Biology as an Axiomatic Process

J. Pozas, Dec. 2017

Abstract

The replication mechanisms of living beings can be compared with the self-replication of automatons in the context of computability theory. In particular, DNA replication, analyzed from the perspective of the recursion theorem, indicates that its replication structure goes beyond biology and the quantum mechanisms that support it.

Thus, according to the recursion theorem, the copy or replication of an object must be structured in at least two independent axiomatic processes. On the other hand, the central dogma of molecular biology clearly shows this behavior, with a sequence of independent biological processes forming a recursive processing structure. These processes are a consequence of the physicochemical properties of molecules and therefore in the quantum properties of matter, allowing DNA replication and protein synthesis and are the foundation of what we know as living beings.

This parallelism clearly indicates that biological processes obey abstract rules in which the axiomatic process of information is fundamental. For this reason, we can consider biology as a paradigmatic example of the principle of reality and therefore of the indistinguishability between reality and information.

Keywords: Recursion theorem, central dogma of molecular biology, DNA replication, axiomatic processing, principle of reality.

The biological process: A brief presentation

Physical chemistry establishes the principles by which atoms interact with each other to form molecules. In the inorganic world the resulting molecules are relatively simple, not allowing establishing a complex functional structure. On the other hand, in the organic world, molecules can be made up of thousands or even millions of atoms and have complex functionality. It highlights what is known as molecular recognition, through which the molecules interact with each other selectively and which is the basis of biology.

Molecular recognition plays a fundamental role in the structure of DNA, in the translation of the genetic code of DNA into proteins and in the biochemical interaction of proteins, which ultimately form the basis on which living beings are based.

The detailed study of these molecular interactions makes it possible to describe the functionality of the processes, in such a way that it is possible to establish formal models, to such an extent that they can be used as a computing technology, as is the case of DNA-based computing.

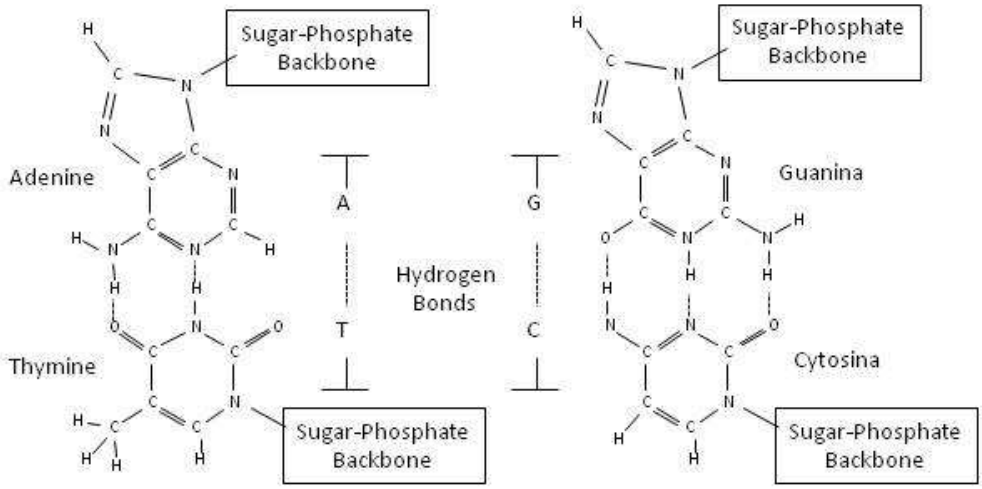
From this perspective, this allows us to ask if the process of information is something deeper and if in reality it is the foundation of biology itself, according to what is established by the principle of reality [1].

For this purpose, this section aims to analyze the basic processes on which biology is based, in order to establish a link with axiomatic processing and thus investigate the nature of biological processes. For this, it is not necessary to describe in detail the biological mechanisms described in the literature. We will simply describe its functionality, so that they can be

identified with the theoretical foundations of information processing. To this end, we will explain the mechanisms on which DNA replication and protein synthesis are based [2].

