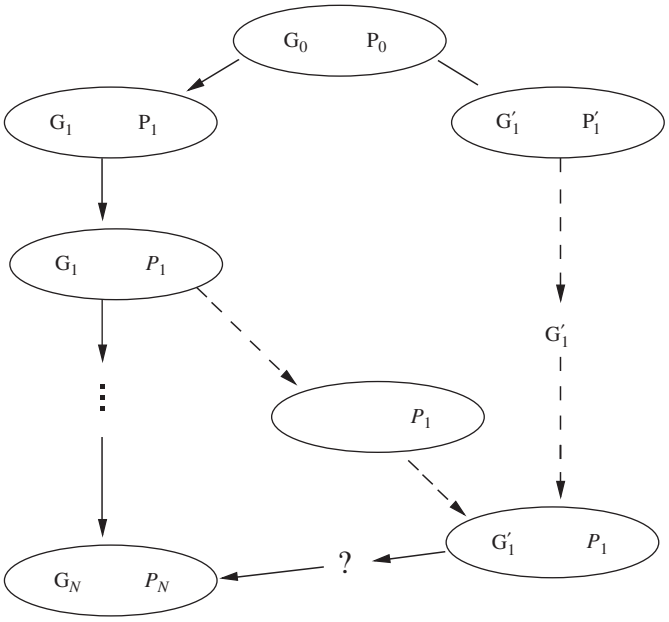

Fig. 3. Experimental demonstration of Informed Generation in evolution. Solid lines represent ordinary life-cycle and reproductive processes; dashed lines represent artificial genetic cloning-type processes. There is a notional separation between the genetic information G and the phenotypic form P of organisms, all of which are considered to be in the same environment and subject to the same selection pressures. A common ancestor (G_0 P_0) has neutrally-related variant offspring, (G_1 P_1) and (G'_1 P'_1), the first but not the second of which undergoes a epigenetic transition to a form (G_1 P_1) that readily adapts, through several stages of Natural Selection, to a superior phenotype (G_N P_N). The hypothetical transition would be an example of Informed Generation and could be shown to have occurred if, when the genome G'_1 was inserted into the P_1 form of the organism, the hybrid (G'_1 P_1) readily evolved into the adapted form (G_N P_N).

capability need not be symmetrical. It is even possible that a human could be produced by a chimpanzee cell when supplied with a genome which does not serve as a codescript for the production of a human when interpreted by a human cell. These things are unknown. What must be considered is whether the particular dynamic state of the denucleated zygote, derived from its particular historical lineage through a series of symmetry-breaking transitions, determines whether a particular genome can be "correctly" interpreted. In other words, are many species boundaries characterized by unique matchings of genomes with the historically contingent dynamic states of matter with which they are uniquely associated?

When Informed Generation is the primary cause of evolutionary change in the lineages of organisms it may leave molecular biological traces, in the same way as the traces of Natural Selection can be seen in nucleic acid phylogenies. Traces of Informed Generation, if it occurs, are likely to show up in the topography and dynamics of intercellular interaction networks, particularly where the states of networks of different kinds, of metabolic, protein-interaction or genetic control, impinge on one another so that the particular state of one set of processes modulates the operation of another set of processes. Some general features of the dynamics of such complex networks, taking into account fluctuations and variations in fluxes, are now beginning to be understood (Argollo de Menezes and Barabási, 2004a, b; Szejka et al., 2008) and the results are being applied to the description of intracellular processes (Luscombe et al., 2004; Nacher et al., 2005; Ochiaia et al., 2007; Samal and Jain, 2008). However, much more direct, detailed and specific information about the operation of networks of intercellular processes will have to be available before a serious attempt can be made to detect historical occurrences of Informed Generation.

