Biosemiotics
DOI 10.1007/s12304-012-9147-3

ORIGINAL PAPER

Code Biology – A New Science of Life

Marcello Barbieri

Received: 14 January 2012 / Accepted: 8 March 2012
© Springer Science+Business Media B.V. 2012

Abstract Systems Biology and the Modern Synthesis are recent versions of two classical biological paradigms that are known as *structuralism* and *functionalism*, or *internalism* and *externalism*. According to functionalism (or externalism), living matter is a fundamentally passive entity that owes its organization to external forces (functions that shape organs) or to an external organizing agent (natural selection). Structuralism (or internalism), is the view that living matter is an intrinsically active entity that is capable of organizing itself from within, with purely internal processes that are based on mathematical principles and physical laws. At the molecular level, the basic mechanism of the Modern Synthesis is *molecular copying*, the process that leads in the short run to heredity and in the long run to natural selection. The basic mechanism of Systems Biology, instead, is *self-assembly*, the process by which many supramolecular structures are formed by the spontaneous aggregation of their components. In addition to molecular copying and self-assembly, however, molecular biology has uncovered also a third great mechanism at the heart of life. The existence of the genetic code and of many other organic codes in Nature tells us that *molecular coding* is a biological reality and we need therefore a framework that accounts for it. This framework is *Code biology*, the study of the codes of life, a new field of research that brings to light an entirely new dimension of the living world and gives us a completely new understanding of the origin and the evolution of life.

Keywords Organic codes · Modern synthesis · Systems biology · Origin of life · Macroevolution

Introduction

Today there are two major paradigms in biology. One is the Modern Synthesis, the paradigm that has unified the ideas of Darwin and Mendel and has put them on the firm mathematical basis provided by the equations of population genetics. The second is Systems Biology, a paradigm that studies living organisms not as collections of molecules but as whole systems, in line with the ideas developed by Wiener (1948), Rosen

M. Barbieri (✉)
Dipartimento di Morfologia ed Embriologia, Via Fossato di Mortara 64a, 44121 Ferrara, Italy
e-mail: brr@unife.it

Published online: 30 March 2012

(1958, 1991), von Neumann (1966), Mesarovic (1968), von Bertalanffy (1969), and Maturana and Varela (1980).

Systems Biology and the Modern Synthesis are the recent versions of two classical biological paradigms that have been discussed ever since the early 19th century. Traditionally they are known as *structuralism* and *functionalism*, but more recently they have also been referred to as *internalism* and *externalism* (Linde Medina 2010).

According to functionalism (or externalism), living matter is a fundamentally passive entity that owes its organization either to external forces (functions that shape organs) or to an external agent (natural selection). Structuralism (or internalism), is the view that living matter is an intrinsically active entity that is capable of organizing itself from within, with purely internal processes that are based on mathematical principles and physical laws.

At the molecular level, the basic mechanism of the Modern Synthesis is *molecular copying*, the process that leads in the short run to heredity and in the long run to natural selection. The basic mechanism of Systems Biology, instead, is *self-assembly*, the process by which many supramolecular structures are formed by the spontaneous aggregation of their components.

In addition to molecular copying and self-assembly, however, molecular biology has uncovered also a third great mechanism at the heart of life. The existence of the genetic code tells us that *molecular coding* is a biological reality and we need therefore a framework that accounts for it. This framework is *Code biology*, the study of all codes of life, from the genetic code to the codes of language.

Despite the existence of the genetic code, however, most biologists do not regard molecular coding as a fundamental mechanism of life, and it is interesting to understand why. The main reason is probably the widely diffused belief that there are only two types of codes in Nature, the genetic code, that appeared at the origin of life, and the cultural codes that arrived more than 3.5 billion years later. If this is true, in fact, one has to conclude that coding appeared only at the beginning and at the end of the history of life, and would be therefore an extraordinary exception not a normal mechanism.

It is important therefore to underline that many organic codes actually exist in Nature (Barbieri 2003, 2008), but even this discovery is not enough, on its own, to prove that coding is a *normal* mechanism of life. In order to accept this conclusion, we need to convince ourselves that the presence of many organic codes in Nature makes a real difference to our understanding of life and in particular to our traditional views of its origin and its history.

This paper is dedicated precisely to illustrate these points and has been divided accordingly into two parts. The first is dedicated to the origin of life and the second to the relationship that exists between organic codes and the great events of macroevolution.

Part 1 Life before the Cell

The Concept of Postchemical Evolution

Ever since the classic experiment by Stanley Miller (1953) it is known that a wide variety of organic molecules can be formed spontaneously under many environmental conditions. Organic molecules have been found in meteorites, for example, and spectroscopy has revealed their presence in comets and in interstellar space. This tells us that organic matter is formed spontaneously whenever and wherever suitable conditions exist, and today we know that those conditions existed on Earth at a very early stage of its history.

For more than a century, now, the problem of the origin of life has been formulated as the problem of describing the origin of the first cells in the environment of the primitive Earth. More precisely, as the problem of describing how primitive genes and primitive proteins appeared on Earth and managed to evolve into supramolecular systems that had the characteristics of primitive cells.

The crucial point in this process is to realize that genes and proteins are *not* produced by spontaneous reactions in any cell. They are produced by molecular machines which physically stick their subunits together according to the order provided by templates. Genes and proteins, in short, are *manufactured* molecules, not spontaneous ones. Spontaneous genes and spontaneous proteins did appear on the primitive Earth but could not give origin to cells because they did not have biological specificity. They gave origin to molecular machines and it was these machines and their products that evolved into the first cells.

The distinction between spontaneous and manufactured molecules allows us to divide the period that preceded the origin of the first cells into two great phases: a first period of *chemical* evolution during which organic compounds were formed exclusively by spontaneous reactions, and a second period of *postchemical* evolution, that started with the appearance of the first molecular machines and of the first manufactured molecules. This distinction is justified by the fact that there is a *fundamental* difference between spontaneous reactions and manufacturing processes. All spontaneous reactions are completely described by physical quantities (space, time, mass, temperature, etc.), whereas manufacturing processes can be fully accounted for only if we take into account additional observables like sequences and codes. Postchemical evolution, in short, is a stage of the history of life that came after chemical evolution but had not yet acquired the full characteristics of biological evolution.

Unfortunately, postchemical evolution is virtually an unexplored field of research, today, and is largely based on speculation, so it is tempting to dismiss it. But speculation on the origin of the cell is not a useless exercise because we understand something only when we understand, at least in principle, how it came into being. That is why origins are so important and why we are so fascinated by them.

