Interdisciplinary Description of Complex Systems 3(2), 59-71, 2005

COMPLEXITY, INFORMATION AND BIOLOGICAL ORGANISATION

Attila Grandpierre

Konkoly Observatory of the Hungarian Academy of Sciences
Budapest, Hungary

Conference paper *Received:* 15 September, 2005. *Accepted:* 15 October, 2005.

SUMMARY

Regarding the widespread confusion about the concept and nature of complexity, information and biological organization, we look for some coordinated conceptual considerations corresponding to quantitative measures suitable to grasp the main characteristics of biological complexity. Quantitative measures of algorithmic complexity of supercomputers like Blue Gene/L are compared with the complexity of the brain. We show that both the computer and the brain have a more fundamental, dynamic complexity measure corresponding to the number of operations per second. Recent insights suggest that the origin of complexity may go back to simplicity at a deeper level, corresponding to algorithmic complexity. We point out that for physical systems Ashby's Law, Kahre's Law and causal closure of the physical exclude the generation of information, and since genetic information corresponds to instructions, we are faced with a controversy telling that the algorithmic complexity of physics is much lower than the instructions' complexity of the human DNA: $I_{\text{algorithmic}}(\text{physics}) \sim 10^3$ bit $<< I_{\text{instructions}}(\text{DNA}) \sim 10^9$ bit. Analyzing the genetic complexity we obtain that actually the genetic information corresponds to a deeper than algorithmic level of complexity, putting an even greater emphasis to the information paradox. We show that the resolution of the fundamental information paradox may lie either in the chemical evolution of inheritance in abiogenesis, or in the existence of an autonomous biological principle allowing the production of information beyond physics.

KEY WORDS

levels of complexity, the computer and the brain, algorithmic complexity, complexity and information, fundamental information paradox of the natural sciences

CLASSIFICATION

APA: 4100

PACS: 82.39.Rt, 89.75.–k, 89.70.+c

*Corresponding author, η: grandp@konkoly.hu; +36 1 391 9360;
Konkoly Observatory of the Hungarian Academy of Sciences, P.O. Box 67, H-1525 Budapest, Hungary.

INTRODUCTION AND THE MEMORY-DNA PROBLEM

Brain's complexity is widely considered in terms of neurons and synaptic connections, e.g. [1]. For a number of neurons $N = 10^{11} - 10^{13}$ [2], taking a value for the number of their interconnections as a few thousand per neuron $c \approx 10^4$ [3], we obtain for the measure of the brain's complexity the number $N_1 = N_{\text{connections}} = c \cdot N_{\text{neurons}} = 10^{15} - 10^{17}$. The general view is assuming a connection (synapse) represents 1 bit of information. In this way, we obtain for the information measure of the brain's complexity a value of $I_1 = I(\text{human brain}) \sim 10^{15} - 10^{17}$ bit. Now since algorithmic complexity may be characterized by the size of the memory, we obtain that $I_1 = I_{\text{algorithmic}}(\text{human brain})$. The biggest supercomputer today, Blue Gene/L, has a memory capacity of 64 TB, corresponding roughly to $5 \cdot 10^{14}$ bit, a value not far from our brain's synoptic capacity I_1. To put the brain's algorithmic complexity into context, we mention a few related measures. Importantly, Maynard Smith [4] noted that the genetic information content of the human DNA corresponds to instructions and it is about $I_2(\text{DNA}) \sim 10^9$ bit. Since the instructions coded in DNA control all cellular processes [5], we may regard genetic information I_2 as acting at least at the algorithmic level or deeper, and so $I_2(\text{DNA}) \leq I_{\text{algorithmic}}(\text{human organism})$. A third important measure is found in [6] measuring a special genetic complexity by the simplest model, in which each gene is either ON or OFF, and so a genome with N genes can theoretically encode $2N$ states. With 30 000 genes indicated to be present in the whole human genome, the arising human genetic complexity is a mere $I_3 = I_{\text{genetic expression}}(\text{human organism}) \sim 3 \cdot 10^4$ bit. We obtained the following result: $I_1 = I_{\text{algorithmic}}(\text{human brain}) \sim 10^{15} - 10^{17}$ bit $\gg I_2(\text{DNA}) \gg I_3 = I_{\text{genetic expression}}(\text{human organism}) \sim 3 \cdot 10^4$ bit.

The first fruit of absorbing these complexity measures arises when we recognize a problem: how can it be that the genetic complexity of the human organism (including the brain) $I_2 \sim 10^9$ bit is smaller than the algorithmic capacity of the human brain, $I_1 \sim 10^{15} - 10^{17}$ bit – if the brain receives merely morphological and no algorithmic information from the environment through our senses during our lifetime? Would it be possible that the brain absorbs somehow algorithmic complexity from the environment as well? Or the complexity measures $I_1 \sim 10^{15} - 10^{17}$ bit and $I_2 \sim 10^9$ bit correspond in reality to different levels of complexity? To solve this problem (the memory-DNA problem), we will need estimations of complexity measures for the brain's dynamic activity as well as for the algorithmic complexity of the environment.

