BioSystems 142-143 (2016) 1–8

Contents lists available at ScienceDirect

BioSystems

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

Review Article

Computational power and generative capacity of genetic systems

Abir U. Igamberdiev[a,*], Nikita E. Shklovskiy-Kordi[b]

[a] Department of Biology, Memorial University of Newfoundland, St. John's, NL A1B 3X9, Canada
[b] National Research Center for Hematology, Moscow 125167, Russian Federation

ARTICLE INFO

Article history:
Received 8 December 2015
Received in revised form 25 January 2016
Accepted 27 January 2016
Available online 29 January 2016

Keywords:
Biological evolution
Generativity
Genetic code
Language
Molecular computation
Physical carriers of mathematical operations

ABSTRACT

Semiotic characteristics of genetic sequences are based on the general principles of linguistics formulated by Ferdinand de Saussure, such as the arbitrariness of sign and the linear nature of the signifier. Besides these semiotic features that are attributable to the basic structure of the genetic code, the principle of generativity of genetic language is important for understanding biological transformations. The problem of generativity in genetic systems arises to a possibility of different interpretations of genetic texts, and corresponds to what Alexander von Humboldt called "the infinite use of finite means". These interpretations appear in the individual development as the spatiotemporal sequences of realizations of different textual meanings, as well as the emergence of hyper-textual statements about the text itself, which underlies the process of biological evolution. These interpretations are accomplished at the level of the readout of genetic texts by the structures defined by Efim Liberman as "the molecular computer of cell", which includes DNA, RNA and the corresponding enzymes operating with molecular addresses. The molecular computer performs physically manifested mathematical operations and possesses both reading and writing capacities. Generativity paradoxically resides in the biological computational system as a possibility to incorporate meta-statements about the system, and thus establishes the internal capacity for its evolution.

Contents

1. Introduction. Genetic language and its operation

It was considered throughout the human history that all known texts were written by humans. However in 1953 it was established that the hereditary information in cells is presented as a DNA code sequence of four letters-nucleotides. The matrix principle of heredity was formulated as a scientific hypothesis which included the double helix idea by Nikolai Koltsov (1927, in more detail 1936), while the first model of the genetic code was suggested by George Gamow (1954). The information in DNA, similarly as in human texts, is presented as a linear order of letters (symbols). Like in books where some information is attributed to pictures and to book shape, the secondary and higher structures of DNA also bear certain information; however for simplicity we usually consider primarily the linear information encoded by four nucleotides. For the reading device, all same signs (letters) are identical.

The genetic language is based on the rigid structure of genetic code which is universal for all organisms on Earth except of minor deviations in mitochondria and several lower organisms.

* Corresponding author.
E-mail addresses: igamberdiev@mun.ca (A.U. Igamberdiev), nikitashk@gmail.com (N.E. Shklovskiy-Kordi).

http://dx.doi.org/10.1016/j.biosystems.2016.01.003

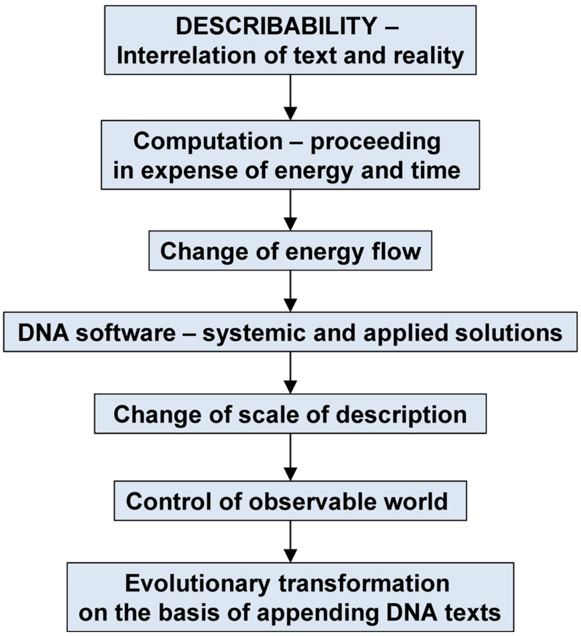

Fig. 1. General scheme showing the process of molecular computation, which is based on reading and interpreting of genetic texts in the expense of energy resulting in the coordination of biosystem's functions in the observable world and in the evolution via writing new genetic texts.