The Twin Problems of the Beginning

The simplest molecular machines that could appear spontaneously on the primitive Earth were *bondmakers*, molecules that could stick monomers together in a random order and produce statistical polymers. In a primitive environment where chemical evolution had already accumulated many varieties of organic monomers, chemical bonds could be formed between all of them, thus giving origin to polymers such as polypeptides, polynucleotides, polysaccharides and countless other combinations of their building blocks.

Statistical polymers could also be formed spontaneously, of course, but bondmakers had at least two important advantages. One is that they could produce polymers at an almost continuous rate thus enormously increasing their number on the primitive Earth. The other is that some bondmakers could acquire the ability to join monomers together no longer at random but in the order provided by templates. Those bondmakers, in short, started making copies of the templates and became *copymakers*. The appearance of bondmakers led therefore to a steady increase in the production of statistical polymers and eventually to the appearance of the first copymakers, the first molecular machines that started populating the Earth with potentially unlimited numbers of copied molecules, in particular of nucleic acids.

Today, the copying of nucleic acids is carried out by huge molecules called *polymerases* but the actual formation of a chemical bond between two organic molecules does not require big dimensions and it is likely therefore that the first polymerases were much simpler than their modern descendants. What really matters, for our purposes, is that the origin of molecular copying does not seem to require extremely improbable events. At an advanced enough stage of chemical evolution, in an environment that had already produced a wide variety of organic molecules, including nucleotides and nucleic acids, the appearance of copymakers was as likely as that of any other average-size molecular structure.

The evolutionary impact of that event, however, was enormous, because the copying of nucleic acids leads in the short run to heredity and in the long run to natural selection. Molecular copying, in short, gives origin to a chain of processes that is potentially unlimited and endlessly open to change. Heredity and natural selection, the very hallmarks of life, are the result of a copying mechanism that is extraordinarily simple.

But there is more to life than heredity. There is metabolism, the actual building and rebuilding of organic structures and the never ending exchange of energy and matter that keeps everything going. And here things are much more complicated, because the molecular agents of metabolism are proteins and proteins cannot be copied. They are manufactured by huge molecular machines, called *ribosomes*, according to the rules of the genetic code, and even the simplest apparatus of protein synthesis requires the coordinated contribution of many dozens of different molecular components.

The origin of genes, in short, can be accounted for by a relatively simple mechanism, but the origin of proteins demands an apparatus whose complexity is almost beyond imagination. These are the twin problems of the beginning, and the second is probably the hardest of all the mysteries that we need to solve before we can say that we understand, at least in principle, how life appeared on our planet.

The Replication Paradigm

An unexpected solution to the problem of the origin of metabolism came in the 1980s with the discovery that some RNAs can function not only as genes but also as enzymes (Cech 1983, 1986; Guerrier-Takada et al. 1983; 1984). This suggested that RNAs capable of enzymatic activity (the "ribozymes"), could well have been the first agents of metabolism in the initial stages of the history of life. As Gilbert (1986) put it: "*There is no need for protein enzymes at the beginning of evolution. One can contemplate an RNA world containing only RNA molecules that serve to catalyse the synthesis of themselves*".

The idea that the first living systems were entirely made of RNAs became a distinct possibility, and this gave a great impetus to the *replication paradigm*, the idea that life has been the result of a process of chemical evolution based on self-replicating molecules. It must be underlined, however, that this idea has been tested in the laboratory by two complementary experiments performed by Spiegelman (1967) and by Eigen (1971), and the results have not supported it.

Spiegelman studied the evolution in the test tube of long RNA molecules obtained from a virus. Initially, these molecules contained 4,500 nucleotides that were coding for a complete set of viral proteins, but in the course of the experiment shorter mutants started appearing by the chance loss of genes that were not essential to replication, and in the end all that remained in the test tube was a relic of 220 nucleotides that could be replicated indefinitely (the *Spiegelman monster*). Eigen, on the other hand, studied the evolution of RNAs starting from the opposite end, i.e., from free nucleotides in a solution that contained also an enzyme that could join them together. In these conditions, the nucleotides were assembled in molecules that grew longer but only up to a limit, and in the end the solution reached a stationary state where the molecules capable of indefinite replication were, on average, 120 nucleotides long. The two experiments arrived in this way at the same conclusion: the evolution of self-replicating RNAs in vitro does not produce an unlimited increase because there is a severe limit to the size of the genes that are capable of indefinite replication. But how general is this conclusion? Can we extend it from the test tube to real life?

Eigen and Schuster (1977) showed that the limit in question is indeed a universal necessity because it is a consequence of fundamental theorems that apply to all self-replicating systems. The maximum length of the molecules is determined by the replication errors that are inevitably present in any replication process, because beyond that limit the system is overtaken by a runaway *error catastrophe* and collapses. Eigen and Schuster proposed that the error catastrophe could be avoided if different molecules combine their replication cycles into a single *hypercycle*, but Ursula Niesert was able to prove that hypercycles too have a size limit beyond which they are swept away by three new types of error catastrophes (Niesert et al. 1981).

Since then, the problem of the maximum size that is attainable by self-replicating molecules has been studied under a variety of conditions, with or without compartments and with or without the arrangement of genes in chromosomes, but in all cases it has not been possible to avoid error catastrophes (Maynard Smith and Szathmáry 1995). All theoretical and experimental results collected so far, in other words, show that the evolution of self-replicating molecules can produce little more than short sequences, and nothing remotely similar to a small cell.

The replication paradigm, in short, does not offer a realistic model for the origin of life. This means that precellular evolution had to take place in systems that did not depend on exact replication, i.e., in systems that had the ability to tolerate errors within very wide limits, because only such systems can avoid the replication catastrophes.

Ribosoids

Molecular phylogeny has shown that the ribosomal RNAs are among the most conserved molecules in evolution (Woese 1987, 2000) and this tells us that they appeared very early on Earth. The first RNAs, furthermore, appeared together with countless other molecules, and started interacting with them. Of all such interactions, those between RNAs and peptides were destined to have long lasting effects because they are still present today. What is particularly important is that in most cases the functions of the RNAs are greatly enhanced by the attachment of even short peptides, and it is that combination that provides the real functional units. This is why all molecules formed by RNAs and by their combinations with peptides and proteins (ribopeptides and ribonucleoproteins) have been referred to as *ribosoids*, and the collection of all ribosoids in a system has been called the *ribotype* of that system (Barbieri 1981).