DYNAMIC MEASURES OF COMPLEXITY

Let us turn again to the computer-brain metaphor to see whether we can recognize some dynamic aspects of complexity that may be useful for further clarification. A measure of dynamic complexity is the number of operations per second. The number of operations per second in the Blue Gene/L in the third quarter of 2005 is 367 Teraflops – i.e., $3{,}67 \cdot 10^{14}$ operations per second. This is to be compared to the number of operations in the human brain per second. Let us take first operations corresponding to neural action potentials. Considering that the visual input into the brain comes through the 10^8 retinal cells, and 10^6 retinal cells are connected to the brain with axons sending 100 spikes of action potentials per second, regarded as carrying 1 bit of information each, one obtains 10^8 bit per second for the visual input into the brain. Assuming that an average neuron processes at a similar rate of operations, or 100 operations per second per neuron, we obtain for the $10^{11} - 10^{13}$ neurons a value of $10^{13} - 10^{15}$ operations per second as the number of "neural operations" in the brain, $N_1 = 10^{13} - 10^{15}$ operations per second. The close agreement of the dynamic complexity of the Blue Gene/L with that of the brain's neural complexity lends certain plausibility to

attempts at modeling the brain in terms of the Blue Gene/L (The Blue Brain project, http://bluebrainproject.epfl.ch).

In reality, neural action potentials do not form a closed chain of events arising from a given initial state. Instead, they are continuously influenced by the information flow coming through the outer senses, and from internal processes extending from cellular chemical reactions up to the level of self-consciousness – by means of processes whose mathematical description far transcends the computational capacity of the Blue Gene/L. Therefore, it appears that although Blue Gene/L may be suitable to simulate brain's neural activity, it is a poor choice for modeling the brain's activity at the molecular level. Actually, Blue Gene/L is planned to simulate protein folding [7].

DYNAMIC COMPLEXITY MEASURES AT THE MOLECULAR LEVEL

To consider complexity measures corresponding to the molecular level, let us try to estimate the number of chemical reactions per second in the human organism. Certainly, the number of chemical reactions per second is larger than the number of ATP molecules produced per second. Kornberg [8] determined that the average daily intake of about 2500 kcal, corresponding to approximately 100 W, translates into a turnover of a whopping 180 kg of ATP. This number translates into $N_2 = N_{ATP}(\text{organism}) \sim 2\cdot10^{21}$ ATP molecule production per second in the human body. Regarding the fact that the ATP is produced in a chain of electron transfer events, and acts though energy coupling that involves the coupling of two reactions occurring at the same time, at the same place, typically utilizing the same enzyme complex, we find it plausible to assume that the rate of ATP production of $N_{ATP}(\text{organism}) \sim 2\cdot10^{21}$ operations per second is smaller than the number of all chemical reactions of the human organism, $N_3 = N_{\text{chemical reactions}}(\text{organism}) > 2\cdot10^{21}$ chemical reactions per second. It is clear that both the production of each ATP molecule together with its reactants has to be timed so that the energy coupling can take effect, and that this timing is not completely pre-programmed because it depends on the cellular, intercellular, and global organizational levels. Each chemical reaction in the cell may occur sooner or later, here or there, therefore, ignoring now the question of redundancy which will be considered later below, one may count that at least 1 bit is necessary for their proper timing. Therefore the flux of biochemical reactions corresponds to a rate of information production $\dot{I}_1 = \dot{I}_{\text{biochem}} > 2\cdot10^{21}$ bit/s. With $6\cdot10^{13}$ cells in the body, we obtain a lower limit $\dot{I}_{\text{lower}}(\text{cell}) > 4\cdot10^{7}$ bit/s. When this measure applies to neurons, we obtain that the dynamic chemical complexity of the brain exceeds by 6 orders of magnitude the complexity of the neural level.

ARGUMENTS EVALUATING THE BIOLOGICALLY UTILISED PERCENTAGE OF THE THERMODYNAMIC CAPACITY

i.) It is well known that the biological efficiency of cellular respiration is about 40 %, and that the general efficiency of the living organism is also about 40 % [9]. While in engineering such a rate of efficiency may be reached, there is a big difference that makes sense for complexity measure considerations. In machines, the energy transfer occurs through a few macroscopic degrees of freedom, corresponding to the moving constituent parts of the machine, in living organisms the energy flux does not flow automatically but is utilized by the living organism for molecular processes. Therefore in living organism the energy is continuously redistributed on microscopic degrees of freedom, on electronic excitation levels, activating just the chemical reactions the occurrence of which is useful for biological activities. Therefore in living organism a significant part of microscopic degrees of freedom corresponds to the dynamic biological information flux flowing from DNA to cellular reactions. This means that the approximately 40 % biological efficiency

is related to an astronomically high information flow corresponding to the app. 40 % utilization of the thermodynamic capacity of the living organism.

ii.) Now let us estimate the thermodynamic capacity of the human organism. With a metabolism rate of L(organism) ~ 100 W the human body can mobilize an extropy flow $\Pi = L_{out}/T_{out} - L_{in}/T_{in} \sim 3{,}3 \cdot 10^{-1}$ $J \cdot K^{-1} \cdot s^{-1}$ [10], and this translates in information units to $\dot{I}_2 = \dot{I}_{TD}$(organism) $\approx 3 \cdot 10^{22}$ $bit \cdot s^{-1}$. This means that the lower limit of information flux we obtained above, $\dot{I}_1 = \dot{I}_{biochem} > 2 \cdot 10^{21}$ $bit \cdot s^{-1}$ is within an order of magnitude to the thermodynamic limit, a fit that may be regarded as quantitatively underpinning our argument. Nevertheless, we find it worthwhile to mention some further theoretical and quantitative arguments and tests on this point.

