First International Conference in Code Biology – Paris 2014

Introduction to Code Biology

Marcello Barbieri
Dipartimento di Morfologia ed Embriologia
Via Fossato di Mortara 64a, 44121 Ferrara, Italy
brr@unife.it

As an introduction to the Conference, I would like to discuss three general issues.

Issue 1

The first point is the place that codes have in modern biology. Let us keep in mind that the great biological theories – on evolution, heredity, embryonic development, etc. – have all been built *before* the discovery of the genetic code, and this gives us a problem: how did people account for the genetic code in a theoretical framework that does not contemplate the existence of codes?

The explanation that was given at the time, and that is still widely accepted today, is that the genetic code is a set of *constraints*. This idea was suggested by the distinction that exists in physics between the *laws of nature* and the boundary conditions that are collectively known as 'constraints'. All planets, for example, are formed according to universal physical laws, and yet they are all different because their individual features were determined by the particular local constraints that affected their historical development.

The conclusion that the genetic code is a set of constraints is formally correct because a code is indeed a set of limitations since it consists of a small number of rules that are selected from a virtually unlimited number of possibilities. On top of that, the idea was attractive because physical constraints cannot be changed and this appeared to explain the fact that the genetic code has been 'frozen' since the origin of life.

There is no doubt that physical constraints impose limitations on living systems and it is true that coding rules are constraints, but there are two other arguments that must be taken into account.

(1) The first is that the rules of the genetic code are *biological* constraints, not *physical* ones. They are *biologically generated* rules and in no way can be assimilated to physical constraints. This is because the genes of the genetic code are constantly subject, like all other genes, to mutation and neutral drift. They are in a continuous state of flux and the fact that they have been highly conserved in evolution means that there is a biological mechanism that *actively and continuously* restores their original structure. The conservation of the genetic code, in other words, is not the passive result of physical constraints that somehow 'freeze' it. It can only be the result of an active biological mechanism that is continuously at work, a mechanism that has been referred to as *codepoiesis*.

(2) The second argument is that the genetic code is much more than a set of constraints. It is the novelty that gave origin to the world of proteins, to the first macroevolution in the history of life. This means that coding is not only a set of constraints. It is a generative mechanism, a mechanism of evolution.

Natural selection is the long term result of the copying of the genes and according to the Modern Synthesis, is the sole mechanism of evolution. The discovery of the genetic code proved instead that there are *two* distinct molecular mechanisms at the basis of life, the *copying* of the genes and the *coding* of proteins. This was its real revolutionary message, but it went unnoticed. Biologists did not conclude that copying and coding are two fundamental mechanisms. They concluded instead that copying is a fundamental mechanism and coding is a set of constraints. This is what prevented them from realizing that copying and coding are equally important mechanisms that have *complementary* roles in life.

In a way, this complementarity recalls the relationship that exists between particles and waves, in the sense that one cannot exist without the other and cannot be reduced to the other. But nobody in physics has ever suggested that the existence of particles and waves can be explained by saying that waves are constraints on particles or particles constraints on waves. Particles and waves are complementary realities, and we must accept this as a fact even if it is not intuitively appealing. In the same way, copying and coding are complementary mechanisms of molecular change and there is no point in denying this fact by saying that one is a mechanism and the other a set of constraints.

The first goal of Code Biology consists precisely in bringing to light the full nature of coding, in showing that coding is not just a set of constraints but a fundamental mechanism of life, the mechanism that brought absolute novelties into existence and gave origin to the great events of macroevolution.

Issue 2

The second point that I propose to your attention is the criteria that allow us to prove the existence of organic codes in Nature. It is widely accepted that a code is a mapping between the objects of two independent worlds, and this implies the existence of a third type of objects, called *adaptors*, that provide a bridge between them. In biology, the discovery of adaptors is the queen proof that an organic code exists, and it is therefore the first thing that we should look for. Unfortunately, it is not always possible to obtain it.

In the case of the organic codes that evolved in free-living cells, the adaptors are molecules and their identification is always possible. In the genetic code, for example, the adaptors are the transfer-RNAs, and in signal transduction codes they are the membrane receptors that establish a bridge between first messengers and second messengers. In the case of embryonic development, however, the mappings are much more complex and the adaptors are difficult to identify. Let me illustrate this point with two examples: the histone code and the *Hox* codes.

The histone code is a system of post-translational modifications of the histones that provide ‘marks’ which are recognized by specialized proteins and used to alter the structure of chromatin in highly specific ways. The result is that the cells of a developing embryo acquire ‘memories’ that allows them to ‘remember’ past events and develop accordingly. The histone code is therefore a set of rules that create a mapping between histone modifications and biological memories. The adaptors of the histone code have remained a mystery for a long time after the announcement of the code, in the year 2000, and only recently Stefan Kühn and Jannie Hofmeyr have been able to identify them. This shows that a code can be identified even when its adaptors are not yet known.