One need only contemplate the complexities of transcription and translation, especially the production and action of multitudes of non-coding RNA molecules and their influences on gene expression in eukaryotes (Amaral et al., 2008), to see that the essential meaning of genetic information has enormous potential variability. The existence of alternative morphogenetic pathways for the interpretation of genomes, in all probability involving different overall states of expression of genetic information and different dynamic states of the nucleus, is thought to be the basis of morphogenetic plasticity (West-Eberhard, 2003) and a significant cause of evolution (Newman, 2002). The existence of rarely crossed transitions in the molecular biological dynamics of adult organisms is amply demonstrated in the etiology of sporadic cases of prion diseases (Wills, 1986, 1989). In a complex organism like an animal, are there a multitude of such pathways separated by relatively high transition barriers? If so, then Informed Generation could be a significant cause of evolutionary change.

The apparent redundancy that contributes to the robustness of biological systems at all levels of their structure and functionality (Wagner, 2005) points to the possibility of multiple dynamic states of high complexity in organisms and could be a major factor contributing to their "evolvability" (Kirschner and Gerhart, 1998; Wagner, 2008) as well as their inherent capacity to evolve in an apparently open-ended fashion. However, it is usual to assume that redundant encodings are not open to alternative interpretations of any evolutionary significance. Thus, Toussaint and von Seelen (2007) describe the genotype–phenotype relationship as "a constant map determined by the physics and chemistry of nature that maps any possible genetic system…as a whole to the corresponding organism". The work of Toussaint (2003a) is the most comprehensive work on the evolution of genotype–phenotypes maps through Natural Selection to date, but it does not consider the possibility that genotypes can be mapped onto multiple phenotypes depending on the state of operation of internal dynamics of the system, even when offspring are selected according to the quality of their parents (Toussaint, 2003b).

A great deal of effort has gone into adapting the theory of selection to take into account all sorts of cooperation among individual organisms and genes. It has often been found to be convenient to reformulate the description of the system by defining a new "level" at which Natural Selection acts on a more complex phenotype (Keller, 1999; Michod, 1999; Okasha, 2006; Sakar, 2008). We effectively did this with the GRT system portrayed in Fig. 2, when we considered the effect of confining the processes within a chemoton-type protocell membrane so that the phenotype on which selection was considered to act became the whole composite system, not just the genetic sequences. However, once a new level of selection has been defined, cooperation between the higher-level individuals has to be considered. In the case of GRT systems, cooperation between individual protocells, for example, through the exchange of genetic elements or proteins, may come into play and it may then be necessary to consider an even higher "level of selection" in order to explain the dynamic behavior of the system in terms of Natural Selection. In any case, the possibility of Informed Generation would have been obscured even though it is precisely in the scenario of complex networks of cooperating units that multiple dynamic states of genotype–phenotype mappings are likely to be found.

The same line of reasoning is relevant to the evolution of organisms within ecosystems. When the population dynamics of diverse species are deeply entangled, typically through symbiotic relationships that depend on special phenotypic properties of more than one organism, the selective values of the variants of a gene in one organism may be a function of variant phenotypic

properties which depend on the expression of otherwise unrelated genes in other organisms. Complex biological systems of this sort are typically robust because essential functionalities are realised in apparently degenerate structures or pathways in the network (Wagner, 2005, 2007). The possibility of essentially irreversible functional self-organization taking place in a system of this sort, independent of any genetic or environmental change, cannot be discounted, especially when the network of interactions is as complex as is often found in long-isolated ecological networks. One process that could trigger this sort of self-organization is horizontal gene transfer (Kunin et al., 2005). When a gene is transferred to a new host its expression could have novel phenotypic effects that facilitated a transition in the dynamics of the whole system. Woese (2004) has proposed that self-organization involving horizontal gene transfer was germaine to the emergence of the universal genetic code and it has been suggested by Goldenberg and Woese (2007) that the new field of environmental metagenomics will uncover complex networks of genetic relationships that will bring into doubt the very concept of "species of organisms" on which Darwin's theory of inheritance with variation is based.