iii.) Ashby [11] pointed out that organization means conditionality, and since biological organization extends to the whole of the organism, every molecule's behavior is conditional, contingent on every other molecule's activity in the cell. There are strong indications that biological organization acts at the molecular level, e.g. [12]. Certainly, a significant part of the molecules of the cell has to follow highly specific pathways. The findings of proteomics, systems biology, and structural biology indicate that the organization of chemical reactions occurs simultaneously in intimate interactions between the molecular, cellular and higher levels. To make these complex interactions possible, Davies [13] noted that biological signals released by nucleic acids do the job to instruct ribosomes to assemble proteins, freeing protein assembly from the strictures of chemistry and permitting life to choose whatever amino acid sequences it needs. The complex of instructing biological signals influence chemical reactions of the cell in a way that is highly non-redundant. At present, little is known about how cells integrate signals generated by different receptors into a physiological response [14], yet it is clear that biological organization at the level of the cell contributes as well as higher and lower levels (corresponding to DNA, its genetic and nucletypic roles, cells, individuals, populations, species). Petricoin et al. [15] formulated that the ultimate goal of proteomics is to characterize the information flow through protein networks that interconnect the different and numerous regulatory systems of the organism. There are eleven major body regulating systems in human physiology: the circulatory, digestive, respiratory, urinatory, skeletal, muscular, integumentary, immune, nervous, endocrine and reproductive systems [16] and all of them influences each cell's chemical reactions. Regarding the non-redundant character of chemical reactions of the cells we note that evolutionary studies had shown that biology attempts to optimize resources. Therefore, it is not implausible to conjecture that biological organization may approach its thermodynamic limits, at least regarding informational resources.

iv.) Aoki [17] estimated the entropy production of the human body as 0,259 $J \cdot K^{-1} \cdot s^{-1}$. This dynamic complexity measure is to be compared with the extropy flow [10] utilised by the whole organism. A food intake of 100 W corresponds to 0,325 $J \cdot K^{-1} \cdot s^{-1}$. On the basis of these crude approximations, we derived the result that nearly 20 % of the total thermodynamic capacity of the human organism can be actually utilized for biological organization. In contrast, the general view is that the net efficiency of the utilized energy income is around 20 – 50 %, and in certain cases it may be even higher. For our present purposes, it suffices to recognize that biological organization utilizes a significant part (say 20 – 50 %) of the thermodynamic informational capacity. This result also fits well to our estimation of the complexity measure $\dot{I}_1 > 2 \cdot 10^{21}$ $bit \cdot s^{-1}$, the information flux present in biochemical reactions, as compared to the thermodynamic capacity of the organism, $\dot{I}_2 = \dot{I}_{TD}$(organism) $\approx 3 \cdot 10^{22}$ $bit \cdot s^{-1}$.

LEVELS OF BIOLOGICAL INFORMATION

Having obtained a quantitative and confirmed result for the dynamic chemical complexity measure, let us now consider how the levels of complexity are interrelated. Maynard Smith [4] realised that one could quantify biological information at three levels. First, at the genetic level, the biological information content is app. 2 bit per base. Second, at the selection level, a value of $\dot{I}$(evolution) ~ 0,2 bit·year^{-1} is found [18]. We add that one could expect that the appearance of the first living cells on the Earth, allegedly by abiogenetic way, would contribute to an enormous acceleration of the accumulation of genetic information, in comparison to the merely chemical evolution. Apparently, as the above obtained numerical measures of complexity show, the case is different. If the first life form has a similar complexity to the smallest genome yet found in free living organism, marine α-proteobacterium (Pelagibacter ubique), having a genome consisting of 1 308 759 base pairs, corresponding to app. $1{,}3\cdot 10^6$ bit, than it had to evolve certainly in less than hundred million years, and so its rate of developments had to be (much) higher than $\dot{I}_{\text{lowerlimit}}$(abiotic) ~ 0,013 bit·year^{-1}, a value comparable with Kimura's $\dot{I}$(evolution) ~ 0,2 bit·year^{-1}. These comparable values show a sharp contrast with plausible expectations that life is enormously more efficient in accumulating information then prebiotic processes. Third, biological information can be quantified at the morphological level. But to consider the morphological level, one has to be careful, for the genome is not a description of the adult form, but a set of instructions on how to make it. Maynard Smith emphasizes that the genome is a recipe, not a blueprint. We note that the genetic level corresponds to a complexity level at the algorithmic complexity or to a yet deeper level of complexity, regulating the algorithms. We think this is one of the main reasons why complexity sciences like cellular automata and self-organization, etc., enter into the scene: These sciences also recognize that it is possible to generate apparently complex products at the phenomenal level by means of simple physical or mathematical rules. Therefore, the real question is not what the degree of complexity at the morphological level is, but how complex an organism is at the algorithmic and at deeper levels.

MEASUREMENT OF BIOLOGICAL COMPLEXITY

Maynard Smith and Szathmáry [19, p.5] presented Table 1.1 summing up the genome sizes and percentages of coding DNA for bacterium (E. coli), Yeast, nematode, fruit fly, newt, human, lungfish, and flowering plants. They realized that when we allow for the fact that a varyingly small proportion of the DNA codes for anything, we may obtain a combined measure as a function of genome size as well as the percentage of coding DNA, a measure that makes sense. This measure, the size of the coding DNA, shows a progressive increase from bacteria to humans, with some minor exceptions only (lungfish). They noted that what this biological complexity measure tells us about structural and functional complexity is very limited [19, p.5]. On the basis of their Table 1.1, the coding part of human genome has N_{bp}(coding) ~ $6\cdot 10^8$ base pairs.