The *Hox* codes are a more difficult case because their definition is not based on mapping. They have been defined as *patterns of Hox genes expressions*, and a pattern is a feature in one world, not a mapping between two independent worlds. This means that the *Hox* codes are not true codes, and should be referred to as *Hox* patterns. The fact remains, however, that they determine the relative positions of the cells in a developing embryo and take part in relationships that have all the distinctive features of coding rules.

In principle, the number of arrangements that embryonic cells can form in space is unlimited, so it is imperative to make choices. These choices are the instructions that specify a body plan. More precisely the cells are instructed that their position is anterior or posterior, dorsal or ventral and proximal or distal *in respect to the surrounding cells*. These instructions are carried by genes, in particular by *Hox* genes, and consist in proteins that are referred to as the *molecular determinants* of the body axes.

The key point is that there are countless molecular determinants in the animal world and all function as arbitrary labels because their relationships with the body axes are not dictated by physical necessity. This means that they are conventional rules and the organic codes that express these rules have been referred to as *codes of the body plans*.

These codes create a mapping between molecular determinants and body axes, but the mapping is not between two types of molecules, as in the genetic code. It is between molecules and body axes, and this makes the mapping extremely complex because body axes are relationships between *populations* of cells.

The *Hox* codes, in other words, are not codes in themselves, but are an integral part of the codes of the body plans, and these are real codes because they create real mappings between molecular determinants and body axes. We conclude that it is the presence of mapping that proves the existence of organic codes in Nature, even in those cases where adaptors are not easy to identify.

Issue 3

The third point that I want to illustrate is the existence of different types of codes in life. The organic codes were the sole codes that existed on Earth in the first three billion years of evolution, when our planet was inhabited exclusively by microorganisms, but eventually two higher types of semiosis did appear.

One evolved in nervous systems and gave animals the ability to *interpret* the world (*animal semiosis*). Interpretation is essentially what Peirce called an 'abduction', a process that is neither induction nor deduction but a 'generalization from limited data' (Peirce 1906). More precisely, animals learned to interpret the world by using the two types of signs that Peirce called *icons* and *indexes*. They did not, however, exploit the third type of sign, the *symbols*. Only our species evolved that ability and developed a new type of semiosis that is based on language (*human semiosis*).

The evolution of life was characterized therefore by three great innovations: (1) the origin of organic semiosis, (2) the origin of animal semiosis, and (3) the origin of human semiosis.

This fits nicely with the idea of the 'Three Worlds' proposed by Karl Popper (1972, 1979) because these worlds corresponds to three distinct types of semiosis, as illustrated in the Table.

The Three Worlds of Life

Popper's Worlds	*Type of Semiosis*	*Mechanisms*	*Codes*
WORLD 1	Organic Semiosis	Coding	Organic codes
WORLD 2	Animal Semiosis	Coding Interpretation	Organic codes Neural codes
WORLD 3	Human Semiosis	Coding Interpretation Language	Organic codes Neural codes Language codes

This Table gives us a panoramic view of our field because it describes all types of codes that appeared in the history of life. On top of that, it makes us realize that there are two distinct versions of Code Biology: a restricted version and a general one. Code Biology in the strict sense is the study of all *organic* codes. Code Biology in the general sense is the study of *all* codes of life: organic codes, neural codes and language codes. This distinction is important because there are substantial differences between the two versions.

Code Biology in the strict sense is a fully scientific research field where the organic codes are studied with rigorous experimental methods and mathematical models. Code Biology in the general sense is, at the moment, a more speculative field, because experiments on neural codes and language codes are much more difficult and in most cases we only have indirect evidence about them.

At the same time, we cannot concentrate exclusively on the organic codes. We must not lose sight of the larger picture formed by all codes of life, and this is why Code biology must study all of them, from the genetic code to the codes of culture, from the origin of life to the origin of language.

Code Biology, in conclusion, is a deeply interdisciplinary field and necessarily requires the contribution of different scholars: biologists, neuroscientists, ecologists, linguists, philosophers, mathematicians and computer scientists. This is why we have speakers from many different disciplines in this conference, and all contributions are welcome.

Let me end this talk with a final concept. I said at the beginning that the great theories of biology have all been built *before* the discovery of the genetic code, and for this reason they have nothing to do with codes. After their discovery, organic codes have been regarded as constraints, complications, frozen accidents and the like, never as *normal fundamental components* of life, and this is an attitude that will not change easily.

We are going therefore to have difficult times ahead, but there is something that we can hang on to. It is the experimental fact that the organic codes are the great invariants of life, the sole entities that have been conserved for billions of years in evolution while everything else has been changed. This means that they are what really matters in living systems, and we can take comfort in the idea that what is at the centre of life will also be, one day, at the centre of biology.