The theories on the origin of life have been traditionally divided into gene-first and protein-first models, but the ribosoids suggest a third possibility. In addition to the gene-first and to the protein-first scenarios, we can consider the idea that genes and proteins appeared and evolved *together* on the primitive Earth. A particular version of this idea is *the ribotype theory* on the origin of life (Barbieri 1981, 1985), a theory that here is re-proposed in a slightly updated form.

The chemical properties of RNAs and peptides were undoubtedly favouring their interactions, but that is not enough to account for their evolutionary *partnership*. Why then did RNAs and peptides evolve together?

An important clue has come from the discovery that short pieces of ribosomal RNA can form peptide bonds between amino acids, thus joining them together and producing polypeptides (Nitta et al. 1998). This means that some primitive RNAs could produce random strings of peptides, and the interaction of these peptides with other RNAs could allow them to perform a variety of functions, including the ability to form phosphodiester bonds between nucleotides, thus joining them together and producing polynucleotides. Some ribosoids, in short, became *polimerizing ribosoids* and started producing increasing amounts of random polypeptides and random polynucleotides. There is no biological specificity in these relationships, and yet the mechanism for a sustained partnership is already there. *RNAs and peptides evolved together because one class of molecules provided bondmakers for the other.*

The polimerizing ribosoids that could stick amino acids together and produce random proteins were the ancestors of the modern ribosomes, and it is natural to expect that they were small molecules at the beginning, but in reality there is no need for this assumption. Modern ribosomes are huge molecular machines with molecular weights between 2 and 4 million daltons, and yet they can all arise by self-assembly from their components. In this case, the mechanism of self-assembly is perfectly capable of producing high-molecular weight structures from small molecules, which means that large-size ribosoids could well have appeared fairly early. It is important, furthermore, to realize that the polymerizing ribosoids did not need to have the same structure in order to perform the same function. Ribosomes of different species can have very different proteins and yet they all translate messenger-RNAs with the same accuracy, as if they were identical, and a similar polymorphism can legitimately be attributed to their ancestral precursors.

We come in this way to a major conclusion on the early history of life: a random form of replication was possible long before the origin of replicating cells. This statistical replication mechanism has been referred to as *quasi-replication* (Barbieri 1981) because the descendants were not exact copies of the progenitors. The peptide-polymerizing ribosoids, or *protoribosomes*, were capable of quasi-replication because they could produce a wide variety of random proteins some of which could interact with ribosomal RNAs and self-assemble into new protoribosomes which had the same function as their predecessors even though their structures were different. Quasi-replication, in short, allows a system to reproduce its functions without an exact replication of all its components, and this is important because it means that quasi-replicating systems can evolve and increase their size and complexity without being swept away by the error catastrophes that are inevitably associated with exact replication.

Nucleosoids

Ribonucleic acids and proteins can give origin not only to ribonucleoproteins and to large assemblies of ribonucleoproteins like ribosomes. They can also make giant scaffoldings made of tens of thousands of molecules and produce supramolecular systems whose dimensions can reach the size of a small cell. The formation of these supramolecular clusters is based on self-organizing processes, so it could well have taken place in primeval solutions, particularly when these became enriched by RNA-driven synthesis of proteins and by protein-driven synthesis of RNAs. The limits to the dimensions of such clusters are anybody's guess but the best example that we have today is represented by the nucleoli, and for this reason they have been given the name of *nucleosoids* (Barbieri 1981).

Nucleosoids are, therefore, large clusters of ribosoids, and form a highly heterogeneous family whose members can have shapes and dimensions of countless different types. Most of the nucleosoids that appeared on the primitive Earth were bound to be dead-ends from the point of view of evolution, but some of them could have an evolutionary potential. The comparison with modern nucleoli cannot be taken literally, of course, but it is nonetheless useful because it allows us to discuss some general properties that can be attributed to all nucleosoids.

(1) The first property is that a cluster of ribosoids is by no means an inert scaffolding. For one thing, its ribonucleoproteins can perform physical movements from one point in space to another and undergo a variety of conformational changes despite their being part of a supramolecular structure.
(2) Secondly, clusters of ribosoids provide microenvironments which trap molecules and localise their interactions. This is particularly important in all cases where many components take part in a biological process because molecules need to be close to each other in order to interact, and micro-compartments are ideally suited to keep them at close distances.
(3) Thirdly, nucleosoids form three-dimensional backbones which can have all intermediate dimensions between the nanometer and the micron, thus providing a bridge between molecules and cells.

Internal movements, internal compartments and a wide range of dimensions were properties of many types of nucleosoids and gave them a great evolutionary potential. It may be pointed out that nothing is of value in evolution if it has no lasting effect and however interesting the nucleosoids were, they were not capable of replication.

The answer, once again, is *quasi-replication*. We know that nucleoli of different species have widely different sizes, shapes and components and yet they all perform the same function as if they were identical. It is perfectly legitimate to assume that a similar polymorphism was present in primitive nucleosoids, and we can assume therefore that they too were capable of quasi-replication.

Nucleosoids, in short, could synthesize a wide variety of molecules and their supramolecular organization could trap the synthesizing ribosoids in a confined space, so that the cycles of synthesis could go on for long periods of time. Many nucleosoids were presumably short-lived, but on purely statistical grounds we can assume that a fraction of them were preferentially synthesizing other ribosoids and these did have an advantage over the others. A nucleosoid that was able to synthesize its own components could *grow* and reach whatever dimensions were physically attainable. Eventually, however, it would become unstable, break apart in smaller pieces and in some of these the ribosoids that were responsible for the previous syntheses would go on repeating the original process.

It may be worth underlining, furthermore, that the breaking up of the nucleosoids would distribute a wide set of fragments in space, including small pieces of RNAs that could travel and colonize other nucleosoids, thus behaving like the viruses that are responsible for what has become known as "horizontal gene transfer" (Doolittle 1999; Woese 2000).

The evolution of nucleosoids, however, could not go on indefinitely. Their quasi-replication mechanisms were producing an unpredictable variety of random components and were bound to lead them either towards extinction or towards yet more complex structures.

Heterosoids

The nucleosoids are, by definition, aggregates of ribosoids, but the statistical nature of their reactions implies that many other types of molecules could appear in them. As long as these non-ribosoidal components had only a temporary association with the nucleosoids, their presence was a sort of random noise and can be ignored. Eventually, however, some 'contaminations' proved to be useful for quasi-replication purposes, and that set in motion a process that turned them into stable components of the system. The nucleosoids evolved in this way into structures that became permanently associated with non-ribosoidal components, a process that can be summarized by saying that the nucleosoids became *heterogeneous nucleosoids* or *heterosoids* (Barbieri 1981).