COMPLEXITY JUMPS IN THE HISTORY OF LIFE AND THE PROBLEM OF ABIOGENESIS

We find it remarkable that the size of the coding DNA shows a mere hundredfold increase from bacteria to humans, from $4\cdot 10^6$ base pairs to $6\cdot 10^8$ base pairs. It is widely thought that terrestrial life were already present within 100 million years after the solidification of the Earth's crust. In this context, it is important to take into account the fundamental fact that the laws of physics has a very low information content, since their algorithmic complexity can be characterized by a computer program less than a thousand characters [20]. In a personal

communication, Chaitin wrote [21]: "My paper on physics was never published, only as an IBM report. In it I took: Newton's laws, Maxwell's laws, the Schrödinger equation, and Einstein's field equations for curved spacetime near a black hole, and solved them numerically, giving 'motion-picture' solutions. The programs, which were written in an obsolete computer programming language APL2 at roughly the level of Mathematica, were all about half a page long, which is amazingly simple". Now one may estimate the complexity of a page as approximately $2 \cdot 10^3$ bit, since the average rate of information processing in reading is about 50 bit·s^{-1} [22] and so reading 1,5 pages in one minute the information content of a page is about 10^3 bit. In this way we obtain that the algorithmic complexity of physical equations is surprisingly low, $I_{\text{algorithmic}}$(physical equations) ~ 10^3 bit. Certainly, the observed flow of environmental information is enormous, but it is morphological information, and, apparently, it may arise from a much smaller algorithmic complexity through self-organization [23]. Now since we cannot expect that Big Bang (or recycling) cosmological models obtained initial conditions corresponding to an algorithmic complexity higher than the algorithmic complexity of physical laws themselves, we can estimate that the complexity measure of physics, initial and boundary conditions and physical equations included, is also about I(physics) ~ 10^3 bit.

This means that there is a much larger complexity jump between the early Earth without life and the first bacteria (from 10^3 bit to $4 \cdot 10^6$ bit, a jump of $J_1(10^8$ years) ~ $4 \cdot 10^3$, within about 10^8 years) than between the first bacteria and humans (from $4 \cdot 10^6$ bit to $6 \cdot 10^8$ bit, a jump of $J_2(4 \cdot 10^9$ years) ~ 150, during $4 \cdot 10^9$ years). This fact seems strange, since chemical abiogenesis may be thought as apparently unable to accelerate the evolution of complexity much faster than life itself. The question inevitable arises: How could chemical evolution reach a twenty-seven times higher increase in complexity within a forty times shorter time period, than life, if one would expect that biological complexity increase should be relatively (much) faster? This is the problem what we count as the problem of abiogenesis.

QUANTITATIVE RELATION BETWEEN GENOMIC AND DYNAMIC INFORMATION

Maynard Smith and Szathmáry [19, p.5] noted that the number of base pairs of the coding DNA is a measure of genomic complexity that makes sense, but what these numbers tell us about structural and functional complexity is very limited. It is a general view that DNA contains the information necessary to govern biological organization, e.g. [24 – 26]. The DNA stores information that controls all cellular processes [5].

Now the requirement that the DNA information $I_2 \sim 10^9$ bit should control $\dot{I}_1 \sim \dot{I}_{\text{biochem}} > 2 \cdot 10^{21}$ bit·s^{-1} can be satisfied only if we allow that in every time-steps the activation state of any base pairs of the DNA may change. Indeed, in order to regulate and control all the cellular reactions, DNA has to represent functional information. It was Abel [27] who introduced functional sequence complexity (FSC) which is a succession of algorithmic selections leading to function, besides the random sequence complexity (RSC) that can be simplistically defined as a mathematical function of the number of equiprobable potential alphanumeric symbols that could occupy each locus times the number of loci in that sequence of symbols and the ordered sequence complexity (OSC) which is exampled by polymers such as polysaccharides. Bits of functional information represent binary choices at successive algorithmic decision nodes. Algorithms are processes that produce a needed result, whether it is computation or the end products of biochemical pathways. Such strings of decision node selection are anything but random, and they are certainly not self-ordered by redundant cause-and-effect necessity. Abel [27] pointed out that questions relating to the origin of FSC are among the most difficult in biology, if not all science. If one would ask, does the FSC

originates from OSC and RSC, the best answer would probably be the slang expression: "No way!". The genetic information content of DNA does not originate from the chemical sequence of amino acids. Instead, FSC can only be quantified in its relation to biological functions actualized in the instantaneous internal and external environment. This is why FSC (and its counterpart, our dynamic complexity measure $\dot{I}_{\text{biochem}}$) is not highly redundant. Abel [27, p.65] adds: "There is a cybernetic aspect of life processes that is directly analogous to that of computer programming." We may realize that computer programming represents a yet deeper complexity level than the algorithmic complexity corresponding to the memory capacity. Analogously, the DNA complexity ($I_2 \sim 10^9$ bit) has to correspond to a deeper level of complexity than the algorithmic level of the brain (corresponding to $I_1 \sim 10^{15} - 10^{17}$ bit), and in this way we obtained a solution to the memory-DNA problem. Actually, if it is the DNA that plays the dominant role in governing cellular chemical reactions, it has to couple, coordinate and determine the timing of chemical reactions. This means that DNA corresponds to the deepest complexity level of the organism where the coordinating cellular reactions, what are themselves governed by algorithms related to couplings of chemical reactions, occurs. The complexity level of DNA corresponds to the regulation of the algorithmic complexity of cellular reaction pathways; therefore the genetic complexity is deeper than the algorithmic complexity of the memory.