8. Discussion and conclusions

We have presented in some detail what is perhaps the simplest possible model of the establishment of a self-organized genotype–phenotype mapping: the emergence of genetic coding from a pre-existing state of disordered information-processing. The simplicity of the model and the rather elementary dynamic behavior of the corresponding system, in which the individual components of the system have only a weak dynamic coupling to one another, should not detract from the generality of the problem we have attempted to formulate. The conclusion we have drawn concerning the true causes of evolutionary change and the incompleteness of Neo-Darwinian explanations of evolution has not been rigorously proven, and nor is it likely to meet with universal approval, but it is offered as a hypothesis worthy of careful consideration. The illustrative example of genetic coding, besides being the simplest, is also probably the most plausible, and it might be conceded, even by Neo-Darwinists, that Informed Generation could have played a role at special periods corresponding to major transitions in evolution (Maynard Smith and Szathmáry, 1995). However, the idea that symmetry-breaking transitions in the dynamics of biological systems play a continuing role in determining their specific characteristics challenges Neo-Darwinian theory at its core. The nature of the controversy concerning the completeness of Neo-Darwinian theory can be seen very easily through discussion of the ideas of De Duve (2005) and Woese (1965, 2004) concerning genetic coding.

De Duve (2005) admits that there is a conundrum concerning the origin of the genetic code through Natural Selection. Mutations that change the specificity of codon to amino acid assignments are expected scarcely to be tolerated by organisms, yet they must have been numerous for the code to have become optimized, as it is, better to withstand the effects of mutation and errors in translation (Wagner, 2005). De Duve argues that optimization of the code occurred very early in the evolution of protein synthesis and he envisages some kind of evolutionary development in which code and functional products were jointly subjected to selection. A process of the kind envisaged by De Duve would have required widespread functional redundancy among the coding assignment catalysts present in the system and such a level of redundancy would in turn have required the presence of multiple variant copies of genes. The processes described by Wills (1993) and Füchslin and McCaskill (2001) satisfy the requirements specified by De Duve but they are not totally amenable, as he would apparently wish, to an explanation in terms of Natural Selection alone.

Over a period spanning decades, Woese (1965, 2004) has espoused a point of view of the evolution of coding and specificity completely different from the orthodox Darwinian view. He says that evolution is basically reticulate, meaning that it is ultimately impossible to trace lineages of discrete biological functions or characteristics through lines of inheritance that never intersect. In contrast to the picture painted by selectionists, Woese (2004) envisages a multiplex evolutionary strategy that had special effect in an early era when horizontal gene transfer between proto-organisms was a more dominant cause of phenotypic change than adaptation as a result of Natural Selection. According to this view the last universal common ancestor (LUCA) of present day organisms is essentially a fictitious entity—the three main kingdoms, archea, bacteria and eukaria, eventually established separate lines of inheritance from an earlier state in which their predecessors shared genes freely.

De Duve (2005) condemns this "communal" picture of the LUCA on the grounds that it lacks the mainspring of evolutionary progression, selection. As he puts it "Collectivism is the antithesis of competition, and competition is the essence of Darwinian selection." and further

> The whole developmental history of protein synthesis and progressive lengthening, of the genetic code, of the enzymes that came to catalyse the first reactions of metabolism, and all of the other improvements that led to the LUCA was…dominated by selection processes that mandated the participation of competing protocells endowed with distinctly more stringent genetic individuality than characterizes the hypothesized progenotes or precells.

De Duve goes so far as to accuse advocates of the communal hypothesis of deliberately omitting to specify the mechanism whereby useful novelties were separated from the useless or deleterious majority, before the "Darwinian threshold", as Woese (2004) calls it, was crossed.

The operation of a general cooperative process, whereby new functional forms of systems bootstrap themselves into existence rather than waiting for novelty to arise in the form of mutated variants, has long been invoked, though not always explicitly, in discussions of the evolution of the AARSs and the emergence of genetic coding (Bedian, 1982; Hoffmann, 1974; Vetsigian et al., 2006; Wills, 1993; Woese, 1965, 2004). The communal hypothesis finds empirical support in phylogenetic analysis of the AARS coding enzymes (Woese et al., 2000; O'Donoghue and Luthey-Schulten, 2003). The main problem seems to have been that the general theoretical significance of the dynamics of coding evolution has never been explicitly enough laid out.