One such contamination, for example, was DNA. The natural tendency of DNAs to split into two complementary parts could have greatly helped the splitting of the nucleosoids into fairly similar pieces, and this can explain why DNA became a component of quasi-replicating systems even when it was not carrying genetic information.

In addition to DNA, there was at least another class of molecules that was likely to become associated with nucleosoids. The research in prebiotic chemistry has shown that lipid and lipid-like molecules can be obtained very easily in simulation experiments and are frequently present in meteorites. This means that they are high-probability molecules, and it is likely therefore that they were formed in great quantities on the primitive Earth. Lipids, on the other hand, are capable of spontaneous self-assembly into supramolecular structures such as vesicles and membranes, and lipid membranes could easily reappear in different generations of quasi-replicating systems because they have a sort of hereditary potential of their own. Most membranes in fact are generated from other membranes and this means that they do not have to be created *ex novo* at each generation (Luisi and Varela 1989; Morowitz 1992; Szathmáry 1999).

One of the most important characteristic of the lipid membranes is their ability to act as catalysts in a variety of reactions such as ion transport and signal transduction. In particular, they can incorporate pigment molecules that capture the energy of light and set in motion those chains of energy transduction processes that eventually led to photosynthesis and respiration.

Lipid membranes, in short, could provide the sites of energy transformations and this was extremely valuable in RNA-based systems because RNAs have no ability to capture energy from the environment. As a matter of fact, the capacity to act as energy sources was so important that the chemical evolution of lipid systems was probably a precondition for the evolution of ribosoids and nucleosoids. It is possible, in other words, as suggested by Segré et al. (2001) that "*a Lipid World may have preceded the RNA world*".

Whatever was the historical sequence of the events, we conclude that a close association between lipid membranes and nucleosoids was bound to take place. There have been therefore two major developments in the evolution of the nucleosoids: DNA molecules started appearing inside them, and lipid membranes started providing energy sources and protective coats around them. These processes of amalgamation gave origin to the first 'modular' structures, i.e., to systems made of subsystems that were performing different functions but that managed to became integrated into a single working whole.

The first heterosoids, however, were not yet the first cells, because they were still synthesizing statistical proteins and dividing themselves into unequal parts by quasi-replication. In order to divide themselves into equal parts, they had to able to transmit the same proteins to their descendants, which means that they had to translate genes into specific proteins and this could be done only by developing a translation apparatus that was able to make proteins according to the rules of a genetic code.

The Co-evolution Theory of the Genetic Code

The origin of the genetic code is still a mystery, but we do have some interesting clues. Let us start from the fact that all experiments that simulate prebiotic conditions have always produced *less that half* of the 20 canonical amino acids that are found in living systems. This suggests that not all canonical amino acids appeared spontaneously on the primitive Earth, a conclusion that is strongly supported by the discovery that the amino acids which are missing in laboratory syntheses are also missing from carbonaceous meteorites.

The same conclusion is suggested by the fact that many bacteria are capable of synthesizing all 20 canonical amino acids from inorganic molecules, but they do this with two very different processes. In one case the pathways start from inorganic molecules and go all the way up to the so-called 'primary' amino acids. In the other case the pathways start from primary (or 'precursor') amino acids and terminate with the synthesis of 'secondary' (or 'product') amino acids.

The crucial point is that only primary amino acids are produced in laboratory syntheses, which strongly suggests that less than 10 primary amino acids appeared spontaneously on the primitive Earth, and that the first proteins were made exclusively from them. The implication is that the evolution of the genetic code started with less than 10 amino acids and went all the way up until it reached the canonical set of 20 amino acids that has been strongly conserved ever since. One of the best explanation of this pattern, so far, is the theory proposed by Jeffrey Tze-Fei Wong, the so called *co-evolution theory* of genetic code and amino acid biosynthesis (Wong 1975, 1980).

Wong noticed that, in the genetic code, the primary amino acids occupy codons that are almost always adjacent to those occupied by secondary amino acids. This correlation is overwhelmingly non random, and suggests that the evolution of the genetic code went hand in hand with the evolution of new amino acid pathways. More precisely, Wong proposed that the first genetic code that appeared on Earth was codifying only primary amino acids and all available codons were assigned to them. Later on, when other amino acids became available by new biosynthetic pathways, the primary amino acids conceded some of their codons to the secondary amino acids that were derived from them. During the evolution of the genetic code, in other words, the number of secondary amino acids was steadily increased by new biosynthetic pathways and they received their codons from the primary amino acids that served as precursors in their biosynthesis.

The explanation suggested by Wong is that the addition of new amino acids was a primary force in the evolution of the genetic code. At a time when translation was still prone to errors, the incorporation of new amino acids was favoured presumably because it increased the accuracy of translation by reducing the number of errors. This suggests that new amino acids continued to be admitted into the genetic code up to the point in which translation became virtually immune to errors. From that point onwards, no new amino acid was ever allowed into protein synthesis and the genetic code stopped evolving.

The evolution of the genetic code, in conclusion, was driven by the incorporation of new amino acids into protein synthesis, presumably because this process increased the accuracy of translation, and came to an end when the translation apparatus became the virtually error-free machine which has guaranteed biological specificity throughout the whole history of life.

The First Code

The co-evolution theory takes as a starting point a genetic code where all codons were assigned to less than 10 amino acids, but how did this *first code* come into existence? This is probably the most crucial issue because it is about the *origin* of the first genetic code, whereas Wong's theory is about the *evolution* of that first code into the universal code made of 20 canonical amino acids. The origin of the first code is clearly a different problem, and here we immediately come up against a formidable obstacle: how can a code evolve in stages if its typical features appear only when all its rules are present?

In order to deal with this problem let us start from the fact that a code is always part of a larger system that has been referred to as its *codemaker*. In our case, the codemaker of the genetic code is the apparatus of protein synthesis, so it is the evolution of this apparatus that we need to reconstruct.

Since ribosomes can arise by self-assembly, and ribosomal RNAs have the ability to produce peptide bonds, it is possible that a primitive apparatus had the ability to stick amino acids together at random and produce statistical proteins. In this case, the transfer RNAs would have served only as carriers of amino acids, and the precursors of the messenger RNAs would have provided only a temporary attachment site for the transfer RNAs. In such a situation, was there any point in using some transfer RNAS as adaptors that could start associating codons to amino acids? Without a complete genetic code the apparatus would have continued to produce statistical proteins, but even a small number of adaptors (i.e. a small number of coding rules) would had reduced the "randomness" of these random proteins.