Requiring that the static but deeper-than-algorithmic complexity of DNA is expressed through the mediation of activations of the $N_{\text{bp}}(\text{DNA}) \sim 10^9$ base pairs of the DNA, its information measure $I_2 \sim 10^9$ bit have to correspond to $\dot{I}_1 \sim \dot{I}_{\text{biochem}} > 2 \cdot 10^{21}$ bit·s^{-1}, and so from this requirement we became able to determine the length of the time necessary to activation of base pairs as $\Delta t \approx 4{,}2 \cdot 10^{-13}$ s. Actually, this timescale may be realistic for light-induced transfer of electrons Δt(electronic transitions) $\approx 10^{-12}$ s [28, p.6, Figs. 1-7]. This physical requirement seems to fit well with activation timescales. In this way, we obtained a conversion between the different forms of information, of its static and dynamic forms, converting $I_2 \sim 10^9$ bit to $\dot{I}_1 \sim \dot{I}_{\text{biochem}} > 2 \cdot 10^{21}$ bit·s^{-1}. It is this manner by which the DNA can fulfill the natural requirement to be differently activated in the different cells, if necessary, in each timesteps. In this way, DNA becomes able to supply the task of timing, to determine which chemical reactions should occur in the next timestep. Certainly, the DNA cannot do the timing alone, and its activity should be coherent with cellular organization supplying the necessary chemicals in the necessary places in the right moments, utilizing also a significant part of their thermodynamic capacities. But, in the way we obtained, the dynamic DNA can still preserve its key role to allow genetic control over the cellular reactions.

THE INFORMATION PARADOX: AN APPARENT CONFLICT BETWEEN PHYSICS AND BIOLOGY

While the morphological information of a circle is enormous, its algorithmic complexity is minuscule. The basic importance of the fact that simple rules may govern the appearance of high phenomenal complexity are already recognized in self-organizing systems, computer games, cellular automata [23], and it is widely thought that life's apparent complexity may appear as a product of certain yet-to-be-discovered, presumably simple physical rules.

Now Ashby's Law [11] states that "The variety of outputs of any deterministic physical system cannot be greater than the variety of inputs; the information of output cannot exceed the information already present in the input." In accordance, Kahre's "Law of Diminishing Information" reads: Compared to direct reception, an intermediary can only decrease the amount of information [29, p.14]. Moreover, it is a widely held view nowadays that the chain of physical causes forms a closed circle. The hypothesis of the causal closure of the physical

[30] maintains (roughly) "that for any event E that has a cause we can cite a physical cause, P, for its happening, and that citing P explains why E happened". Therefore, not only Ashby's and Kahre's laws but the causal closure thesis is in conflict with the complexity measures found in physics and in biology. Now if the algorithmic complexity of one human brain is already around $I_1 \sim 10^{15} - 10^{17}$ bit, the information paradox consists in the fact that the information content of physics is I(physics) ~ 10^3 bit while that of the whole living kingdom is $I_4 = I$(biology) >> $I_3 = I_{algorithmic}$(one human organism) > $I_1 = I_{algorithmic}$(brain) ~ $10^{15} - 10^{17}$ bit. Taking into account also that physics is hopelessly far from being able to cope with the task to govern even one human person's biological activity $\dot{I}_1 \sim 2 \cdot 10^{21}$ bit per second, it becomes clear that at present, modern cosmological models' algorithmic complexity is much less than the above obtained complexity measures characterizing life. Actually, the origin of biological information is widely thought to be in evolutionary biology as arising from the environment through natural selection. The problem is now: where does the high algorithmic information of the environment comes from, in a universe the behavior of which – as it is widely assumed – can be described by physical laws corresponding to a mere I(physical laws) < 10^3 bit? In other words: If the genome obtains its high information content from the environment, as it is assumed in evolutionary studies nowadays, how this environment could achieve an algorithmic complexity of biological size if it should correspond to the much lower algorithmic complexity measure of physics? We may realize that we are faced with a complexity paradox corresponding to the relation of physics to biology. Apparently, the informational resources of physics are far lower than the complexity measures of the brain and, in general, living organisms.

Certainly, the thermodynamic capacity of modern cosmological models allows the development of an information generation process producing information in an astronomical rate. Our Sun has a luminosity L(Sun) ~ $4 \cdot 10^{26}$ $J \cdot s^{-1}$, corresponding to a thermodynamic information capacity of $\dot{I}_3 = \dot{I}$(Sun) ~ 10^{38} $bit \cdot s^{-1}$ [31, p.183]. There are N(stars) ~ $7 \cdot 10^{22}$ stars in the observable universe, offering an information flux capacity $\dot{I}$(stars) ~ 10^{61} $bit \cdot s^{-1}$. What percentage is utilized from this astronomically high information capacity in the universe? What kinds of agents are necessary to utilize the thermodynamic capacity of the universe? The problem is: how this vast thermodynamic information capacity is utilized in the universe in the nowadays widely assumed absence of cosmic life? One could expect that the thermodynamic capacity to generate information can be utilized only by symbol-generating agents capable of generating, recognizing, handling and accumulating information. Again, it seems that abiotic processes should generate and accumulate information – in sharp contrast with the fundamental law of cybernetics (Ashby's Law [11]); with the fundamental law of the mathematical theory of information, the Law of Diminishing Information [29]; and with the dogma of the causal closure of the physical.

In this context, an example may be enlightening. Hoyle [32] pointed out that to solve the Rubik cube by one random step in every second, it would take $1{,}35 \cdot 10^{12}$ years. The chance against each move producing perfect color matching for all the cube's faces is about $5 \cdot 10^{19}$ to 1. Now if an intelligence is present, telling after each move if it is successful or not, reckoning 1 minute for each successful move and, say, 120 moves to reach the solution, the solution of the same Rubik cube may be reached within 2 hours. Certainly, the abiotic processes are not completely random – modifying the success ratio with and without intelligence from about 10^{16} to somewhat lower.

