

Recorded Versus Organic Memory: Interaction of Two Worlds as Demonstrated by the Chromatin Dynamics

Anton Markoš · Jana Švorcová

Received: 26 November 2008 / Accepted: 21 January 2009 /

Published online: 29 April 2009

© Springer Science + Business Media B.V. 2009

Abstract The “histone code” conjecture of gene regulation is our point of departure for analyzing the interplay between the (quasi)digital script in nucleic acids and proteins on the one hand and the body on the other, between the recorded and organic memory. We argue that the cell’s ability to encode its states into strings of “characters” dramatically enhances the capacity of encoding its experience (organic memory). Finally, we present our concept of interaction between the natural (bodily) world, and the transcendental realm of the digital codes.

Keywords Chromatin structure and dynamics · Natural and digital worlds · Creation and interpretation of codes

Introduction

Two different modes of semiosis can be distinguished according to Marcello Barbieri (2007):

1. In the frame of a system of “organic codes”, semiosis comprises the trinity of sign, meaning, and code (and associated operations like coding and decoding, or deriving meaning). Codes have been enacted in history, and therefore they are *not* deducible from the laws of physics. However, once a code system comes into existence, it behaves deterministically and is fully comprehensible by the standard approaches used in (natural) science. The only enigma that evades scientific understanding, then, is the actual, contingent process, or event, that gave origin to the coding rules. This system of organic codes knows no interpretation: codes themselves are context-free; essentially, their *meaning* is a

Anton Markoš and Jana Švorcová contributed equally to this communication.

A. Markoš (✉) · J. Švorcová

Department of Philosophy and History of Science, Faculty of Sciences, Charles University, Viničná 7,
CZ 128 44 Praha 2, Czechia
e-mail: markos@natur.cuni.cz

- question of decoding. Evolution of the system proceeds via adding new organic codes, and/or by building nested hierarchies thereof.
2. In systems capable of interpretation, i.e., hermeneutic systems, *meaning* involves contexts, history, memory, learning, and experience. Such systems are not directly accessible to scientific scrutiny; their study belongs mainly to the realm of humanities. The prototype here is the natural (vernacular) language, as well as human culture and history.

Heated discussions have debated for decades whether the phenomenon of life is fully comprehensible by scientific objective standards (point 1) — i.e., whether it can be accommodated and handled by biology, or whether life possesses also characteristics which go beyond the reach of science (point 2).

Biosemioticians who work on *scientifically* tangible problems assume, often silently, that the first possibility holds true, and thus that all phenomena of life can ultimately be reduced to semiosis of the first type. However, the second possibility may be no less justified. In Markoš et al. (2007, 2009) we proposed the notion of “Barbieri’s platform” for a level of life organization where both approaches may meet. The platform can be climbed “from below”, by assembling parts according to established organic codes (or grammars), or touched down from the heights of holistic sciences such as systems biology or hermeneutics. At this level, both tendencies are in equilibrium, or have equal rights, so to speak. The question remains whether such a platform is a realistic model or not. Here we attempt to demonstrate, with the example of chromatin dynamics, that it is indeed justified.

Codes Belonging to the Realm of Science

Two examples of scientifically manageable code systems are given in Barbieri (2003). The first is protein synthesis, where mRNA and polypeptide are not connected causally (e.g. by chemical correspondence); they correspond only via an established code, which is implemented by a set of adaptor molecules, the aminoacyl—tRNAs. Such machinery can be (at least for some proteins) assembled in a test tube; yet this is not a classical chemical reaction, as its components came into existence as a result of long evolutionary tinkering filtered by natural selection.

The second example of a coding system is provided by the rules that connect written and spoken word in any given language. Here, the alleged adaptors reside in the mind of the person who is fluent and literate in that particular language.¹

The existence of a plethora of analogous coding systems has been proposed in living beings. The above examples introduce however a novel problem. The protein synthesis can be taken — at least at a first approach — as synchronous, i.e., performed by predetermined machinery: the decoding rules borne by adaptors are constant; their whole set is known, and present, from the beginning of the task; and

¹ Today, they can be realized also by computer programs transferring string of letters into sounds and even vice versa; such programs, of course, were created, and mirror what existed in the creators’ minds.

the set of rules is manageable. In processes like cell differentiation or development (not to speak of language), however, the complexity grows with time: novel rules and novel adaptors appear, and the neat model becomes cumbersome, even unusable. Turner (2007) touched the point in comparing three systems containing signs, adaptors and outputs: (i) The red/green traffic signs will lead in any system acquainted with the traffic code (a driver, a schoolchild, a computer with appropriate sensors) to two possible outputs — stop or go. (ii) In protein synthesis, 64 codons constituting innumerable possible strings will be translated by some 45 adaptors into 21 outputs, also combined into innumerable incidences. (iii) But how to manage, asks Turner, the histone code in chromatin (see below), or processes involved in development, when the number of elementary inputs, outputs, and adaptors may go to hundreds or even millions, and all three sets may change in time? Why should we speak about coding when we can neither write down the table of the code nor quantify the number of components, and the rules of the system? How can any system memorize so many commands and shortcuts, and — even more important — how can it consult them in real time? Yet, attempts have been made to prove such a predetermined, synchronous superstructure of nested, overlapping codes in cells (see, e.g. Trifonov 2008; Popov et al. 1996).

Codes Enacted *En Passant*

Barbieri gives an answer by introducing the concept of *organic memory*, and illustrates it on the early embryonic development. Development starts with a hardwired (coded) program and proceed mechanistically up to the “platform” — the stage of the *phylotype*. Besides, the developing germ is endowed also with an *organic memory*, which may be taken for empty at the beginning. It takes inputs from the developmental program and from the environment, and ensures coordination of species-specific processes, thus increasing enormously the amount of information and interpretative rules, as compared with those available at the moment of fertilization. This “bootstrapping” between the program and the developing memory will — at the phylotype stage — lead to a takeover of the affairs by the memory.