To this purpose, let us recall that amino acids are divided into four great groups (acidic, basic, hydrophilic and hydrophobic) by the chemical properties of their side chains, and it is the sum total of these groups that determine the overall chemical properties of the proteins. Aspartic acid and glutamic acid, for example, belong to the acidic family, and a higher number of their codons would have generated more acidic proteins. Histidine and lysine, on the other hand, are basic amino acids, and if their frequency had been increased the result would have been a world of more basic proteins. In a similar way, other changes in codon frequencies would have given origin to more hydrophobic or more hydrophilic systems, and so on.

These consideration allow us to figure out the problems that primitive systems were facing when the genetic code was still evolving and the apparatus of protein synthesis was producing only statistical proteins, not specific ones. Statistical proteins could not function as specific enzymes, but could provide structural support, could modify the local microenvironment, and above all could maintain the systems in a state of continuous metabolic activity. In those circumstances, a change in the coding rules would have changed the overall physico-chemical conditions of the primitive systems and therefore their ability to grow and to split. Different RNAs, for example, would have favoured the synthesis of different families of statistical proteins, thus promoting group-properties rather than individual features.

We find a similar pattern in the evolution of language. The sounds uttered by the first speakers were probably random combinations of phonemes that were divided into a few great categories (sounds of friendship, enmity, fear, satisfaction, etc.), and gradually evolved into more specific expressions. The evolution of the rules of language went on hand in hand with the evolution of the words, and the two processes, although intrinsically different, evolved in parallel.

We conclude that in the last phase of *postchemical* evolution, the first genetic code probably arose as a set of rules that were affecting the *collective* properties of the primitive systems. The underlying mechanism was a combination of selection and conventions, a mixture of individual change and collective change, like the processes that shape the evolution of language.

The Ribotype Theory on the Origin of Life

This is a theory of *postchemical* evolution (Barbieri 1981). It describes a *logical* sequence of events that presumably took place on the primitive Earth when chemical evolution brought into existence the first molecular machines. These are molecules that were capable of producing other molecules by sticking their subunits together, first in a random order (*bondmakers*) and then in the order provided by a template (*copymakers* and *codemakers*). The appearance of manufactured objects started an entirely new phase in evolution because it brought into existence molecules that were no longer produced by the spontaneous assembly of their components, but by the manufacturing activity of other molecules.

The primary goal of the ribotype theory was to find a realistic solution to the problem that the replication paradigm had been unable to solve: *how did primitive systems manage to increase their size and complexity without being destroyed by error catastrophes?* The answer of the ribotype theory was based on three points.

(1) Polymerizing ribosoids could spontaneously form aggregates of very high molecular weight by self-assembly. This was a formidable mechanism because it could produce compounds that had dimensions of millions of daltons, as proved by the fact that viruses and ribosomes can still achieve these dimensions entirely with processes of self-assemby.

(2) It is known that ribosomes of different species contain different proteins, and yet they are all functioning as ribosomes. In this case, *the same function does not require the same components*, but only structures that collectively belong to a large family of molecules. This means that the ribosome function can reappear in many generations of ribosoids even without an exact replication of the molecules, a mechanism that was called *quasi-replication*. Polymerizing ribosoids, in other words, could generate various generations of protoribosomes even without mechanisms of exact replication.

(3) A system which contains heavy molecular components can avoid the error catastrophes because *high molecular weight structures absorb thermal noise*, and are immune to a wide range of perturbations. This conclusion is based on a general engineering principle that Burks (1970) expressed in this way: *"there exists a direct correlation between the size of an automaton – as measured roughly by number of components – and the accuracy of its function"*. In the case of protein synthesis, this means that, in order to be precise, ribosomes must be immune to thermal noise and must therefore be heavy.

This then is the solution of the ribotype theory: in order to avoid the error catastrophes in the journey toward exact replication, it was necessary to have high molecular weight protoribosomes, and the production of these ribosomes for an indefinite number of generations was possible, before exact replication, because ribosoids could achieve it with processes of self-assembly and quasi-replication. The development of high molecular weight protoribosomes took place therefore during *postchemical* evolution, simply because it could be realized with processes that were both natural *and* primitive.

Together with polymerizing ribosoids, ribogenes were also evolving, which means that precellular systems had genomes predominantly made of RNAs. The dimensions of ribogenomes could also increase, because the mechanism of quasi-replication was avoiding error catastrophes, and eventually ribogenomes and protoribosomes became sufficiently large for both to be immune to thermal noise. At that point, the genetic code underwent the last phase of its evolution and gave origin to a translation system that was virtually immune to errors. That turned quasi-reproduction into exact reproduction, and postchemical evolution was replaced by what we now call biological evolution.

The Ribotype Model of the Cell

According to modern biology, the cell is a duality of *genotype* and *phenotype*. This model was proposed by Johannsen (1909), but was accepted only in the 1950s, when it became clear that genes carry *linear information* whereas proteins function in virtue of their *three-dimensional structures.* The genotype-phenotype duality is a model that divides two different biological functions (heredity and metabolism), and two different physical quantities (information and energy), and describes the cell as a biological computer made of organic software and organic hardware. It is the simplest and most general way of defining a living system, and has become the foundational model of modern biology, the scheme that transformed the *energy-based* biology of the 19th century into the *information-based* biology of the 20th.

In the 1950s and 60s, however, the study of protein synthesis revealed that genes and proteins are not formed spontaneously in the cell but are manufactured by a system of molecular machines based on RNAs. In 1981, the components of this manufacturing system were called *ribosoids* and the system itself was given the collective name of *ribotype* (Barbieri 1981, 1985). The cell was described in this way as a system made of genes, proteins and ribosoids, i.e., as a trinity of *genotype, phenotype and ribotype*.

The category of the ribotype may appear redundant, at first, because the ribosoids are usually regarded as components of either the genotype or the phenotype, but there is a good reason for keeping it distinct. The reason is that we can speak of genotype and phenotype only when a translation apparatus is present, because it is this apparatus that *defines* the genes and *synthesizes* the proteins. Genotype and phenotype, in other words, cannot exist without a translation apparatus, whereas the RNAs of the ribotype do not need that apparatus. They only need to be copied in order to perform their functions. The evolutionary consequence of this fact is that the ancestors of the ribotype had to come before the translation apparatus and were instrumental to its evolution. The ribotype, in short, is distinct from genotype and phenotype because it had an evolutionary priority over them.