We point out that the production of algorithmic information seems not to be possible from phenomenal information arriving through the outer senses. The generation of genetic information from sensory data seems to be even more implausible. Although it is shown [23] that simple algorithmic rules can produce high amount of phenomenal complexity, certainly,

the opposite process, the production of algorithmic information from phenomenal information is not shown to be possible yet, especially not in the absence of agents that are able to follow their own interests and not merely the laws of physics. One cannot expect that rules of games will develop from mere aggregate of phenomenal data. Mathematical operations like addition and multiplication does not arise from numbers alone. If laws could develop from aggregation of phenomena, it would be nomic emergence. Nomic emergence is something completely different from property emergence. Not only a different level of phenomena should emerge, but casual laws should also emerge simultaneously. But there is no basis "to accept emergent causal powers that magically emerge at a higher level of which there is no accounting in terms of lower-level properties and their causal powers and nomic connections" [33]. Nomic connections are based on generation of algorithmic complexity corresponding to the emergent laws. But no algorithmic complexity comes for free. Laws cannot be generated in the universe of phenomena. Although chemical symbiosis may be present in abiogenesis [4, p.35], even if it could increase the algorithmic complexity of chemical information (a process that already requires the existence of agents – and agents should be the output of chemical evolution and not its input), it could not generate genetic complexity, since the ordered sequence complexity, as Abel [27] had shown, is much simpler than the functional one and functional sequence complexity cannot be produced form ordered sequence complexity.

There exists a popular example of monkeys that can type Shakespeare's complete oeuvre on a typewriter. Actually, to type only one sentence from the Hamlet, consisting of 40 letters, each selected from 30 possibilities, it would be necessary to realize $30^{40} \sim 10^{59}$ trials. Let us assume that we have ten billion monkeys – that is, rather more monkeys than there are currently people in the world. And let us imagine each monkey hits one key per second. Let us further assume that they never stop to sleep or eat or anything else. It will still take more than 10^{49} seconds before one of the monkeys has the luck to hit on the right sequence. Now one year is about 32 million seconds, so it will take our world population of monkeys about $3 \cdot 10^{41}$ years to get there. Now how would it be possible that the absence of monkeys and typewriters, corresponding to the case of chemical abiogenesis, would accelerate the process to write an amount of information corresponding to Shakespeare's whole Hamlet, within a mere 10^8 years? Certainly, one cannot expect that chemical evolution would be able to produce useful amount of genetic complexity in the absence of agents. Even in the presence of "inanimate agents" it seems highly implausible to expect that the accumulation rate of genetic information by chemical abiogenesis in an assumedly *physical* environment (information accumulation in physical systems is excluded by Ashby's Law, Kahre's Law and causal closure) could produce much higher jump in genetic information than the jump produced by life during its $4 \cdot 10^9$ years of evolution. Why should "inanimate agents", if they may exist at all, be much more efficient than living agents possessing much higher genetic complexity? Especially, if the number of regulatory genes grows approximately with the square of the total number of genes, as it is shown by genetic experiments [34].

SOLUTIONS FOR THE INFORMATION PARADOX

Within the present state of biology, it seems that there are only two ways out of the informational paradox of biology. The established way is that of the abiogenesis [19, 35, 36]. They realised a foundational work concerning the details of the chemical evolutionary process. The chemoton theory has the ambitious aim to follow chemical evolution until life's development. We think that chemoton theory is basic and will remain fundamental even when we turn our attention to a complementary aspect relative to chemical evolution: to the quantitative understanding of the origin of genetic information. In the light of the results of this paper, it seems that the problems of chemical evolution are larger now than the problem

of protein folding was fifty years ago. In this case, we find it reasonable to shed some light on another important aspect of the problem that is the weak point of evolutionary theories: how to handle complexity. In this paper we try to characterize by numerical measures the process coupling autocatalytic cycles into hypercycles and co-operating hypercycles and representing genetic information. Only further developments of chemical evolution theories (see [36, 37]) may help to understand chemical evolution at the molecular level. With the present paper we would like to present a complementary global picture which may shed some light to the quantitative aspects of complexity at the algorithmic and genetic levels where coupling process occur, in the hope that the simultaneous development of progress from the aspects of molecular level and that of the global level may facilitate to bridge the gap between these levels much earlier than to proceed in one direction only.

Our proposal for answering the information paradox concerns the nature of first principles. Let us consider the important point that when complex forms develop from simple rules in self-organization, a static algorithmic complexity generates developing structures, a dynamic information of the morphological level. If the algorithmic complexity of the "simple laws" of our real world has to be much higher than that of the physical laws, then certainly we will need complex rules instead of simple rules producing biological blueprints. But perhaps these complex rules (together with their high algorithmic information content) may arise from simple laws themselves – again from a deeper level of information. The development of physics in the twentieth century had shown that physical laws arise from a much deeper concept: the concept of the first principle of physics – the action principle. Apparently, there is an intimate connection between the three levels of biological complexity, the morphological, algorithmic and genetic levels, and the three levels of science: the phenomenal level of observable phenomena, the level of laws, and the level of the first principle of physics, the action principle [38]. Morphological complexity seems to be related to the phenomenal, algorithmic complexity to the nomic level, and genetic complexity to the principal level.

The action principle is formulated by Feynman's path-integral method as arising from virtual processes covering a multitude of possible pathways and the resulting physical path will be the simple sum or integral of these paths. The integral form of the action principle contains a non-negligible surplus over its formulation in differential equations. Differential equations need definite initial conditions, while the integral formalism – virtually – includes informative interactions with a large set of the environment. Integral principles are independent from coordinates, and therefore they can cope with time-dependent boundary conditions as well. The apparent teleological behavior of living organisms may correspond to computational processes determined at the organism level, where the organism acts as an agent, following its own interests and biological needs like survival. Once the biologically favourable endpoint of a biological process is prescribed by the organism, the biological problem will be simplified, and with the help of the action principle of physics it becomes possible to determine the trajectory to be followed, and the organism can realize the biological needs by rearranging its internal physical environment.