Barbieri (2003) gives no clues as to the “reification” of the memory (or its parts). For our purpose here, we can identify it with somewhat vague concepts like “fine tuning”, “setup” of the “living state”. Actually, we shall take the memory for the very realm of semiotic and hermeneutic processes, and from *here* we shall try landing on the Barbieri’s platform “from above”: from the realm of historically established “cultural” conventions, which lend the mechanically erected platform a much subtler, ornamented, baroque, species-specific and individuum-specific pattern. It follows that the conventions established should be taken for fuzzy and malleable, language-like: it is here where meaning is generated, and the process cannot be — as in the case of codes — executed by machine-like contraptions, automatons. In Markoš et al. (2007) we argue that the organic memory is never empty; it comes with the bodies of the germ cells (and is, at least in part, contributed by the mother) and is really responsible for procuring the individual pattern (body) subject to natural selection.

The model of the “platform” can serve as a good heuristic tool for our understanding of life. In short: codes are hardwired, whereas semiotic processes “from above” mollify and adjust their impact, and extract a meaning (sensu 2) of the whole process. Codes and memories work *always* in tandem; *we never encounter a “read only” code, or an empty memory matrix.* Only in this way the massive parallel processing that is taking place in the body can be managed in real time. Let us now approach, with this concept in mind, the model case of chromatin.

The Structure of Chromatin and the Histone Code

Nucleosome

DNA in the eukaryotic nucleus is folded into a higher-order structure — the nucleosome (Fig. 1); and a major role in this folding is played by proteins known

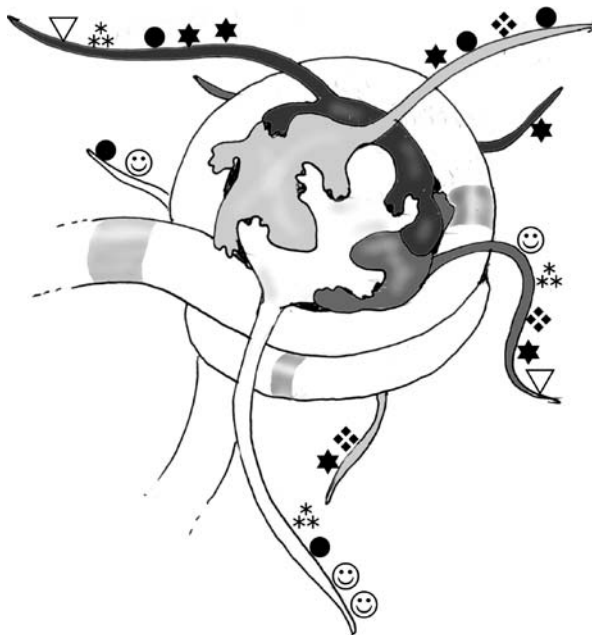


Fig. 1 A cartoon of the nucleosome (approximately in-scale). The “hose” represents DNA, wrapped to create a nest containing a “brood” of 8 histone proteins of four kinds. The pattern of marks on protruded tails of the histones is produced by a “zoo” of hundreds of protein species making up the chromatin ecosystem; other proteins can bite the marks off, or sink their teeth into them and remain stuck. The bodies of all proteins are, or may be, also decorated by similar marks. The result comprises the organic memory at this level of description: a huge, dynamic multiprotein complex around each nest (not shown). Three highlighted segments of DNA represent motifs attractive for yet other proteins; the assembly of such proteins will decide whether that particular area of DNA will be transcribed or not. Such an assembly is a part of, and its composition highly depends upon, the overall meta-assembly of the intracellular “ecosystem” of molecules

as histones. The nucleosome particle consists of 147 base pairs of DNA wrapped around a histone octamer core, comprising pairs of histones H2A, H2B, H3, and H4.

Each histone complex is linked (zipped) to DNA through zillions of so-called weak interactions.² The contact of histones with DNA is independent of the DNA sequence, i.e., any part of the long, linear molecule can be wrapped onto the structure. At this level of description, the main task of histone proteins consists in stabilizing the DNA molecule (6 billions of base pairs in a human nucleus) by condensing it into millions more easily manageable packages.

Besides this elementary “zipping” role, nucleosomes play a very important role in short-time and long-time regulation of gene expression, by controlling the recruitment of the protein machinery required for the process. First, the very arrangement of nucleosome units along the DNA strand decides accessibility of particular elements for that machinery; second, the fine tuning of nucleosome shape restricts or enables selective accessibility of the particular DNA segment for higher-order regulatory systems (for a review see Allis et al. 2007). This second function of the nucleosomes is attained by putting bookmarks (“bar codes”) on histone tails, which stick out from the octamere core, and are thus available to inspection (or rather palpation) from outside. By their *reversible* chemical modifications, multifarious patterns will be induced in the histone backbone; in turn, the backbone will change its competence as to the shaping of attached DNA, and the docking of regulatory protein(s) present in the nucleus. Hence, selective combinatorics of histone modifications has far-reaching effects resulting in cell differentiation, tissue modifications and organogenesis, and in maintaining such states in a cell (or in a cell lineage) for long periods of time. Such modifications — epimutations — are responsible also for the highly flexible and dynamic responses of chromatin to external cues.

Epimutations

It is well known that a mutation in DNA resulting in a haphazard replacement of a single amino acid may introduce havoc in the resulting protein structure/function. Yet, in a strictly controlled manner, such replacements (like the above-mentioned “bar codes”) are widespread at the level of “adult” protein molecules. The difference between both kinds of mutations is as follows:

- (i) Nascent proteins are synthesized from a constant set of 20 amino acids according to a sequence code hardwired in DNA. If a mutation occurs at the level of the code, it cannot be detected and repaired, and it will yield misconstrued or truncated proteins forever.
- (ii) Epimutations, in contrast, will change amino acid composition by *reversible* chemical modification of amino acid residues already assembled in translation. The effect is fully comparable to that of mutation, as can be seen in the two examples shown in Fig. 2. Thus, phosphoserine is an amino

² For non-biologists: “Weak”, because they can be broken by mild treatment, like change of acidity, or elevated temperature.