Neo-Darwinists (Dawkins, 1976, 1986; De Duve, 2005; Ruse, 2006) take adaptation through Natural Selection acting on populations of genetic variants to be the sole significant cause of biological evolution. It has been our purpose in enunciating the principle of Informed Generation to go beyond the Neo-Darwinian position and seek to circumscribe a further general process through which biological systems become progressively more differentiated and complex, not as a result of some occult teleology, *élan vital* or Intelligent Design, but because they arrive in new states of structural organization by undergoing irreversible symmetry-breaking transitions in the dynamics of their information-processing that are not necessarily caused by any antecedent genetic change. Rather, we have argued that the genetic information stored in biological systems cannot be regarded as some arbitrary initial condition from which phenotypes can be

generated through a fixed mapping in the manner envisaged by Dawkins (1986, p. 73) when he described how a genetic engineer should, in principle, be able to transverse the mathematical space of genetically determined forms from a pigeon to a dodo. In our view, it is fundamental to the character of biological systems that the meaning of any genomic sequence is defined exquisitely by the historically contingent physical system in which it is found or placed. It represents whatever its surrounding physical system generates from it. The same applies to any physical configuration which serves as information in the sense that it is a member of a functionally defined class (Pattee, 1995; Wills, 2001). For that reason, consideration of the principle of Informed Generation is potentially relevant to the evolutionary phenomenology of complex biological systems and processes ranging from genetic coding and morphogenesis to neural structures and language.

In organisms the execution of the symbolic codescript that their genes represent is carried out by the physical results of that execution. This can only occur once the physical representation of the codescript has been fixed by means of "decoupling" from the dissipative processes occurring within the system (Pattee, 1995; Ruiz-Mirazo et al., 2008). If we take this property of functional reflexivity as a necessary part of any definition of "life", then we can fairly judge that proponents of synthetic biology and artificial life are a long way from reaching their goal, except by pirating extant designs which have resulted from biological evolution. Even the most ambitious plans to produce replicating protocells, whose construction and composition is subject to adaptive Natural Selection (Olasagastia et al., 2007), fail to show how a process even akin to Informed Generation will operate in the proposed systems. Anything that is produced by tinkering with nucleic acid and protein sequences, that perhaps even includes a restyling of the translation apparatus, cannot be regarded as a truly artificial form of life any more than the life of genetically engineered organisms can be regarded as having been created by humans. Engineers who wish to claim to have created a new form of life, especially one as novel as the chemically alternative astrobiologies whose existence has sometimes been suggested (Bains, 2004), will have to design *de novo* some way of conferring on replicating systems a form of autonomous control that is derived from internally stored symbolic information. A much deeper understanding of Informed Generation will have to be achieved before it is possible to embark on such a venture with any real hope of success. In the interim, wise caution should be exercised in activities that significantly alter the mode of expression of phenotypic traits whose effects influence the dynamics of ecological and biogeographical processes. The possible operation of Informed Generation, as well as Natural Selection, should be taken into account in evaluating the ecological and environmental effects of new organisms, including genetically engineered organisms.

Acknowledgments

The author is grateful to the many colleagues and friends who have offered encouragement, to The University of Auckland for the use of office, computing and library facilities, and to the anonymous referees for their comprehensive critical discussion of an earlier version of this paper. The Alexander von Humboldt-Stiftung is also thanked for its continued support. Sven Findeiß kindly drew Figs. 2 and 3.

Endorsement: It is the author's wish that no agency should ever derive military or purely financial benefit from the publication of this paper. Authors who cite this work in support of their own are requested similarly to qualify the availability of their results.