Therefore, the solution we offer as an alternative to solve the information paradox of physics and biology is to allow agents to follow their biological needs. Agents had been introduced into biology, and they are indicated to be present already at the subcellular level [39, 40]. Szathmáry [36, 41] pointed out that there are some enzymes can adapt different shapes: in A it works as an enzyme, in B it does not. The enzyme function of hormones like adrenalin is not determined by physics at all. Nothing makes the structure of adrenalin to act like it does. Actually, adrenalin could be used as the opposite effect. An arbitrary coupling occurs between the enzyme and its function, and it is just such arbitrary coupling that is at the heart of symbolic communication that appears at agents (see also [42]).

If our results will be confirmed, it will turn out that biology cannot be reduced simply to physics, since its genetic, algorithmic and symbolic information content is much higher than that of physics. Our proposal not only allows biology to follow its own, and, necessarily, autonomous first principle not derivable from physics, but allows also to approach biology from a viewpoint that can make theoretical biology to develop into a science with exactness almost reaching the exactness of physics. The first principle of biology (the Bauer principle [43]) may be only one step only beyond the action principle of physics, and can be understood as its generalization. We propose that biological organization manifests itself in a way that decisions are made at the biological level first, and these decisions determine the endpoints of physical processes to be reached. By our proposal, biological laws can harness physical laws with the help of their enormous and effective information content.

ACKNOWLEDGMENT

It is a great pleasure to acknowledge about the inspirations, encouragement and English lecturing of a previous version of this paper to my friend, Jean Drew.

REFERENCES

[1] Stripling, J.: *Introduction to Cognitive Science*.
Neuroscience, 2004,
http://comp.uark.edu/~jstripli/CogSci-JS-L1-web.pdf,

[2] Smith, T.E.: *Molecular Cell Biology*.
In: Devlin, T.M., ed.: *Textbook of Biochemistry*. 4th edition. Wiley-Liss, New York, Ch. 22, p.921, 1997,

[3] Koch, C. and Laurent, G.: *Complexity and the Nervous System*.
Science **284**, 96-98, 1999,

[4] Maynard Smith, J.: *The Concept of Information in Biology*.
Philosophy of Science **67**, 177-194, 2000,

[5] Woski, S.A. and Smith, F.J.: *DNA and RNA: Composition and Structure*.
In: Devlin, T.M., ed.: *Textbook of Biochemistry*. 5th edition. Wiley-Liss, p.28, 2002,

[6] Claverie, J.-M.: *What if there are only 30,000 human genes?*
Science **291**, 1255-1257, 2000,

[7] Allen, F.; et al.: *Blue Gene: A vision for protein science using a petaflop supercomputer*.
IBM Systems Journal **40**(2), 2001,

[8] Kornberg, A.: *For the love of enzymes*.
Harvard University Press, Cambridge, p.65, 1989,

[9] Vander, A.J.; Sherman, J.H. and Luciano, D.S.: *Human Physiology*. 5th edition.
McGraw-Hill Publ. Co., New York, p.95. 1990,

[10] Martinás, K.: *Entropy and information*.
World Futures **50**, 483-493, 1997,

[11] Ashby, W. R.
In: Foster, H.V. and Zopf, G.W., eds.: *Principles of Self-Organization*. Pergamon Press, Oxford, 255, 1962,

[12] Bialek, W. and Setayeshgar, S.: *Physical limits to biochemical signalling*.
Proc. New York Acad. Sci. 102, 10040-10045, 2005,

[13] Davies, P.: *Physics and Life*.
In: Chela-Flores, J.; Owen, T. and Raulin, F., eds.: *The First Steps of Life in the Universe*. Kluwer Academic Publishers, Dordrecht, 2001,

[14] Aebersold, R.: *Molecular Systems Biology: a new journal for a new biology*?
Molecular Systems Biology **1**, 1, 2005,

[15] Petricoin, E.; Wulfkuhle, J.; Espina, V. and Liotta, L.A.: *Clinical proteomics: Revolutionizing Disease Detection and Patient Tailoring Therapy*.
Journal of Proteome Research **3**, 209-217, 2004,

[16] Sherwood, L.: *Human Physiology. From Cells to Systems*. 2nd edition.
West Publ. Co., Minneapolis-St. Paul-New York, p.7, 1993,

[17] Aoki, I.: *Entropy production in living systems: from organisms to ecosystems*.
Thermochimica Acta **250**, 359-370, 1995,

[18] Kimura, M.: *Natural Selection as a Process of Accumulating Genetic Information in Adaptive Evolution*.
Genetical Research **2**, 127-140, 1961,

[19] Maynard Smith, J. and Szatmáry, E.: *The Major Transitions in Evolution*.
W. H. Freeman – Spektrum, Oxford, 1995,

[20] Chaitin, G.J.: *An APL2 gallery of mathematical physics – a course outline*.
Proc. Japan '85 APL Symposium, Publ. N:GE18-9948-0 IBM Japan, 1-26, 1985,

[21] Chaitin, G.J. *Personnal communication*. 2004,

[22] Breuer, H.: *Informatik*.
DTV-Atlas zur Informatik. Deutscher Taschenbuch Verlag. Kg, München, p.13, 1995,