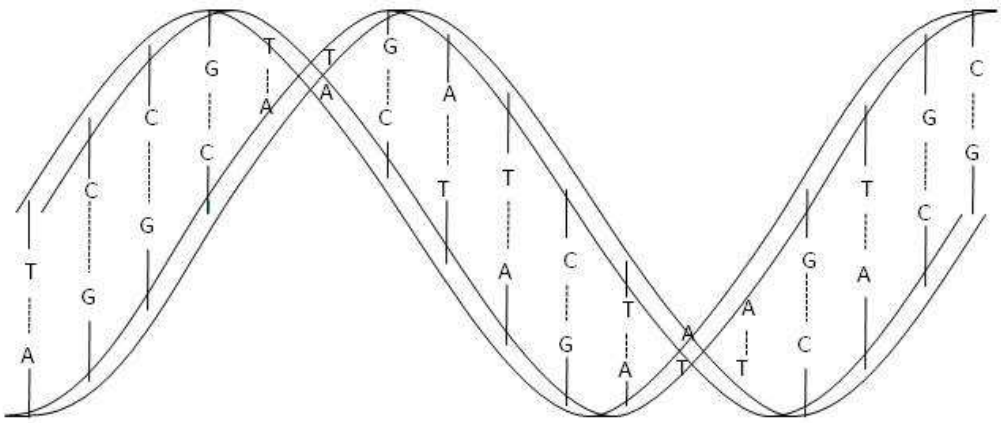
DNA and RNA molecules are polymers formed from the ribose and deoxyribose nucleotides, respectively, bound by phosphates. On the basis of this nucleotide chain, one of the four possible nucleic acids can be linked. There are five different nucleic acids, adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U). In the case of DNA, nucleic acids that can be coupled by covalent bonds to nucleotides are A, G, C and T, whereas in the case of RNA they are A, G, C and U. As a consequence, molecules are structured in a helix shape, fitting the nucleic acids in a precise and compact way, due to the shape of their electronic clouds [3].

The helix structure allows the nucleic acids of two different strands to be bound together by hydrogen bonds, forming pairs A-T and G-C in the case of DNA, and A-U and G-C in the case of RNA, as shown in the following figure.



Base-pairing of nucleic acids in DNA

As a result, the DNA molecule is formed by a double helix, in which two chains of nucleotides polymers wind one on top of the other, remaining together by means of hydrogen bonds of nucleic acids. Thus, each strand of the DNA molecule contains the same genetic code, one of which can be considered the negative of the other.



Double helix structure of DNA molecule

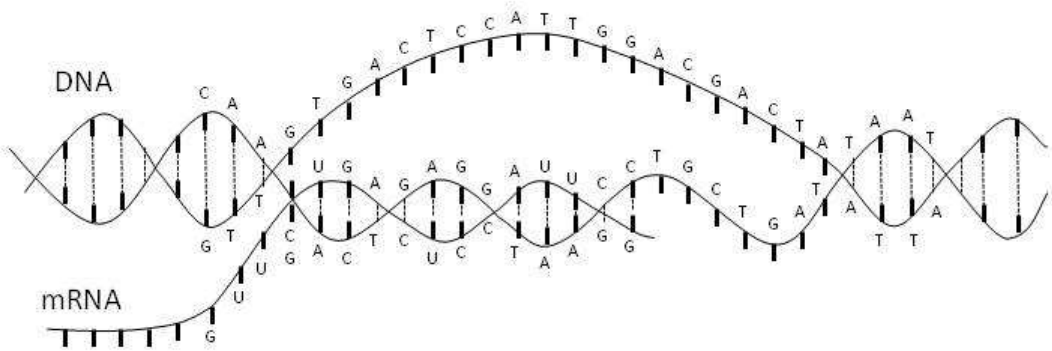
The genetic information of an organism, called a genome, is not contained in a single DNA molecule, but is organized into chromosomes. These are made up of DNA strands bound together by proteins. Thus, in the case of humans, the genome is formed by 46 chromosomes, and so, the number of bases in the DNA molecules that compose it being about $3x10^9$. Since each base can be encoded by means of 2 bits, the human genome, considered as an object of information, is equivalent to $6x10^9$ bits.

The information contained in the genes is the basis for the synthesis of proteins, which are responsible for executing and controlling the biochemistry of living beings. The proteins are formed by the bonding of amino acids, through covalent bonds, which is done from the sequences of the bases contained in the DNA. The number of existing amino acids is 20 and since each base codes 2 bits, 3 bases (6 bits, 64 combinations) are necessary to be able to code each one of the amino acids. This means that there is some redundancy in the assignment of base sequences to amino acids, in addition to control codes for the synthesis process (Stop), as shown in the following table.