References

Amaral, P.P., Dinger, M.E., Mercer, T.R., Mattick, J.S, 2008. The eukayotic genome as an RNA machine. Science 319, 1787–1788.
Argollo de Menezes, M., Barabási, A.-L., 2004a. Fluctuations in network dynamics. Phys. Rev. Lett. 92, 028701.
Argollo de Menezes, M., Barabási, A.-L., 2004b. Separating internal and external dynamics of complex systems. Phys. Rev. Lett. 93, 068701.
Bains, W., 2004. Many chemistries could be used to build living systems. Astrobiology 4, 137–167.
Bak, P., 1997. How Nature Works: The Science of Self-organized Criticality. Oxford University Press, Oxford.
Bedau, M.A., 1996. The nature of life. In: Boden, M. (Ed.), The Philosophy of Artificial Life. Oxford University Press, Oxford, pp. 332–357.
Bedian, V., 1982. The possible role of assignment catalysts in the origin of the genetic code. Origins Life 12, 181–204.
Behe, M., 2007. The Edge of Evolution: The Search for the Limits of Darwinism. Free Press, New York.
Bird, A., 2007. Perceptions of epigenetics. Nature 447, 396–398.
Blount, Z.D., Borland, C.Z., Lenski, R.E., 2008. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. Proc. Natl. Acad. Sci. USA 105, 7899–7906.
Boogerd, F.C., Bruggeman, F.J., Hofmeyer, J.S., Westerhoff, H.V., 2007. Systems Biology Philosophical Foundations. Elsevier, Amsterdam.
Copley, S.D., Smith, E., Morowitz, H.J., 2007. The origin of the RNA world: co-evolution of genes and metabolism. Bioorg. Chem. 35, 430–443.
Crick, F.H.C., 1958. On protein synthesis. Symp. Soc. Exp. Biol. 12, 138–163.
Crick, F.H.C., 1970. Central dogma of molecular biology. Nature 227, 561–563.
Daley, A.J., Girvin, A., Kauffman, S.A., Wills, P.R., Yamins, D., 2002. Simulation of a chemical autonomous agent. Z. Phys. Chem. 216, 41–49.
Dawkins, R., 1976. The Selfish Gene. Oxford University Press, Oxford.
Dawkins, R., 1986. The Blind Watchmaker. Longman, Harlow.
De Duve, C., 2005. Singularities: Landmarks on the Pathways of Life. Cambridge University Press, Cambridge.
Delarue, M., 2006. An asymmetric underlying rule in the assignment of codons: possible clue to a quick early evolution of the genetic code via successive binary choices. RNA 13, 161–169.
Dembski, W.A., Ruse, M., 2004. Debating Design: From Darwin to DNA. Cambridge University Press, Cambridge.
Edelmann, J.B., Denton, M.J., 2007. The uniqueness of biological self-organization: challenging the Darwinian paradigm. Biol. Philos. 22, 579–601.
Eigen, M., 1971. Self-organization of Matter and the Evolution of Biological Macromolecules. Naturwiss 58, 465–523.
Eigen, M., Schuster, P., 1979. The Hypercycle. Springer, Berlin.
Eigen, M., Winkler-Oswatitsch, R., 1981. Naturwiss 68, 282–292.
Fernando, C., Rowe, J., 2007. Natural selection in chemical evolution. J. Theor. Biol. 247, 152–167.
Fernando, C., Rowe, J., 2008. The origin of autonomous agents by natural selection. Biosystems 91, 355–373.
Füchslin, R.M., McCaskill, J.S., 2001. Evolutionary self-organization of cell-free genetic coding. Proc. Natl. Acad. Sci. USA 98, 9185–9190.
Gánti, T., 2003. The Principles of Life. Oxford University Press, Oxford.
Gatenby, R.A., Frieden, B.R., 2007. Information theory in living systems, methods, applications, and challenges. Bull. Math. Biol. 69, 635–657.
Gesteland, R.F., Cech, T., Atkins, J.F. (Eds.), 2006. The RNA World, third ed. Cold Spring Harbor Laboratory Press.
Glansdorff, P., Prigogine, I., 1971. Thermodynamic Theory of Structure, Stability and Fluctuations. Wiley-Interscience, London.
Goldenberg, N., Woese, C., 2007. Biology's next revolution. Nature 445, 369.
Hoffmann, G.W., 1974. On the origin of the genetic code and the stability of the translation apparatus. J. Mol. Biol. 86, 349–362.
Hodgson, G.M., Knudsen, T., 2008. Information, complexity and generative replication. Biol. Philos. 23, 47–65.
Hopkin, M., 2007. UK set to reverse stance on research with chimeras. Nature Medicine 13, 890–891.
Johannes, F., Colot, V., Jansen, C., 2008. Epigenome dynamics: a quantitative genetics perspective. Nature Rev. Genet. 9, 883–980.
Jordan, I.K., Katz, L.S., Denver, D.R., Streelman, J.T., 2008. Natural selection governs local, but not global, evolutionary gene coexpression networks in *Caenorhabditis elegans*. BMC Syst. Biol. 2, 96.
Kauffman, B.B., Yang, Q., Mettetal, J.T., van Oudenaarden, A., 2007. Heritable stochastic switching revealed by single-cell genealogy. PLoS Biol. 5, 1973–1980.
Kauffman, S.A., 1986. Autocatalytic sets of proteins. J. Theor. Biol. 119, 1–24.
Kauffman, S.A., 1993. The Origins of Order. Oxford University Press, Oxford.
Kauffman, S.A., 1995. At Home in the Universe. Oxford University Press, Oxford.
Kauffman, S.A., 2000. Investigations. Oxford University Press, Oxford.
Keller, L., 1999. Levels of Selection in Evolution. Princeton University Press, Princeton, NJ.
Kirschner, M., Gerhart, J., 1998. Evolvability. Proc. Nat. Acad. Sci. USA 95, 8420–8427.
Kornberg, A., 1987. The universal language. Biotechnology 5, 520.
Kornberg, W. (Ed.), 1983. DNA: The Master Molecule. Avery, New York.
Kunin, V., Goldovsky, L., Darzentas, N., Ouzounis, C.A., 2005. The net of life: reconstructing the microbial phylogenetic network. Genome Res. 15, 954–959.