[23] Wolfram, S.: *A NewKind of Science*.
Wolfram Media, 2002,

[24] Crick, F.: *Central dogma of molecular biology*.
Nature **227**, 561-563, 1970,

[25] Campbell, N.A.: *Biology*. 4th edition.
The Benjamin/Cummings Publishing Company, Menlo Park, p.4, 1996,

[26] Miyano, S.: *Gene Networks for Systems Biology*.
http://www.inrialpes.fr/helix/people/sagot/bertinoro2005/bertinoro.html,

[27] Abel, D.L.: *Is life reducible to complexity*?
In: Pályi, G.; Zucchi, C. and Cagliotti, L., eds.: *Fundamentals of Life*. Elsevier, Paris, 2002,

[28] Stryer, L.: *Biochemistry*.
W.H. Freeman & Comp., p.6 & Figs 1-6, 1995,

[29] Kahre, J.: *The Mathematical Theory of Information*.
Kluwer Academic Publishers, Boston/Dordrecht/London, 2002,
http://www.matheory.info/chapter1.html,

[30] Cameron, R. J.: *Teleology in Aristotle and Contemporary Philosophy of Biology: An Account of the Nature of Life*. PhD Thesis.
University of Colorado, p.244, 2000,
http://hosting.uaa.alaska.edu/cameron/research/diss/diss.pdf,

[31] Tribus, M. and McIrvine, E.C.: *Energy and Information*.
Scientific American **225**(3), 179-188, 1971,

[32] Hoyle, F.: *The Intelligent Universe*.
Michael Joseph Ltd., London, 1983,
see also http://en.wikipedia.org/wiki/Fred_Hoyle,

[33] Kim, J. et al., eds.: *Supervenence and Mind*.
University of Cambridge, Cambridge, p.326, 1993,

[34] Schuster, P.: *The disaster of central control. An impressive example from nature*.
Complexity **9**, 13, 2004,

[35] Gánti, T.: *The Principles of Life*. With a commentary by James Griemeser and Eörs Szathmáry.
Oxford University Press, Oxford, 2003,

[36] Szathmáry, E.: *In search of the simplest cell*.
Nature **433**, 469-470, 2005,

[37] Schuster, P.: *From self-organization to evolution of RNA molecules: The origin of biological information*.
Solid State Phenomena **27**, 97-98, 2004,

[38] Grandpierre, A.: *The Dynamics of Time and Timelessness: Philosophy, Physics and Prospect for our Lives*.
In: Buccheri, R.; Saniga, M. and Stuckey, W.M., eds.: *The Nature of Time: Geometry, Physics and Perception*. NATO Science Series II. Mathematics, Physics and Chemistry, Vol. 95, Kluwer Academic Publishers, Dordrecht-Boston-London, 383-392, 2003,

[39] Monod, J.: *Chance and Necessity*.
Vintage, 1972,

[40] Emmeche, C.: *The Agents of Biomass*.
In: Jürgensen, A. and Ohrt, C., eds.: *The Mass Ornament. The mass phenomenon at the turn of the millenium*. Kunsthallen Brandts Klædefabrik, Odense, 64-79, 1998,
http://www.nbi.dk/~emmeche/p.biosem.html,

[41] Szathmáry, E.: *Information and information-processing in biological systems*.
talk at the European Forum, Alpbach, 2005,

[42] Abel, D.L. and Trevors, J.T.: *Three subsets of sequence complexity and their relevance to biopolymeric information*.
Theoretical Biology & Medical Modelling **2**, art. no. 29, 2005,
http://www.tbiomed.com/content/2/1/29,

[43] Bauer, E.: *Theoretical Biology*. In Hungarian.
Akadémiai Kiadó, Budapest, 32-78, 1967,

KOMPLEKSNOST, INFORMACIJA I BIOLOŠKA ORGANIZACIJA

A. Grandpierre

Opservatorij Konkoly Madžarske akademije znanosti
Budimpešta, Madžarska

SAŽETAK

Raširena konfuzija oko koncepta i prirode kompleksnosti, informacije i biološke organizacije motivirala nas je na koordinarana konceptualna razmatranja kvantitativnih mjera prikladnih za izdvajanje značajki biološke kompleksnosti. Kvantitativne mjere algoritamske kompleksnosti superračunala poput *Blue Gene/L* su uspoređene s kompleksnossti mozga. Pokazujemo da je i računalu i mozgu pridružena fundamentalnija, dinamička mjera kompleksnosti koja odgovara broju operacija u sekundi. Noviji uvidi upućuju na to da izvor kompleksnosti može biti u jednostavnosti na višoj razini, što odgovara algoritamskoj kompleksnosti. Ashbyev zakon, Kahreov zakon i kauzalna zatvorenost fizikalnih sustava isključuju nastajanje informacija. Budući da genetske informacije predstavljaju upute, nailazimo na paradoks da je algoritamska kompleksnost fizike znatno manja od kompleksnosti uputa u ljudskoj DNK: $I_{algoritam}$(fizika) ~ 10^3 bit << I_{upute}(DNK) ~ 10^9 bit. Analizirajući genetsku kompleksnost dobivamo da genetska informacija odgovara stupnju kompleksnosti višem od algoritamske, što dodatno pojačava informacijski paradoks. Naposljetku, pokazujemo da razriješenje informacijskog paradoksa može biti ili u kemijskoj evoluciji nasljeđa u abiogenezi, ili u postojanju autonomnog biološkog principa koji omogućava generiranje informacija van fizike.

KLJUČNE RIJEČI

razine kompleksnosti, računalo i mozak, algoritamska kompleksnost, kompleksnost i informacija, temeljni informacijski paradoks prirodnih znanosti