1st	2nd	3rd	Amino Acid	1st	2nd	3rd	Amino Acid	1st	2nd	3rd	Amino Acid	1st	2nd	3rd	Amino Acid
A	A	A	Lys	G	A	A	Glu	C	A	A	Gin	U	A	A	Stop
A	A	G	Lys	G	A	G	Glu	C	A	G	Gin	U	A	G	Stop
A	A	C	Asn	G	A	C	Asp	C	A	C	His	U	A	C	Tyr
A	A	U	Asn	G	A	U	Asp	C	A	U	His	U	A	U	Tyr
A	G	A	Arg	G	G	A	Gly	C	G	A	Arg	U	G	A	Stop
A	G	G	Arg	G	G	G	Gly	C	G	G	Arg	U	G	G	Trp
A	G	C	Ser	G	G	C	Gly	C	G	C	Arg	U	G	C	Cys
A	G	U	Ser	G	G	U	Gly	C	G	U	Arg	U	G	U	Cys
A	C	A	Thr	G	C	A	Ala	C	C	A	Pro	U	C	A	Ser
A	C	G	Thr	G	C	G	Ala	C	C	G	Pro	U	C	G	Ser
A	C	C	Thr	G	C	C	Ala	C	C	C	Pro	U	C	C	Ser
A	C	U	Thr	G	C	U	Ala	C	C	U	Pro	U	C	U	Ser
A	U	A	Ile	G	U	A	Val	C	U	A	Leu	U	U	A	Leu
A	U	G	Met	G	U	G	Val	C	U	G	Leu	U	U	G	Leu
A	U	C	Ile	G	U	C	Val	C	U	C	Leu	U	U	C	Phe
A	U	U	Ile	G	U	U	Val	C	U	U	Leu	U	U	U	Phe

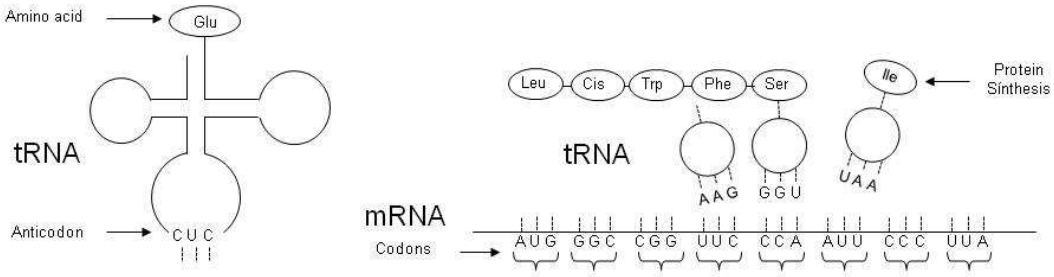
Translation of nucleic acids (Codons) to amino acids

However, protein synthesis is not done directly from DNA, since it requires the intermediation of RNA. This is called the translation process and involves two types of different RNA molecules, the messenger ARM (mRNA) and the transfer RNA (tRNA) [4]. The first step is the synthesis of mRNA from DNA. This process is called transcription, in such a way that the information corresponding to a gene is copied into the mRNA molecule, which is done through a process of recognition between the molecules of the nucleic acids, carried out by the hydrogen bonds, such as shows the following figure.



DNA transcription

Once the mRNA molecule is synthesized, the tRNA molecule is responsible for mediating between mRNA and amino acids to synthesize proteins, for which it has two specific molecular mechanisms. On the one hand, tRNA has a chain of three amino acids called anticodon at one end. On the opposite side, tRNA binds to a specific amino acid, according to the translation table of nucleic acid sequences into amino acids. In this way, tRNA is able to translate mRNA into a protein, as shown in the figure below.



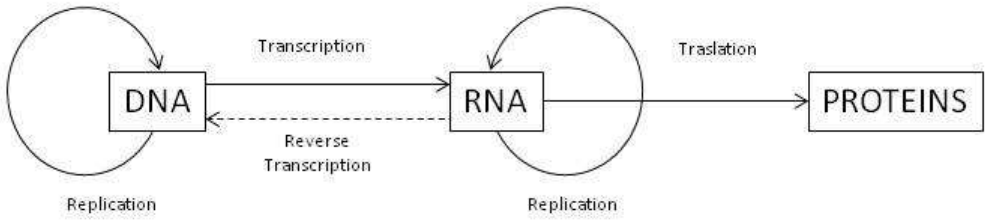
Protein synthesis (mRNA translation)

But perhaps the most complex process is undoubtedly DNA replication, so that each molecule produces two identical replicas. Replication is performed by unwinding each strand of the molecule and inserting the nucleic acid molecules on each of the strands, in a similar way to that shown in the mRNA synthesis. DNA replication is controlled by enzymatic processes supported by proteins [2]. Without going into detail and in order to show its complexity, the table below shows the proteins involved in the replication process and their role.