Lartigue, C., Glass, J.I., Alperovich, N., Pieper, R., Parmar, P.P., Hutchison III, C.A., Smith, H.O.J., Venter, C., 2007. Genome transplantation in bacteria: changing one species to another. Science 317, 632–638.
Lenski, R.E., Rose, M.R., Simpson, S.C., Tadler, S.C., 1991. Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations. Am. Nat. 138, 1315–1341.
Lenski, R.E., Travisano, M., 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. Proc. Nat. Acad. Sci. USA 91, 6808–6814.
Luscombe, N.M., Badu, M.M., Yu, H., Snyder, M., Teichmann, S.A., Gerstein, M., 2004. Genomic analysis of regulatory network dynamics reveals large topological changes. Nature 431, 308–312.
Markowitz, S., Drummond, A., Nieselt, K., Wills, P.R., 2006. Simulation model of prebiotic evolution of genetic coding. In: Rocha, L.M., Bedau, M., Floreano, D., Goldstone, R., Vespignani, A., Yaeger, L. (Eds.), Artificial Life X. MIT Press, Cambridge, pp. 152–157.
Maynard Smith, J., Szathmáry, E., 1995. The Major Transitions in Evolution. Oxford University Press, Oxford.
McMullin, B., 2001. John von Neumann and the evolutionary growth of complexity: looking backward, looking forward. Artif. Life 6, 347–361.
Michod, R.E., 1999. Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality. Princeton University Press, Princeton, NJ.
Morowitz, H.J., 1968. Energy Flow in Biology. Academic Press, New York.
Morowitz, H.J., 1992. Beginnings of Cellular Life: Metabolism Recapitulates Biogenesis. Yale University Press, New Haven.
Munteanu, A., Attolini, C.S.-O., Rasmussen, S., Ziock, H., Solé, R.V., 2007. Generic Darwinian selection in catalytic protocell assemblies. Philos. Trans. R. Soc. B 362, 1847–1855.
Munteanu, A., Solé, R.V., 2006. Phenotypic diversity and chaos in a minimal cell model. J. Theor. Biol. 240, 434–442.
Nacher, J.C., Ochiai, T., Akutsu, T., 2005. On the relation between fluctuation, and scaling-law in gene expression time series from yeast to human. Mod. Phys. Lett. B 19, 1169–1177.
Newman, S.J., 2002. Developmental mechanisms: putting genes in their place. J. Biosci. 27, 97–104.
Nieselt-Struwe, K., Wills, P.R., 1997. The emergence of genetic coding in physical systems. J. Theor. Biol. 187, 1–14.
Ochiaia, T., Nachera, J.C., Akutsua, T., 2007. Emergence of the self-similar property in gene expression dynamics. Phys. A 382, 739–752.
O'Donoghue, P., Luthey-Schulten, Z., 2003. On the evolution of structure in aminoacyl-tRNA synthetases. Microbiol. Mol. Biol. Rev. 67, 550–573.
Olasagastia, F., Moreno, A., Peretó, J., Morán, F., 2007. Energetically plausible model of a self-maintaining protocellular system. Bull. Math. Biol. 69, 1423–1445.
Okasha, S., 2006. Evolution and Levels of Selection. Oxford University Press, Oxford.
Orgel, L.E., 2004. Prebiotic chemistry and the origin of the RNA World. Crit. Rev. Biochem. Mol. Biol. 39, 99–123.
Pattee, H.H., 1995. Evolving self-reference: matter, symbols and semantic closure. Comm. Cogn. Artif. Intell. 12, 9–27.