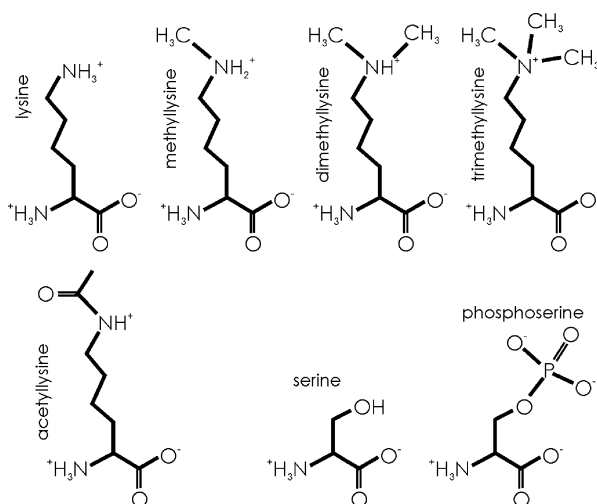


Fig. 2 Modifications (epimutations) of amino acids, as exemplified by lysine and its methylated and acetylated forms, and serine and its phosphorylated form. All such “mutations” of amino acids in a protein are reversible, provided that the cell expresses enzymes responsible for particular transformations

acid with properties drastically different from original serine (S); the same with many derivatives of lysine (K) residue, which can give rise to acetyllysine, mono-, di-, or trimethyllysine, hydroxylysine, or even can be coupled to whole *proteins*.

In a similar manner, almost all amino acid residues in a protein can be derivatized, often each sister copy (product of the same mRNA) differently: the uniform population of nascent proteins will soon give rise to a plethora (even dozens) of proteins differing in shape and hence performance. However, the instructions to introduce such specific modifications are not stored in any coded string. Epimutations result from the collective action of a protein “ecosystem” in any given compartment — the nucleus in case of histones.

Such modification marks may facilitate (or prevent) the binding of various classes of proteins, “readers of the code”; these in turn can recruit whole cascades of proteins bound to such already bound “adaptors”, to become yet higher-order adaptors, etc. Their lifetime may be very short (seconds in case of the so called signal particles) or the process leads to the establishment of big irreversible complexes that literally “immure” long sections of DNA, even whole chromosomes.

Examples of Histone Marks

Most histone modifications known today were discovered in budding yeast (Liu et al. 2005), but information available from mouse or human models show that they remain very *conservative across divergent organismal lineages* (Bernstein et al. 2006). Let us illustrate the phenomenon on the protein histone 3, whose “tail” of 29

amino acids (out of about 220) serves as a board accommodating labels. The nascent sequence reads as follows:

. . . . 5 0 5 0 5
ARTKQTARKSTGGKAPRKQLATKAARKSA

Below we give several hypothetical examples how such reversible chemical modifications of the sequence may look like:

AρTκQTARασTGGKAPRKQLATKAARμSA
ARTλQTARKSTGGKAπRKQLATKAARKSA
ARTμQTARλSTGGKAPRKQLAτKAARλSA
ARTKQTARKσTGGKAPρKQLATKAARKSA
ARTαQTARηSTGGKAωRKQLATKAARμSA

(α — acetyl lysine, η — hydroxy lysine, κ — monomethyl lysine, λ — dimethyl lysine, μ — trimethyl lysine, ρ — methyl arginine, π — hydroxy proline, ω — proline isomer, τ — phospho threonine, σ — phospho serine)

It is to be noted that:

1. Each such modification requires a specific enzyme, which, in addition, may be site-specific. This means that trimethylation of lysine 4 (K4) is carried out in three steps, by three specific methyl transferases; performing the same task with K27 may require a different battery of transferases. The resulting pattern of modifications thus depends on what enzymes are present in the set of proteins present at the time — it is “agreed” within such an “ecosystem” of proteins: negotiated, not encoded.
2. To keep modifications reversible, each modifying enzyme must be accompanied by enzymes with a reverse action (e.g., methylases, removing methyl groups).
3. The changes of shape brought to the protein by such modifications give rise to different shapes, hence very specific antibodies can be raised against each modification. This enables researchers to detect such variations across vast expanses of chromatin, and draw conclusion as to the state of chromatin in different parts of the genome, or in different cells.

We now approach an extremely important part of our explanation: the “bar codes”, the “diacritics” of modified amino acid “characters”, are reversible, i.e. they can be erased, edited, or rewritten. Protein epimutations can appear (and disappear) on chromatin *within minutes* upon arrival of a specific signal. Proof of the biological consequence depending on the individual combination of modification is, however, not always easy to provide and is often based on correlation: proving causality for a modification involves demonstrating that catalytic activity of the enzyme that mediates that modification is necessary for the biological response (Kouzarides 2007).

Examples of Epimutations³

Probably the best-known histone modification is methylation of lysine #9 on histone 3 (H3K9me);⁴ in animals it will initiate a cascade of events resulting in attachment of dozens of proteins of so-called *Polycomb* group; this leads to tighter condensation and silencing of that region of DNA fiber (Giannis et al. 2005). Modulations like H3K9 and H3K27 are responsible for silencing of the chromosome X in mammalian females.

Three lysine methylations are linked with repression of transcription (silencing): H3K9, H3K27 and H4K20. But methylation H3K9 can be found also in transcriptionally active chromatin; in context with H3K4 and H4K20 it helps holding the chromatin active for transcription by binding of the chromatin remodeling complex (Margueron et al. 2005).

In general, three methylation sites on histones are implicated in activation of transcription: methylation of lysine #4 on histone 3 (H3K4), H3K36 and H3K79. H3K4me and H3K36me play also role in transcriptional elongation. However in budding yeast another exception can be found: methylation H3K4 is involved in DNA silencing (Bryk et al. 2002). The location of such a modification is also important: H3K36me has a positive effect only when on the coding region, and a negative one on the promoter (Vakoc et al. 2005).

Existing modifications may promote further labeling: thus phosphorylation of H3S10 facilitates H3K9 and H3K14 acetylation and thereby inhibits H3K9 methylation (Giannis et al. 2005; Kouzarides 2007). H4K20 methylation and H4K16 acetylation were found to preclude each other (Allis et al. 2007). Trimethylation of H3K4 requires ubiquitylation of H2BK123 and reversely deubiquitylation of H2BK123 leads to trimethylation of H3K27 (Schuettengruber et al. 2007).

Binding of a protein could also be disrupted by a subsequent histone modification: H3K14 acetylation accompanied by H3S10 phosphorylation will dissociate *Polycomb* group proteins from methylated H3K9 (Fischle et al. 2005).

Identical modifications, even in the same region of chromatin, may not necessarily lead to the same output: their context and position is crucial here.

Histone modifications represent part of the cellular epigenetic memory, i.e., information that must be built up in ontogeny and is heritable through the cell lineages. Many cellular phenotypes are transmitted and maintained in this way, including genomic imprinting, X chromosome inactivation, heterochromatin formation or gene silencing, or the expression state of Hox genes involved in

³ Non-interested reader can skip this section and proceed to the next part (Two worlds).