Enzime	Function in DNA replication
Helicase	Separates the two strands of DNA at the Replication Fork behind the topoisomerase
Polymerase	Catalyzes the addition of nucleotide substrates to DNA. Performs proof-reading and error correction
DNA Clamp	Prevents elongating DNA polymerases from dissociating from the DNA parent strand
SSB proteins	Prevent the DNA double helix from reannealing after DNA helicase unwinds it
Topoisomerase	Relaxes the DNA from its coiled nature
Gyrase	Relieves strain of unwinding by DNA helicase
Ligase	Reanneals the semiconservative strands and joins Okazaki Fragments of the lagging strand.
Primase	Provides a starting point for DNA polymerase to begin synthesis of the new DNA strand
Telomerase	Lengthens telomeric DNA

The role of proteins in the DNA replication process

The processes described above are defined as the central dogma of molecular biology and are usually schematically represented schematically as shown in the following figure. It also depicts the reverse transcription that occurs in retroviruses, which synthesizes a DNA molecule from RNA.



Central dogma of molecular biology

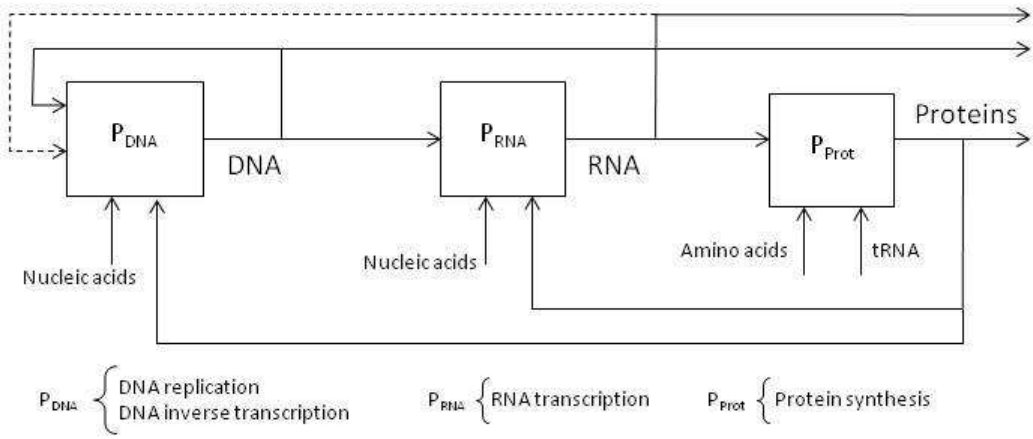
The biological process from the perspective of computability theory

Molecular processes supported by DNA, RNA and proteins can be considered from an abstract point of view as information processes. As a result, input statements corresponding to a language are processed resulting in new output statements. Thus, the following languages can be identified:

- DNA molecule. Sentence consisting of a sequence of characters corresponding to a 4-symbol alphabet.
- RNA molecule - protein synthesis. Sentence consisting of a sequence of characters belonging to a 21-symbol alphabet.
- RNA molecule-reverse transcription. Sentence composed of a sequence of characters belonging to a 4-symbol alphabet.

- Protein molecule. Sentence composed of a sequence of characters belonging to a 20-symbol alphabet.

This information is processed by the machinery established by the physicochemical properties of control molecules. To better understand this functional structure, it is advisable to modify the scheme corresponding to the central dogma of biology. To do this, we must represent the processes involved and the information that flows between them, as shown in the following block diagram.



Functional structure of DNA replication

This structure highlights the flow of information between processes, such as DNA and RNA sentences, where the functional blocks of information processing are the following:

- P_{DNA}. Replication process. The functionality of this process is determined by the proteins involved in DNA synthesis, producing two replicas of DNA from a single molecule.
- P_{RNA}. Transcription process. It synthesizes a RNA molecule from a gene encoded in DNA.
- P_{Prot}. Translation process. It synthesizes a protein from an RNA molecule.

This structure clearly shows how information emerges from biological processes, something that seems to be ubiquitous in all natural models and allows the implementation of computer systems. In all cases this capacity is finally supported by quantum physics. In the case of biology in particular, this is produced from the physicochemical properties of molecules, which are determined by quantum physics. Therefore, the information process is something that emerges from an underlying reality and ultimately from quantum physics. This is true as far as knowledge goes.

This means that, although there is a strong link between reality and information, information is simply an emerging product of reality. But biology provides a clue to the intimate relationship between reality and information, which are ultimately indistinguishable concepts [1]. If we look at the DNA replication process, we see that DNA is produced in several stages of processing:

DNA → RNA → Proteins → DNA.