Prigogine, I., Defay, R., 1954. Chemical Thermodynamics. Everett, D.H., (Trans.), 1962. Longmans, London.
Prigogine, I., Nicolis, G., 1971. Biological order, structure and instabilities. Q. Rev. Biophys. 4, 107–148.
Rodin, S.N., Rodin, A.S., 2006. Partitioning of aminoacyl-tRNA synthetases in two classes could have been encoded in a strand-symmetric RNA world. DNA Cell Biol. 25, 617–626.
Ruiz-Mirazo, K., Umerez, J., Moreno, A., 2008. Enabling conditions for 'open-ended evolution'. Biol. Philos. 23, 67–85.
Ruse, M., 2006. Darwinism and its Discontents. Cambridge University Press, Cambridge.
Sakar, S., 2008. A note on frequency dependence and the levels/units of selection. Biol. Philos. 23, 217–228.
Samal, A., Jain, S., 2008. The regulatory network of *E. coli* metabolism as a Boolean dynamical system exhibits both homeostasis and flexibility of response. BMC Syst. Biol. 2, 21.
Schneider, E.D., Kay, J.J., 1994. Life as a manifestation of the second law of thermodynamics. Math. Comp. Modelling 19, 25–48.
Schrödinger, E., 1944. What is Life? Cambridge University Press, Cambridge.
Smith, E., 2008a. Thermodynamics of natural selection I: energy flow and the limits on organization. J. Theor. Biol. 252, 185–197.
Smith, E., 2008b. Thermodynamics of natural selection II: chemical Carnot cycles. J. Theor. Biol. 252, 198–212.
Smith, E., 2008c. Thermodynamics of natural selection III: Landauer's principle in computation and chemistry. J. Theor. Biol. 252, 213–220.
Szathmáry, E., Santos, M., Fernando, C., 2005. Evolutionary potential and requirements for minimal protocells. Top. Curr. Chem. 259, 167–211.
Szathmáry, E., 2006. Coevolution of metabolic networks and membranes: the scenario of progressive sequestration. Philos. Trans. R. Soc. London B 362, 1781–1787.
Szejka, A., Mihaljev, T., Drossel, B., 2008. The phase diagram of random threshold networks. New J. Phys. 10, 063009.
Tessier, P.M., Linsquist, S., 2007. Prion recognition elements govern nucleation, strain specificity and species barriers. Nature 447, 556–562.
Timoféeff-Ressovsky, N.W., Zimmer, K.G., Delbrück, M., 1935. Über die Natur der Genmutation und der Genstruktur. Nachr. Ges. Wiss. Göttingen, Fachgruppe VI N. F 1, 189–245.
Toussaint, M., 2003a. The evolution of genetic representations and modular adaptation. Dissertation zur Erlangung des Grades eines Doktors der Naturwissenschaften in der Fakultät für Physik und Astronomie der Ruhr-Universität Bochum.
Toussaint, M., 2003b. On the evolution of phenotypic exploration distributions. In: De Jong, K.A., Poli, R., Rowe J. (Eds.), Foundations of Genetic Algorithms 2003 (FOGA 7) Morgan Kaufmann, pp. 169–182.
Toussaint, M., von Seelen, W., 2007. Complex adaptation and system structure. BioSystems 90, 769–782.
Toussaint, O., Schneider, E.D., 1998. The thermodynamics and evolution of complexity in biological systems. Comp. Biochem. Physiol. A 120, 3–9.