⁴ Glossary: H for histone — H3 in our example; K is the abbreviation for amino acid lysine — K9 means lysine residue in position 9 of the histone protein chain (similarly, e.g., R17 is arginine in position 17); “me” is an abbreviation for a methyl group; if more methyl groups are attached, the number will indicate how many, e.g. me2. Examples: H2BK123uq — histone 2B ubiquitylated at lysine in position 123; H3S10p — histone 3 phosphorylated on serine residue in position 10.

specifying cell identity along the axes of segmented animals (Kouzarides 2007; Schuettengruber et al. 2007; Costa 2008).

How Does the Memory Manipulate the Code?

Bernstein et al. (2006) detected the so-called bivalent DNA domains in mouse embryonic stem cells. Such domains (residing mainly in the promoter region of given genes) contain both activating and repressive modifications of histones (for example H3K27me3 is implicated in chromatin silencing, whereas H3K4me3 in its activation). Transcription factors that control certain differentiation processes are in this manner kept in a poised, low-level expression within embryonic stem cells. When cells differentiate, the bivalent domains tend to switch either towards the repressive H3K27 state, or to the activating H3K4 modification.

Also in the case of Hox genes the active state is typically distinguished by continuous stretch of di- or trimethylation of H3K4 in the surrounding chromatin, whereas silent genes are marked by trimethylation of H3K27 (Swigut and Wysocka 2007). Specific demethylases are involved in switch from silencing to activating marks during activating of Hox genes expression (Lan et al. 2007). By contrast, monovalent domains (promoters with mark H3K4me3) are associated with the so called “housekeeping” genes (genes of basic functions-replication, transcription, metabolism) (Mikkelsen et al. 2007).

Polycomb and *Trithorax* Complexes and the Cell Memory

The *Polycomb* and *Trithorax* group of proteins belong among key regulators in defining cell identity in eukaryotes. *Polycomb* genes encode a group DNA-binding proteins, histone modifying enzymes, or chromatin repressive factors with affinity for H3K27me3 (Kingston and Tamkun 2007). H3K27 trimethylation is often distributed over large chromosomal domains, sometimes covering hundreds of kilobases (i.e. thousands of nucleosomes in a row), which might provide the basis for epigenetic inheritance of *Polycomb*-dependent silencing during cell division (Schuettengruber et al. 2007). These proteins control the silencing of target genes (e.g., chromosome X inactivation, repressing the Hox genes activity). One group of *Polycomb* proteins components has a histone-modifying function (methylation of H3K27 and H3K9), whereas the others bind to these modifications and change the chromatin structure. In *Drosophila*, mouse and human the H3K27me3 is highly correlated with binding of *Polycomb* group proteins (Schuettengruber et al. 2007).

Products of *trithorax* genes exhibit the opposite activity — they are transcriptional factors or chromatin-remodeling enzymes, which are involved in maintaining the chromatin in an active state (via methylation of H3K4). *Polycomb* and *trithorax* complexes are highly evolutionary conserved; they are supposed to be crucial for the cell differentiation and cell fate plasticity. But they represent only the tip of the iceberg — of massive parallel processing of proteins in chromatin domains, on million of nucleosomes contained in it.

The lesson from our histone inquiry is that various kinds of “bar codes” inscribed onto the protein molecule during an individual’s life are not inherited in a ready-

made state: they come into existence by bootstrapping processes between the hardwired genetic message and the organic memory of the body. Hence, the buildup of organic memory (*sensu* Barbieri) is accompanied by “taking notes” in the form of a sequence of modifications — epigenetic counterparts of inherited informational molecules (codes).

The fact that such quasi-digital texts can be created — written and edited — during the lifespan of an organism, is the central starting point for our further investigation.

Two Worlds

Our model example invites an extremely interesting question concerning the relationship between the natural world and the world of digital coding. A comprehensive study in this area was provided by Emmeche and Hoffmeyer (1991); here we try to develop their views further, by distinguishing between the real and the transcendent world, their roles for living beings, and their “ontological” status.

Before we start our investigation, however, a terminological insertion is necessary, to avoid shaggy interpretations of terms which have been in use, literally, for ages. We need to distinguish clearly between *characters*, *signs*, and *symbols*. Our division is not original and we not pretend introducing a new, or correcting some of older systems: we simply need to clarify our usage of terms.

- (i) Character (digit, mark, tag) is a member of some finite alphabet (or table), and its *single* qualification is its *position* (its coordinate) in the given alphabet (or table). Characters have no meaning except (i) their *membership* of the set, and (ii) their *position* in that set; they are *neither* signs or symbols. No additional member may be inserted between two alphabetical places (no position left in between), and no transition characters (e.g. half U, half V) are allowed; the character is *absolutely unmistakable* among the other digits of a given alphabet. Thanks to this, it can be copied, and distinguished in a string of characters with absolute accuracy.

Characters dwell in the *transcendent (virtual) realm* (see below and Fig. 3): only here they retain their clear-cut definition, and can be lined up into strings which can be copied, ideally, i. e., without errors. As to their *appearances*, i.e., the images that they assume in *our natural world*, these are absolutely arbitrary as soon as their alphabetic coordinates are known. The string

“This sentence is a string of characters.”

remains the same when put in italics:

“This sentence is a string of characters.”;

in courier:

“This sentence is a string of characters.”;

or in Morse:

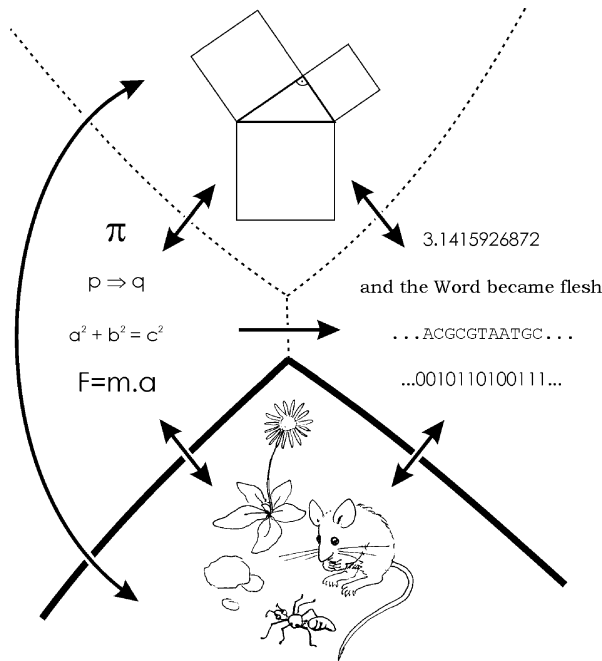


Fig. 3 The two worlds. The domain below the thick line represents the natural world, with its living beings and inanimate objects, languages, games, and history, all in endless becoming and extinction. The domain above is the transcendent world, subdivided to fit our text. The uppermost part represents the world of ideas, geometry, mathematics, logic, and all non-alphabetical scripts and symbols; here also dwells God of those religions which are based on Torah. Its items can be “zipped” into linear strings of *alphabetic* characters, which can be manipulated according to pre-established rules (left). Alphabetic strings, however, may exist also on their own, as singular, nominable entities (right). The items of sub-domains can be interchangeable, but their interpretation is always bound to the world natural

If such a string carries some meaning (in this case for people who speak English), the meaning remains the same upon such transformations, but transformations pay no regard to meaning. To illustrate our point: where Cowley (2008, 321) gives the scheme

$$\text{adaptor (e.g. Morse)} \rightarrow \boxed{\text{process}} \rightarrow \text{output (e.g. English)}$$

we prefer a reading:

$$\text{string (English alphabet, Morse font)} \rightarrow \boxed{\text{click}} \rightarrow \text{string (English alphabet, some other font)}.$$

The same holds for the Morse-Latin example of coding-decoding provided by Barbieri (e.g. 2003, 93).

The strings of characters may bear no meaning in whatever language, yet *their mutual transformation will proceed correctly*, provided the code is supplied. It should be also stressed that natural language is in no way *produced* by constructing such strings of characters; on the contrary, *some* aspects of language can be mapped on such strings. Only formal languages, i.e., those *created* by humans, can be developed “from below”, by starting with strings and a formal grammar.

- (ii) A sign (representamen, synthema) will be taken in the ordinary usage: as something standing for something else, in a given frame of contexts. In the frame of naval codes, SOS stays for “help”; in the genetic code AUG means “methionine”; for liver cells, insulin means “take up glucose from the bloodstream”; in the Highway Code an inverted triangle means “yield”; in the world of a hunter a footprint means “game”. Words expressed in a given language are signs as well. In contrast to Characters, Signs may dwell in both transcendental and real worlds, but their interpretation is *always* coupled to the real world: *no interpretation is possible in the transcendent realm only*. Signs can be used in both formal and natural languages.
- (iii) A symbol (omen) is reserved only for entities loaded with a long cultural, historical, religious, etc., tradition. The meanings of symbol may keep a whole community together, without long deciphering and explanations. Examples: national flags and anthems, the Christian cross, the Red cross, etc. It follows that we can speak about symbols only in connection with natural languages, and we shall not use the term here.⁵ In connection with our casuistry, we shall try to stick to the proposed terminology, in order not to use the same term in multiple contexts (like, e.g., in a recent article by Pattee 2007, where “symbol” is used in several meanings, not easily discernible from each other).

Transcendent and Natural

In Fig. 3 we propose an existence of two worlds: the natural world we live in, and the world transcendental, in a Platonic sense, an ideal “otherworld”. For the sake of our discussion, we invite the reader to consider the divide between them to be as sharp as possible.

The domain below the line represents the natural, bodily world of cosmos, life, events; the world of our experience where we feel at home. Here, things may be similar, but *never* identical; they cannot be copied, only imitated. Any attempt of digitalization in the natural world is only approximate, hence error-prone. Life, natural language (including its possible analogues in non-human species, cells, etc.) is the product of this natural, or bodily, world (Markoš 2002; Cvrčková and Markoš 2005).

The uppermost domain of our scheme is the ideal world of geometrical objects, ideas, mathematics and logic; entities that behave orderly, much more obediently than those of the natural world. The ideal world does not fill any place in our world; it is a transcendent world, *invented* apparently by the Greeks at the dawn of Western civilization, and cultivated ever since. In the following centuries, people found that

⁵ Note: the same appearance may play a role in all three contexts. Hence, Ω is (i) last character of the Greek alphabet, (ii) a sign for the unit of electric resistance (ohm), and (iii) a Christological symbol in theology; similarly “666” is (i) string of three digits, (ii) a sign for a number (with a meaning, in a decimal system, “six hundred and sixty six”), or (iii) a symbol in Kabala and in apocalyptic mysteries. A reverse path is not possible: obviously the Christian cross does not belong to any alphabet, so the character “+” (plus) belonging to the set alphabet of arithmetic digits, has nothing in common with Christian symbolic.

knowledge invented in the transcendental domain is endowed with the power to influence the course of our natural world, and hence can be useful.⁶ From such positions, it is easy to adopt a belief that the world is actually *ruled* by this transcendental realm — be it laws of God or, in our days, laws of Nature.

Ideas and philosophical systems were recorded in stone, papyrus, parchment, or paper using alphabetic *characters*, an invention and heritage of Semitic cultures. Apparently alphabetic characters were not invented to record utterances; soon, however, they were adjusted to communicate also language expressions and numerals. Much later, modern science came to the astonishing discovery that the entities of the transcendental realm — ideas, geometry, math, etc., — can all be reduced to a common denominator, i.e., to strings of alphabetic characters. This means that the transcendent world can be communicated to the real world via strings and decoded there, to influence the behavior of the material world. In reverse, experiences gained in the world can be frozen — mapped into such strings.

This brings our attention to the domain in the middle of our scheme (but still safely on the transcendent side of the line). It represents the space of such character strings: virtual, devoid of any sign of bodily existence, digital and accurate. Here they can be copied and trans-coded into different alphabets with an utmost accuracy (never attainable in the real world).

For our purposes, the subspace of strings is subdivided into two parts. We start our examination in the left one, in the realm of mathematics and logics, analytical geometry, ballistic curves etc., the realm of “natural laws”; note that we place *natural* laws *outside* nature, in the transcendental. Analytical geometry, logics, mathematics etc. found their refuge here, neatly expressed in a linear sequence of marks. Here science was born and here she has her seat, here she construes theories, paradigms and hypotheses. This is the space that allows for generalization: the calculation of force from mass and acceleration is universal, the function contains virtually *all* possible combinations of these variables, given initial and boundary conditions; the Pythagoras’ theorem is valid for *all* right-angled triangles in a Euclidean space, etc.