The RNAs and the proteins that appeared spontaneously on the primitive Earth produced a wide variety of ribosoids, some of which were synthesizing ribosoids whereas others were ribogenes and others were riboproteins (or ribozymes). The systems produced by the combination of all these molecules, therefore, had a ribotype, a ribogenotype and a ribophenotype. Eventually, evolution replaced the ribogenes with genes and the riboproteins with proteins but the synthesizing ribosoids of the ribotype have never been replaced. This shows not only that the ribotype is a distinct category of the cell but also that it is a category without which the cell simply cannot exist.

It is an experimental fact, at any rate, that every cell contains a system of RNAs and ribonucleoproteins that makes proteins according to the rules of an organic code, and that system can be described therefore as a *code-and-template-dependent-protein-maker*, i.e., as a *codemaker*. That is the third party that makes of every living cell a trinity of genotype, phenotype and ribotype. The genotype is the seat of heredity, the phenotype is the seat of metabolism and the ribotype is the codemaker of the cell, the seat of genetic coding.

Part 2 Codes and Macroevolution

The 'Fingerprints' of the Organic Codes

From time immemorial it has been taken for granted that codes and conventions exist only in the world of culture whereas nature is governed by universal laws. Our first problem, therefore, is to understand if we can *legitimately* use the term 'code' at the molecular level. To this purpose, let us start by underlining that there is an important difference between *copying* and *coding*, a difference that in molecular biology is particularly evident in *transcription* and *translation*.

In transcription, a DNA sequence is used as a template to assemble an RNA sequence, and in this case a normal biological catalyst (an RNA polymerase) is sufficient, because each step requires a single recognition process. In translation, instead, two independent recognition processes must be performed at each step (one for a codon and the other for an amino acid), and the assembling system (the ribosome) needs special molecules, first called *adaptors* and then transfer RNAs, in order to associate codons to amino acids.

This conclusion should now be generalized. We are accustomed to thinking that all biochemical processes are *catalyzed* reactions, but in reality we should sharply distinguish between catalyzed and *codified* reactions. Catalyzed reactions are processes (like transcription) that require only one recognition process at each step, whereas codified reactions require (like translation) two independent recognition processes at each step and a set of coding rules. The crucial point is how can we distinguish between them? How can we prove the existence of codified reactions and organic codes in Nature?

The answer comes from the fact that any code is a set of rules that create a correspondence between the objects of two independent worlds, which means that it is necessarily implemented by structures that perform two independent recognition processes at each step. The genetic code, for example, is a set of rules that link the world of nucleotides to the word of amino acids, and its adaptors are the transfer-RNAs. The adaptors are required because there is no necessary link between the two worlds, and a set of rules is required in order to guarantee the specificity of the correspondence.

The adaptors, in short, are the key molecules in all organic codes. They are the molecular *fingerprints* of the codes, and their presence in a biological process is a sure sign that that process is based on a real code. This gives us an *objective* criterion for the discovery of the organic codes, and their existence in Nature is no longer a matter of speculation. It is, first and foremost, an experimental problem.

Signal transduction, for example, creates a link between first and second messengers, but these molecules belong to two independent worlds because any first messenger can be coupled with any second messenger (Alberts et al. 2007). The signal receptors of the cell membrane have in fact the defining characteristics of true adaptors and this reveals the existence of *signal transduction codes* (Barbieri 1998, 2003). Molecular adaptors have also been found in many other biological processes, thus bringing to light the existence of *splicing codes*, *cell compartment codes* and *cytoskeleton codes* (Barbieri 2003, 2008). Other organic codes have been discovered with different criteria. Among them, the *metabolic code* (Tomkins 1975), the *sequence codes* (Trifonov 1987, 1989, 1996, 1999), the *adhesive code* (Readies and Takeichi 1996; Shapiro and Colman 1999), the *sugar code* (Gabius 2000, 2009), the *histone code* (Strahl and Allis 2000; Turner 2000, 2002; Gamble and Freedman 2002), the *tubulin code* (Verhey and Gaertig 2007), and the *splicing code* (Pertea et al. 2007; Barash et al. 2010; Dhir et al. 2010).

The living world, in short, is teeming with organic codes, and this means that at some point in time they came into being. Our goal, at this point, is to find out the implications of those historical facts.

The First Cells

There is no trace of the first cells in the fossil record, but molecular biology has allowed us to reconstruct some of their most general features. More precisely, the comparative analysis of the molecular data has shown that all known cells belong to three distinct primary kingdoms, or domains, that have been referred to as *Archaea*, *Bacteria* and *Eucarya* (Woese 1987, 2000). Another major result is that all cells have a virtually universal genetic code, which means that this code appeared in the ancestors of all cell domains, and evolved therefore in systems that had not yet acquired a modern cell design. These systems are collectively known as the *common ancestor*, because it was from that primordial population that the genetic code has been transmitted to all living creatures that have appeared on Earth. The origin of the genetic code, however, was not enough to create a modern cell and there had to be at least another macroevolution to bring that about. According to Woese, the ancestral systems that developed the genetic code were not proper cells because they had not yet crossed what he called the *Darwinian threshold*, an unspecified crucial limit beyond which a full cell organization could come into being (Woese 2002).

The existence of many organic codes in Nature, however, suggests a different explanation. A good clue comes from the fact that the combination of a signal transduction code with the genetic code allows the cell to regulate protein synthesis according to the signals that come from the environment and respond to them with a context-dependent behaviour (Jacob and Monod 1961). A signal transduction code was therefore of paramount importance to the ancestral systems, which makes it very likely that they made various independent attempts to develop it.

It is an experimental fact, at any rate, that Archaea, Bacteria and Eucarya have three different types of membranes and three distinct signalling systems, and this brings us to the idea that the three cell domains came into being by the combination of the universal genetic code with three distinct signal-transduction codes. This amounts to saying that a modern cell design requires at least two organic codes: a genetic code for protein synthesis and a signal transduction code for a context-dependent behaviour. The three primary kingdoms of life, in other words, were the result of three independent attempts performed by the descendants of the common ancestor to evolve a signal transduction code.

We realize in this way that the genetic code was instrumental to the origin of life and that the signal transduction codes were instrumental to the origin of the modern cell designs. In these two cases, in other words, there has been a deep relationship between organic codes and the great events of macroevolution.