Turing, A.M., 1936. On computable numbers, with an application to the Entscheidungsproblem. Proc. London Math. Soc. 42, 230–265.
Turing, A.M., 1952. The chemical basis of morphogenesis. Philos. Trans. R. Soc. B (London) 237, 37–72.
Vetsigian, K., Woese, C., Goldenfeld, N., 2006. Collective evolution and the genetic code. Proc. Natl. Acad. Sci. USA 103, 10696–10701.
Von Neumann, J., 1949. In: Burks, A.W. (Ed.), Theory of Self-Reproducing Automata. University of Illinois Press, Urbana, pp. 29–87.
Wagner, A., 2005. Robustness and Evolvability in Living Systems. Princeton University Press, Princeton, NJ.
Wagner, A., 2007. Gene networks and natural selection: Is there a network biology? In: Pagel, M., Pomiankowski, A. (Eds.), Evolutionary Genomics and Proteomics. Sinauer Associates, Sunderland, MA.
Wagner, A., 2008. Robustness and evolvability: a paradox resolved. Proc. R. Soc. B 275, 91–100.
Watson, J.D., Crick, F.H.C., 1953. Genetical implications of the structure of deoxyribonucleic acid. Nature 171, 964–967.
Weber, B.H., Depew, D.J., Dyke, C., Salthe, S.N., Schneider, E.D., Ulanowicz, R.E., Wicken, J.S., 1989. Evolution in thermodynamic perspective: an ecological approach. Biol. Philos. 4, 373–405.
West-Eberhard, M.J., 2003. Developmental Plasticity and Evolution. Oxford University Press, Oxford.
Wills, P.R., 1986. Scrapie, ribosomal proteins and biological information. J. Theor. Biol. 122, 157–178.
Wills, P.R., 1989. Quantitative constraints on mechanisms of replication for information-carrying prions. Prog. Clin. Biol. Res. 317, 669–677.
Wills, P.R., 1993. Self-organisation of genetic coding. J. Theor. Biol. 162, 267–287.
Wills, P.R., 1994. Does information acquire meaning naturally? Ber. Bunsenges. Phys. Chem. 98, 1129–1134.
Wills, P.R., 2001. Autocatalysis, information and coding. BioSystems 60, 49–57.
Wills, P.R., 2004. Stepwise evolution of molecular biological coding. In: Pollack, J., Bedau, M., Husbands, P., Ikegami, T., Watson, R.A. (Eds.), Artificial Life IX. MIT Press, Cambridge, pp. 51–56.
Wills, P.R., Marshall, J.M., Smith, P.J., 2004. Genetic information and self-organised criticality. Europhys. Lett. 68, 901–907.
Wills, P.R., Henderson, L., 2000. Self-organisation and information-carrying capacity of collectively autocatalytic sets of polymers: ligation systems. In: Bar-Yam, Y. (Ed.), Unifying Themes in Complex Systems: Proceedings of the First International Conference on Complex Systems (Perseus Books) pp. 613–623.
Woese, C.R., 1965. On the evolution of the genetic code. Proc. Natl. Acad. Sci. USA. 54, 1546–1552.
Woese, C.R., 2004. A new biology for a new century. Microbiol. Mol. Biol. Rev. 68, 173–186.
Woese, C.R., Olsen, G.J., Ibba, M., Soll, D., 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiol. Mol. Biol. Rev. 64, 202–236.
Wiley, E.O., Brooks, D.R., 1982. Victims of history: a non-equilibrium approach to evolution. Syst. Zool. 31, 1–24.