The genetic code possesses all basic characteristics of the linguistic semiotic system that were formulated by Saussure (1911). One of such characteristics is the arbitrariness of sign, expressed in the fact that the correspondence of triplets of the genetic code to amino acids is not based on any physical similarity and could be different (which is observed at a limited degree in the mitochondrial code). Another important characteristic feature is the linear nature of the signifier, which fully corresponds to the linearity of human alphabetic script. Any language has the syntactic, pragmatic and semantic aspects. In the genetic language, its syntax is represented by the combinatorial rules of interactions between nucleotides; the pragmatics is realized via the context-dependent transcription, while the semantics appears as the function of the transcription products (Witzany, 2016).

For language operation, a device is needed which can read it, and a subject ("self") which can perceive it. Originally the concept of such reading device in biological cells was introduced by Efim Liberman (1972) and developed in his subsequent papers (Vaintsvaig and Liberman, 1973; Liberman, 1978, 1979, 1983, 1989, 1997; Liberman et al., 1998, 2001). According to this concept, DNA is not a compendium of genes but a molecular text representing a program for the molecular computer of the cell (MCC). The necessity of MCC arises from the fact that that no natural code codes itself but needs some competent agents that act on this code (Witzany, 2016). These agents form a machine-like structure which function is to decode genetic statements. MCC is a system consisting of DNA, RNA and proteins addressing them. Its operators cut and crosslink the molecules at certain places determined by the program written on DNA. The enzymatic activity necessary for this function is associated with corresponding protein enzymes and with the enzymatic activity of RNA.

The scheme of MCC operation and of its transformational and evolutionary consequences is presented in Fig. 1. It shows that the process of molecular computation is based on reading and interpreting of genetic texts in the expense of energy resulting in the coordination of biosystem's functions with the events of the observable world. This coordination possesses generative properties and results in the open process of evolution via writing new genetic texts.

The process of natural calculation performed by MCC uses the Brownian movement of molecular structures. Multiple searches of addressed molecular operations of words-molecules are possible due to heat motion (Liberman, 1972). The certainty of interpretation of the sequence and its time-irreversible reading are defined, according to Liberman, by the loss of energy for calculation which represents the price of action and determines the physical limitation of computation (counting). For the meaning of signs, their chemical matter is important only for the process of readout by MCC, while the information by itself is static and independent from its carrier. In the physical reality of the cell, the genetic information receives its interpretation, and its material carriers (nucleotides) exhibit their real physical properties, in particular in the secondary, tertiary and quaternary DNA structures up to chromosomes that possess the species-specific architecture.

The genetic language expresses the living self (corresponding to the "quantum regulator" in Liberman's concept, see Liberman, 1983), while the human language expresses the conscious self. It is widely accepted that the process of computation as well as the process of fixation of the result of measurement are realized by the conscious observer. However this statement may not be valid even at the level of human brain. The experiments of Benjamin Libet (1985) show that consciousness appears as an epiphenomenon of brain states, and the reduction of uncertainty in volition actions takes place at the level of the unconscious before any realization of the awareness. The important examples of the unconscious choice have been established for animal behaviour in the studies of Gunji group (Fukano et al., 2004; Migita et al., 2005).

Gamkrelidze (1989) analyzed the structural isomorphism between the two codes (genetic and linguistic) in relation to the existence of two approaches to explain such isomorphism, one arising to Jacob (1977) and the other to Jakobson (1971). According to Francois Jacob, this isomorphism appeared as a result of the structural coincidence between the two systems bearing similar information functions, while Roman Jakobson derived this isomorphism from the phylogenetic construction of the linguistic code on the basis of the structural principles of the genetic code. Such opposition can be resolved via understanding of the universal principles that are reproduced independently for different digital semiotic systems and arising to the combinatorial rules that were initially formulated in the Chinese "I Ching" book (Petoukhov, 2006, 2016).

Both the genetic and the human alphabetic languages are based on the linear representation, while the interactions based on the secondary and tertiary structure of nucleic acids and proteins resemble the Chinese and Egyptian hieroglyphics (Doerfler, 1982). Ratner (1993) analyzed the genetic language as a collection of rules and regularities of encoding the genetic information in the course of operation of the genetic texts. The genetic language possesses the alphabet, the grammar, the system of punctuation, and semantics. Searls (1997, 2002) pointed the universality of resemblance between the genetic and the human language and emphasized the linguistic basis of the basic methods and approaches of bioinformatics. As Eigen and Winkler (1981, p. 282) stated, "Nature's two great evolutionary processes – the development of all forms of life and the evolution of the intellect – both depended on the existence of language".

2. Generativity and evolution

A key feature of language is generativity, which means a possibility to create an infinite number of meaningful sentences from the fixed and finite number of basic elements (sounds, letters, and words). Alexander von Humboldt (1792, cited by Chomsky, 1966) characterized this feature as "the infinite use of finite means".