Two Evolutionary Strategies

In order to understand the evolution of the first cells we need to reconstruct, at least in principle, the basic properties of the common ancestor, and to this purpose the best clue comes from the discovery that bacteria appeared very early on our planet and some of them have remained substantially the same ever since. This is dramatically illustrated by the fact that modern stromatolites built by cyanobacteria are virtually identical to the 3.4 and to the 1.8 billion year old stromatolites that have been found in the fossil record (Schopf 1999; Knoll 2003). Primitive bacteria, in other words, already had the main characteristics of their modern descendants, and this allows us to draw at least three conclusions about the common ancestor.

(1) The modern bacterial genome is a single circular molecule where all genes are arranged one after the other without interruptions and all of them carry information that is actually used by the cell. Such an organization is surely very efficient, but precisely for this reason it could hardly have appeared very early. A genome made of various DNA pieces where only a few were carrying useful genetic information is definitely more primitive, and it is likely therefore that the common ancestor did have such a genome.

(2) Bacterial protein synthesis is based on *unstable*, or *short-lived*, messengers that allow the cells to adapt very rapidly to changing environmental conditions Again, such a fast-reacting system cannot be primitive because unstable messengers require advanced forms of regulation and must have been the result of some evolution. It is likely therefore that the common ancestor was using more stable messengers because only these molecules are compatible with simple forms of regulation.

(3) In bacteria, the transcription of the genes is immediately followed by translation, to the extent that in most cases protein synthesis starts on primary transcripts that are still attached to DNA. The result is that there is no time for a modification of the primary transcripts, and these are directly used as messenger RNAs in protein synthesis. Such a fast sequence of precisely coordinated steps is hardly primitive and it is likely therefore that in the common ancestor there was no strict coupling between transcription and translation.

We realize in this way that the common ancestor did not have the typical features that we find in bacteria, but much more primitive characteristics. What is most important is that because of these characteristics the descendants of the common ancestor had two evolutionary strategies in front of them, one based on increasing simplification, or streamlining, and one based on increasing complexity.

The cells that adopted a streamlining strategy got rid of all unnecessary components and learned to make the fastest possible use of genetic information by abolishing all intermediate steps between transcription and translation. Other cells, however, conserved most of their primitive features, including a physical separation between the transcription of the genes and their translation into proteins. This allowed them to introduce intermediate reactions between these processes and started performing cutting and sealing operations that eventually evolved into the rules of the *splicing code*.

The cells that adopted a streamlining strategy, in conclusion, became extremely efficient in a relatively short time and have remained substantially the same ever since. Other cells conserved their primitive features, including the potential to evolve new organic codes, and became increasingly complex. In particular they developed splicing, a process that can transform the primary transcripts of a gene into different messenger RNAs, and therefore into different proteins.

The operations of splicing, on the other hand, require a separation *in time* and *in space* between transcription and translation, and this was the precondition for the development of a physical barrier between them, a barrier that eventually took the form of the nuclear membrane. The origin of the nucleus is traditionally regarded as the origin of the eukaryotic cell, and splicing was precisely the process that created the *precondition* for that event. Splicing, on the other hand, is based on the rules of a code and we find, once again, that a new organic code was instrumental in bringing about a major transition in the history of life.

Form and Movement

Organic molecules are formed spontaneously in a wide range of environmental conditions, and today we know that suitable conditions did exist on the primitive Earth. A process of chemical evolution started therefore fairly early on our planet and slowly transformed it into a niche where the first cells appeared together with countless organic compounds and nutrients. Even a large food supply, however, was destined to get depleted, and this created the conditions for the appearance of two very different survival strategies. Some cells adapted their metabolism to smaller and smaller molecules, and eventually learned to perform all metabolic reactions from inorganic matter. In this way they ceased to be *consumers*, and became *producers* of organic matter, and when that happened, the risk that life could become extinct by lack of food was drastically reduced.

Other cells continued to feed on organic matter, but the scarcity of small nutrients forced them to use increasingly bigger compounds. A potentially important source of food was provided by the bodies of other cells, especially dead ones, and the ancestral consumers learned to develop structures that enabled them to ingest bigger and bigger pieces of organic matter. Such a property required a plasma membrane that could engulf the prey, and to that purpose the cells had to be able to change their shape from within, by using molecular structures that eventually evolved into *cytoskeletons*.

The development of a cytoskeleton was probably favoured by the need to improve the ability of the consumer cells to feed on organic matter, but it had other collateral advantages. The operations that build the cytoskeleton are based on molecular adaptors and therefore on the rules of an organic code. New types of adaptors, in turn, can use the same components of the cytoskeleton for new purposes, and that probably explains the fact that the cytoskeleton has acquired a variety of different functions in the course of evolution.

The first cytoskeletons were allowing the consumer cells to perform the movements that were necessary to ingest a prey, but new adaptors allowed them to perform other movements. For example the movements that allow a cell to change its position in space and actively go round searching for a prey rather than passively waiting for it. Another outstanding example is the movements that take place in mitosis and are responsible for the physical displacement of the chromosomes to the opposite end of the mitotic spindle. The key molecules of this process are known and have the typical characteristics of true molecular adaptors, which means that the movements of mitosis and meiosis are based on the rules of an organic code. Another evolutionary development of the cytoskeleton was the process that allowed cells of the same species to fuse together in a system that maintained the ability to divide itself. This was the precondition for the origin of sex and once again we find that a new set of coding rules was instrumental in bringing about another great event of macroevolution.

The Codes of the Body-Plan

The origin of animals gives us the same problem that we face in all major transitions: how did real novelties come into existence? In the case of the first animals the problem is: how did some eukaryotic cells manage to generate those particular three-dimensional structures that we call animals? The solution was obtained by three types of experiments. More precisely, by the attempts to form multicellular structures with one, two or three different types of cells (the *germ layers*). The experiment with one cell type produced bodies which have no symmetry (the sponges); two cell types built bodies with one axis of symmetry (the *radiata* or diploblasts, i.e., hydra, corals and medusae), and three cell types gave origin to bodies with three axes of symmetry (the *bilateria* or triploblats, i.e., vertebrates and invertebrates) (Tudge 2000).

In principle, however, the number of three-dimensional patterns that the first animal cells could form in space was unlimited, so it was imperative to make choices. These choices, or constraints, turned out to be sets of 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 consist of molecules which are referred to as the *molecular determinants* of the body axes (Gilbert 2006). The crucial point is that there are countless types of molecular determinants and yet all triploblastic animals have the same axes (top-to-bottom, back-to-front and left-to-right). This shows that there is no necessary link between molecular determinants and body axes, and that in turns means that the actual links that we find in Nature are based on conventional rules, i.e., on the rules of organic codes that can be referred to as the *codes of the body-axes*.