An important note: the realm can exist *only* because of its interaction, across the barrier, with the natural world. Only from here, meanings come for all virtual contraptions encountered, only from here the initial and boundary conditions are provided; without such an interaction we would be left with a kind of plain algebra (but even here the *rules* must be delivered from the natural world).

Science gains its calculations and hypotheses here, in the transcendent realm, and takes them across the barrier, to the real world, to test them. To do so, it is necessary to create a kind of interface — an artificial world-in-between — embodied in laboratory models: bodies falling in a vacuum, particle colliders or specially constructed cultures of organisms and cells, or models run on computers. If such models behave according to the expectations of the theory, the scientist leaves the lab and starts to examine the external world, where no vacuum exists, forms of life are “wild”, logical theorems do not represent the highest commandment, and

⁶ Up to the point that virtual realms are considered better, flawless, and more logical when compared to the world of our everyday lives. Such views resulted in a gradual development of contempt to the bodily world.

digitalization is possible only to a certain degree (quasi-digitalization, see below). If the world does not behave, it means that there is something wrong either with the theory, or with the testing procedure; then the scientist meekly returns to the lab and starts polishing both. Or she postulates additional factors teeming in the world but not taken so far into consideration by the model; after all, *real* falling bodies experience friction; wild organisms are more inventive than tamed laboratory strains, etc. In this way, with the highest carefulness, we scientists cross the barrier between the two worlds many times. If we keep in mind that one of them — the virtual one — is nothing but *our* construct (built on assumptions pronounced, learned, or silently adopted without much contemplation); if we are aware that testing is performed on artificial fragments of the real world stitched to fit our models; if we realize that the path from the model to reality is dangerous and must be taken with utmost carefulness, then the world will cooperate, will yield to our intrusions and will allow constructing artifacts never seen before — say a liner or a microchip. (Remark: An attempt will be made below to subsume under “we” not only scientists, but all living beings.)

Anyway, the astonishing fact is that the barrier is penetrable, that we can gain inspiration in both worlds, and even state that our virtual constructs — displayed as string of characters — can encode certain properties of the real world. Even more, the world will *obey* rules constructed in the virtual realm. There are many, however, who tend to forget about the real world and to enthrone their virtual construct (objective reality) in its place; some even insist that the virtual realm has existed from time immemorial, and the real, bodily world depends on it. In a way, such a stance is comfortable: if something has existed “from the very beginning”, it is always the same and obeys identical principles in all places. If so, the history of such a world is not due to fate; it is, instead, governed by *chance and necessity*; either strictly prescribed, or a matter of drawing from the wheel — of ready-made, ever-present scenarios. Many scientists would endorse this last view — of the body turned into a pure code.

Now let us shift our attention to the right part of the digital realm in Fig. 3. Here strings constructed from alphabetical characters are dwelling, but no meaning is attached to them. Let us demonstrate the difference between the two areas of the digital world on an alphabet consisting of five positions: [a], [F], [m], [x], and [=]. The right part allows writing any possible strings, like “==xamaFFx”, “FFFFxm=aa”, or “aaaaa”; they may be arbitrarily long, and any combination of characters is allowed, there is no rule or prescription how to compose a string. If we have a string “FFFF...”, we cannot guess which digits will come up on the 5th position — we should simply look. No generalization is possible here; strings are nominable transcendental entities (see Barbieri 2007 on such *nominabilia*). It remains to say that the characters of all such alphabets can be reversibly trans-coded into strings of mere two signs of a binary “alphabet” (usually writ as “0” and “1”), i.e., into strings of binary characters.

Meaning and the World of Strings

Our principal thesis is that *strings may acquire meaning only when confronted with the natural world*. If we return to the left part of the digital realm, here we also

encounter strings of alphabetic characters, but *now* we realize that the strings are not arbitrary — this is because here *strings became signs* — they are connected with *meaning*: installed, however, from behind the barrier, from the natural world. To continue with our example, let [a] designate acceleration, [F] force, [m] mass, [x] is the operator “times” in multiplication, and [=] is the sign for equality; moreover, we define the boundary condition and note that we are in the realm of Newtonian mechanics, and use an algebraic notation allowing commutative rules. Suddenly it turns out that only four strings are *allowed* under such conditions, each of a length 5: $F=axm$, $F=mx a$, $axm=F$, $mx a=F$. Meanings and the framework of boundary conditions drastically reduced the number of possible sequences. This is the reason why, in our scheme, the arrow between both areas of the virtual realm points only in one direction: mathematical and logical formulas, sentences and lemmas *can* be written as a sequence of characters, but strings of digits *cannot* be reduced to mathematical or logical formulas: each such string is unique, and we cannot do more than to *name* it. This can, again, be done only by bestowing the strings with meaning, and meanings (and thus reduction of degrees of freedom in writing) again can enter the digital world only from the natural world: directly, or via the virtual world of ideas. To write them down, we only borrowed the specific marks of some alphabet and put them down into a specific sequence. The nominal realm of strings has also a direct access to the natural world, and the interface, crossing the boundary from *here*, may hold the clue to the mystery of life.

Number, Program, Text: Decoding vs. Reading

The space of sequences (for the sake of simplicity let’s suppose that they all come in binary form) may influence the natural world in three extremely interesting ways: strings may come up as *numbers*, *programs*, or *texts*.⁷ It is not trivial to distinguish in which way a string should be interpreted in the real world. In principle there can be two categories of entities (i) a coder/decoder (or “codec”)⁸, and (ii) a reader able to extract meaning from a string when it (she, he) takes it for a text.

It is in the powers of a reader to switch into the codec regime, but a codec can never “decide” to become a reader — it has no clearance for such a decision. The difference is that decoding proceeds in formal language, reading in natural language.

Program Execution (Formal Language)

As mentioned above, the coding/decoding device must dwell in the natural world, and the string must be delivered to it on a suitable *medium*. For a computer it may have the form of irregularities on a magnetic disk, for its processor it is a succession of electric impulses, for the punched-card reader it is a difference hole/non-hole, for a ribosome the sequence of triplets on mRNA, etc. However, by inscription into a

⁷ But note that the communication between the transcendent and the real is *not* restricted to digital channels operating with strings of characters.