We could consider this to be a specific feature of the biological process. However, computability theory indicates that the replication process is subject to deeper logical rules than the physical processes themselves that support replication. In computability theory, the recursion theorem determines that replication of information requires at least the intervention of two independent processes.

This shows that DNA replication is subject to abstract rules that must be satisfied not only by biology, but by every natural process. Therefore, the physical foundations that support biological processes must verify this requirement. Consequently, this shows that the information processing is essential in what we understand by reality.

The Recursion Theorem: A Review

Kleene's recursion theorem [5] is a fundamental result of computability theory, laying the foundations for the use of computable functions for its own definition. In this context, we are going to focus on a practical aspect of this theorem, known as "quine" or "SELF". A computer program is defined as SELF which, ignoring the input parameters, produces as output a copy of its own source code. For this reason, these programs can be defined as self-replicating programs.

Because of its structural proximity to the biological DNA replication model, the Turing machine model is best suited to analyze the recursion theorem and establish a link between the two models. A substantial part of the recursion theorem corresponds to self-reference, so that a program can obtain its own description and use it for its execution.

According to the recursion theorem, the Turing machine SELF must be structured in two steps by independent processes A and B [6]. The aim is that the execution of process A should produce a description of B and, on the other hand, the execution of process B should produce a description of process A. The tasks of processes A and B are basically the same, but they must be carried out differently, i. e. the processes must be independent. Otherwise, the SELF definition would be circular and therefore logically inconsistent. Without going deeper into the implementation of processes A and B, the important thing in this context is that both processes must be independent. This means that they have an independent axiomatic definition, as can be deduced from the recursion theorem.

If we analyze the central dogma of biology, from the point of view of the recursion theorem, we see that obtaining copies of DNA is structured in three independent processes. So, it verifies, as cannot be otherwise, the recursion theorem. However, the copy of DNA is structured in three different phases: DNA replication, RNA transcription and protein synthesis. The logical interpretation is that a given process can be structured in several sub-processes. Thus, proteins are obtained from DNA in a structured way, so that the information encoded in each gene is transcribed and translated by RNA-specific molecules. For this reason, transcription and translation can be interpreted as sub-processes of a single process and henceforth we can consider them simply as a single process of protein transcription.

Consequently, processes A and B of the recursion theorem can be identified with the protein transcription process and the DNA replication process. But here a fundamental question arises regarding the independence of processes A and B. In the logic model of the Turing machine: What makes these processes independent? The answer lies in the control unit of each of the Turing machines that support processes A and B and is based on their own axiomatic definition.

In the case of the biological model, processes are based on a physical reality determined by the distinct functionality of the molecules that intervene in them and that, as far as knowledge reaches, lies in quantum physics.

What is clear is that the recursiveness of a process is an abstract concept that goes beyond any real process or the perception of reality. This means that the functional structure created by nature to obtain copies of DNA is not something merely casual or a whim of nature. On the contrary, this structure follows a formal rule that all copying processes must comply with, analogously to what happens in the implementation of computer viruses [7] based on self-replication theory. Consequently, the formal copying process requires the definition of two independent axiomatic processes, so that the physicochemical and quantum nature of the biological processes must ultimately be supported by an underlying axiomatic reality.

Conclusions

The functional structure of the central dogma of biology is a case study that allows us to contrast the principle of reality and therefore the indistinguishability between reality and information, as has already been postulated [1]. To this end, the recursion theorem has been used, which transcends the physical processes that support biological processes, so the conclusions reached are independent of biological processes.

The use of the recursion theorem, however, opens up new analysis expectations for the non-cloning theorem established in the field of quantum physics. This indicates that, if the theorem of non-cloning is formally correct, the nature of quantum physics goes beyond computable reality, so that it is not possible to define independent computable processes that can support the cloning of information in the quantum reality layer.

[1] J. Pozas, «Space-time: Reality and information as indistinguishable entities,» May. 2017.

[2] A. Griffiths, S. Wessler, S. Carroll and J. Doeblye, Introduction to Genetic Analysis, W. H. Freeman & Co, 2010.

[3] L. Allison, Fundamental Molecular Biology, John Wiley & Sons Inc, 2011.

[4] B. Alberts, Molecular Biology of the Cell, Garland Science, 2014.

[5] S. C. Kleene, Introduction to Metamathematics, Van Nostrand, Princeton, NJ, 1952.

[6] M. Sipser, Introduction to the Theory of Computation, Wadsworth Publishing Co Inc, 2012.

[7] E. Filiol, Computer Viruses: from Theory to Applications, Springer, 2005.