⁸ We prefer “codec” to Barbieri’s (2003) “codemaker”, which means roughly the same but raises misleading association with some maker, or creator of a code.

medium, the characters cease to be absolutely digital: not always can they be distinguished unequivocally.

We call such material embodiment of string sequences *quasi-digital*. It follows that they can be neither copied nor trans-coded infallibly:⁹ they are prone to mutations — i.e. misreading and mistranscription of their quasi-digits, confusion with a different character of the alphabet, etc. But it is not our task here to discuss the problems of digitalization, i.e., smuggling ideal digits into the quasi-digits of the real world. What is more important:

- (i) A formal language necessary for program execution was derived (created) by entities in command of natural language. Formal language is defined as a set of character strings, and operations upon them (Searls 2002).
- (ii) A non-living device, a machine, will suffice for the task of scanning and executing — thoughtlessly, mechanically. As a drastic example take a missile heading towards its target. In other words, formal language is a domain of trans-coding between virtual strings; *but the trans-coding program (rules) comes from the natural world, from entities with the command of natural language.*
- (iii) A virtual string of marks, when embodied into a suitable medium and scanned by a suitable device, can influence the behavior of the world.
- (iv) The question of the maker who constructed the device and wrote the program in a formal language is easy to answer for man-made machines, but enormous problems arise when contemplating self-reproducing automata (for a discussion, see Pattee 2008). Should we consider living beings as such automata (without adopting the creationist worldview), their coming into being, and evolution, remains an enigma, even though reasonable scenarios had been proposed for the program-first as well as device-first alternative evolutions (see. e.g. Cairns-Smith 1982, 1985, and Kauffman 2000, for respective possibilities).

But the most important, and perhaps most controversial, statement of our analysis is this:

- (v) *Formal languages, as known today, do not know any semiosis — they work on the level of codec. Interpretation is the virtue of natural languages.* For the sake of sharpening our vision of the problem we stick to this statement, even if we are fully aware that linguists, philosophers, and scientist are not united over the problem.

Reading (Natural Language)

To perform reading, a mere device is not sufficient. Reading is a semiotic and hermeneutic task and requires a *community* of living beings (humans or, as we

⁹ Actually, copying, i.e. production of identical entities, is *not* possible in a real world: here no two things are identical — only similar; similar *according to* some criteria.

believe, all living beings, or cells, or even protein ecosystems in case of some texts) — anchored in *language, culture, ecosystem, history*. Natural language is a product of a long evolution of such a community of speakers, and can be taken as a field (potential), from which individual speech acts (expressions) unwind, and understanding is negotiated. Hence, *natural language is a phenomenon of the natural world that cannot be transferred into the transcendent realm, and even less can it be reduced to digits*. We devote a parallel paper to the language metaphor of life; here we list only a few comparisons that highlight the difference between program execution and reading.

- (i) A natural language cannot be produced from scratch, unless we envisage a creator speaking a meta-language (compare with formal languages); a community of speakers, and historical continuity of such a community in time is required.
- (ii) Communication in language proceeds in utterances, which — in contrast to language itself — *can* be parsed, with some reserve, to linear strings of quasi-digital units (morphemes); *grammatical* rules concerning groupings of morphemes can be derived for any given language.
- (iii) To produce speech acts (speaking or writing), speakers are required who are *not* machines; similarly, to perceive the message, i.e., to accomplish understanding, living interpreters are required who are *not* machines;
- (iv) The quasi-digital nature of strings of morphemes allows, with some reserve and with established rules, to *map an utterance in a form of a character string*. By this *artifact making*, the utterance can be saved in the transcendent, atemporal world of characters, and copied infallibly. Artifact-making needs life (Barbieri 2007); life dwells in the real world. As in case of formal languages, strings can be “materialized” by embodying them into a suitable medium (like print on paper, or a file on the hard disk); as in the case of formal languages, the copying of strings and their embodying can be performed by machines.
- (v) Reading such strings requires literate speakers in natural language who are able to transform them back into real-world utterances (spoken or not), in order to understand them.

Whereas a codec will supply a single, deterministic “execution” of a given string, leading every time to identical “interpretative” result, semiotic and hermeneutic abilities of living beings will conjure up, on the same text, a bunch of non-identical interpretations (even if they *may* be quite *similar* in some shared contexts). The result of an interpretation cannot be unequivocally foreseen, because every reader approaches the text influenced by her/his/its previous experience (“organic memory”); he/she/it somehow *understands* the text even before starting reading; if not, reading would be an impossible task. Hence, something deeply interesting occurs during the process of reading. A text written as a string of letters is unequivocal, reproducible, unchanging. Transferred into the real world, however, there pops up — via readers — a great variety of interpretations, very often mutually antagonistic. For writing to have any sense, the community should achieve an agreement on *how* to read. In case of especially important texts the interpretation is even ordained by an authority — religious, political, judicial, scientific, etc. — and in such cases it resembles the program execution. Even in these cases, however, the

consensus often will remain valid only for a limited period of time — no interpretation in natural language is unchangeable and eternal.

Conclusions

If it is true that information flow and inscription/reading constitute the principal distinctions of life, as we believe, we need to suppose that language-like properties exist at various levels of life's organization. Such a conjecture requires the following points:

- (i) The presence of a community of speakers, whose historical continuity in time is ensured by “material” (natural) perpetuation (many generations, long periods of time) of individuals, be it cells, community of cells, individuals in a species, ecosystems, cultures, etc. In our example, it is the network of proteins, which is the heir of such an uninterrupted tradition, and has its ways about how to read, as well as to put down, quasi-digital “notes” on media like DNA and histone molecules. In reference to Fig. 3, such “speakers” are limited to the realm of digital string — we do not pretend their inventing mathematics, logic, or even ideas and God.
- (ii) As in spoken language, “utterances” can be put into one-dimensional form, and be quasi-digitalized. Such strings of quasi-digital units can be parsed and analyzed by methods developed by linguistics; in this way, science can decipher grammatical rules of processing, which constitute the background in most signal-processing units (second messengers, tags on DNA, on histones, or on other proteins of signaling cascades, etc.).
- (iii) All processes of taking notes and bookmarks, as discussed above, can be taken as speech acts (speaking or writing) accomplished by speakers who are *not* machines;
- (iv) The quasi-digital nature of linear (in sugars also branched) aperiodic biomolecules is a characteristics which does not need any comment. What *does* need a comment, however, is the question of the order of events; many authors prefer the primacy of spontaneous origin of “written” macromolecular strings; in such a case the whole edifice of this paper would be in serious difficulty.
- (v) Reading such strings requires speakers in natural language: such features are best demonstrable in the science of evo-devo: they show that a limited genetic toolkit is sufficient to erect all existing animal body plans (e.g. Carroll 2005; Carroll et al. 2006). In our wording: every species has its own hermeneutic rules of meaning-making.

Acknowledgement Supported by the project of the Czech Ministry of education MSM 0021620845, and the GPSS Major Awards Programme, a joint programm of the Interdisciplinary University of Paris and Elon University.

References

- Allis, C. D., Jenuwein, T., & Reinberg, D. (2007). *Epigenetics*. Cold Spring Harbor Laboratory Press.
- Barbieri, M. (2003). *The organic codes. An introduction to semantic biology*. Cambridge UP.
- Barbieri, M. (2007). Is the cell a semiotic system? In M. Barbieri (Ed.), *Introduction to biosemiotics* (pp. 179–207). Springer.

- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, *125*, 315–326.
- Bryk, M., Scott, D. B., Strahl, B. D., Curcio, M. J., Allis, C. D., & Winston, F. (2002). Evidence that Set1, a factor required for methylation of histone H3, regulates rDNA silencing in *S. cerevisiae* by a Sir2-independent mechanism. *Current Biology*, *12*, 165–170.
- Cairns-Smith, A. G. (1982). *Genetic takeover and the mineral origins of life*. Cambridge: Cambridge University Press.
- Cairns-Smith, A. G. (1985). *Seven clues to the origin of life*. Cambridge: Cambridge University Press.
- Carroll, S. B. (2005). *Endless forms most improbable. The new science of evo devo and the making of the animal kingdom*. Norton.
- Carroll, S. B., Grenier, J. K., & Weatherbee, S. D. (2006). *From DNA to diversity. Molecular genetics and the evolution of animal design* (2nd ed.). Malden: Blackwell Science.
- Costa, F. F. (2008). Non-coding RNAs, epigenetics and complexity. *Gene*, *410*, 9–17.
- Cowley, S. J. (2008). The codes of language: Turtles all the way up? In M. Barbieri (Ed.), *The codes of life. The rules of macroevolution* (pp. 319–345). Springer.
- Cvrčková, F., & Markoš, A. (2005). Beyond bioinformatics: Can similarity be measured in the digital world? *J. Biosemiotics*, *1*, 87–105. Reprinted in: *Biosemiotics: Information, codes and signs in living systems*, Barbieri M ed., New York: Nova science, 65–79.
- Emmeche, C., & Hoffmeyer, J. (1991). From language to nature: The semiotic metaphor in biology. *Semiotica*, *84*, 1–42.
- Fischle, W., Tseng, B. S., Dormann, H. L., Ueberheide, B. M., Garcia, B. A., Shabanowitz, J., et al. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature*, *438*, 1116–1122.
- Giannis, A., Biel, M., & Wascholowski, V. (2005). *Epigenetics-An epicenter of gene regulation: Histones and histone modifying enzymes* (pp. 3186–3212). Wiley-VCH Verlag GmbH & Co.
- Kauffman, S. (2000). *Investigations*. Oxford: Oxford UP.
- Kingston, R. E., & Tamkun, J. W. (2007). Transcriptional regulation by Trithorax group proteins. In Allis et al (Ed.), *Epigenetics*.
- Kouzarides, T. (2007). Chromatin modification and their function. *Cell*, *128*, 693–705.
- Lan, F., Bayliss, P. E., Rinn, J. L., Whetstone, J. R., Wang, J. K., Chen, S., et al. (2007). A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature*, *449*, 689–695.
- Liu, C. L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S. L., Friedman, N., et al. (2005). Single-nucleosome mapping of histone modification in *S. cerevisiae*. *PLoS Biology*, *3*, 1753–1769.
- Margueron, R., Trojer, P., & Reinberg, D. (2005). The key to development: Interpreting the histone code? *Current Opinion in Genetics & Development*, *15*, 163–176.
- Markoš, A. (2002). *Readers of the book o life*. Oxford: Oxford University Press.
- Markoš, A., Grygar, F., Kleisner, K., & Neubauer, Z. (2007). Towards a Darwinian biosemiotics. Life as mutual understanding. In M. Barbieri (Ed.), *Introduction to biosemiotics* (pp. 235–255). Springer.
- Markoš, A., Grygar, F., Hajnal, L., Kleisner, K., Kratochvíl, Z., & Neubauer, Z. (2009). *Life As Its Own Designer: Darwin's Origin and Western Thought*. Springer.
- Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, *448*, 553–560.
- Pattee, H. H. (2007). The necessity of biosemiotics: Matter-symbol complementarity. In M. Barbieri (Ed.), *Introduction to biosemiotics* (pp. 115–132). Springer.
- Pattee, H. H. (2008). Physical and functional conditions for symbols, codes, and languages. *Biosemiotics*, *1*, 147–168.
- Popov, O., Segal, D. M., & Trifonov, E. N. (1996). Linguistic complexity of protein sequences as compared to texts of human languages. *BioSystems*, *38*, 65–74.
- Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., & Cavalli, G. (2007). Genome regulation by Polycomb and trithorax proteins. *Cell*, *128*, 735–745.
- Searls, D. B. (2002). The language of genes. *Nature*, *420*, 211–217.
- Swigut, T., & Wysocka, J. (2007). H3K27 Demethylases, at long last. *Cell*, *131*, 29–32.
- Trifonov, E. N. (2008). Codes of biosequences. In M. Barbieri (Ed.), *The codes of life. The rules of macroevolution* (pp. 3–14). Springer.
- Turner, B. M. (2007). Defining an epigenetic code. *Nature Cell Biology*, *9*, 2–6.
- Vakoc, C. R., Mandat, S. A., Olenchock, B. A., & Blobel, G. A. (2005). Histone H3 lysine 9 methylation and HP1 γ are associated with transcription elongation through mammalian chromatin. *Molecular Cell*, *19*, 381–391.