It must be underlined that the relationships of the body axes are between *cells*, and this means that they do not determine only the axes of the body, but also those of all its constituent parts. In the hand, for example, the proximo-distal axis is the direction from wrist to fingers, the antero-posterior axis is from thumb to little finger, and the dorsal-ventral axis is from the outer surface to the palm of the hand. Right and left hands have different symmetries because their axes are one the mirror image of the other. There is therefore a multitude of axes in the animal body, and it turns out that many of them have the same molecular determinants. The products of the gene *Sonic hedgehog* (*Shh*), for example, determine the dorso-ventral axis of the forebrain as well as the antero-posterior axis of the hand, which again shows that molecular determinants are mere labels and represent the conventional rules of a code.

The antero-posterior axis of the body (the head-to-tail direction), is determined by two small depressions that are formed very early on the outer surface of the embryo and that mark the signposts of mouth and anus. Between those two points, a third depression is produced by the movements of a colony of migrating cells that invade the space between the first two germ layers (ectoderm and endoderm) to form the middle germ layer (the mesoderm). The invagination point (the blastopore) can be set either near the mouth-signpost (the *stomodeum*) or near the anus-signpost (the *proctodeum*) and that choice determines the future organization of all organs in the body. The animals where the blastopore is formed near the signpost of the mouth (*stoma*) are invertebrates (technically *protostomes*): they have an outside skeleton, a dorsal heart and a ventral nervous system. The animals where the blastopore is formed away from the mouth signpost are vertebrates (more precisely *deuterostomes*): they have an inside skeleton, a ventral heart and a dorsal nervous system.

The whole organization of the body, in other words, is a consequence of a few parameters that determine the migrations of the mesoderm in respect to the body axes. The crucial point is that these migrations (the *gastrulation* movements) take place in countless different ways in both vertebrates and invertebrates, and this shows that they are not due to physical necessity but to the conventional rules of a *gastrulation code*. We realize in this way that the three-dimensional organization the animal body is determined by a variety of organic codes that together can be referred to as the *codes of the body-plan*.

Cell Fate and Cell Memory

All free-living cells, from bacteria to protozoa, react swiftly to environmental changes, but the cells of multicellular animals have a more sophisticated behaviour. Their reactions do not take into account only their present conditions but also their history. This is because in embryonic development the cells learn not only to become different, but also to *remain* different. They acquire, in short, a *cell memory*. In technical terms, they go through embryonic processes that fix their *histological* fate for the rest of their life.

This great discovery was made by Spemann (1901) by studying what happens when small pieces of tissue are transplanted from one part of an embryo to another. Spemann found that the embryonic cells can change their histological fate (skin cells, for example, can become nerve cells) if they are transplanted *before* a critical period, but are totally unable to do so if the transplant takes place *after* that period. This means that for every cell type there is a crucial period of development in which *something* happens that decides what the cell's destiny is going to be, and that something was called *cell determination.*

Other experiments proved that determination does not normally take place in a single step but in stages, and that the number and duration of these stages vary from one tissue to another. The most impressive property of determination is the extraordinary stability of its consequences. The process takes only a few hours to complete but leaves permanent effects in every generation of daughter cells for years to come. The state of determination, furthermore, is conserved even when cells are grown in vitro and perform many division cycles outside the body. When brought back in vivo, they express again the properties of the determination state as if they had never 'forgotten' that experience (Alberts et al. 2007).

The determination of cell fate, in short, amounts to the acquisition of a *cell memory* that is maintained for life and is transmitted to all descendant cells. The various steps of determination are controlled by molecules, known as *molecular determinants*, which can be passed on by the mother on fertilization or produced by the embryo at various stages of development. The crucial point is that the basic histological tissues are the same in all animals, but their molecular determinants are of countless different types, which shows that the link between determinants and histological fate is not dictated by physical necessity but by the rules of codes that have been referred to as *histological codes*, or *transcriptional codes* (Jessell 2000; Marquardt and Pfaff 2001; Perissi and Rosenfeld 2005; Flames et al. 2007).

This is dramatically illustrated by the most fundamental of all cell distinctions, that between somatic and sexual cells. In *Drosophila*, for example, that distinction is determined by the *pole plasm*, a substance that is deposited by the mother at the posterior end of the egg. All cells that receive molecules from the pole plasm become sexual cells and are potentially immortal, whereas all the others become somatic cells and are destined to die with the body. The distinction between somatic and sexual cells takes place in all animals but is produced by widely different molecules, in some cases produced by the mother and in other cases by the embryo, all of which shows that it is an outstanding example of histological code.

During embryonic development, in conclusion, the cells undergo two distinct processes of determination: one for their three-dimensional pattern and the other for their histological fate. Both processes are totally absent in free-living cells, which again shows that the origin of animals was a true macroevolution. Both processes, furthermore, are based on conventional rules of correspondence between molecular determinants and cell states because the determinants can be of countless different physical types. In all animals, in other words, the body plan and the histological fate of tissues and organs are based on the rules of organic codes.

The Neural Code

There seems to be a deep parallel between the origin of life and the origin of mind. In the origin of life, the key event was the appearance of *proteins* and the genetic code played a crucial part in it because it was instrumental to protein synthesis. In the origin of mind, the key event was the appearance of *feelings* and it seems that a *neural code* was as instrumental to the production of feelings as the genetic code was to the production of proteins (Barbieri 2011). The parallel, therefore, is between feelings and proteins, and this immediately tells us that there are both similarities and differences between the two cases.

Proteins are *space-objects*, in the sense that they act in virtue of their three-dimensional structure, whereas feelings are *time-objects* because they are 'processes', entities that consist of flowing sequences of states. The same is true for their components. Proteins are assembled from smaller space-objects like amino acids, and feeling are assembled from lower level brain processes such as neuron firings and chemical signalling.

But can we really say that a (nearly) universal neural code exists in all animals as a (nearly) universal genetic code exists in all cells?

Let's consider, for example the transformation of mechanical stimuli into tactile sensations. Rats have mechano-receptors on the tip of their whiskers while we have them on the tip of our fingers, and there is no doubt that our tactile exploration of the world is different from theirs, but does that mean that we use a different neural code? The evidence is that the physiological processes that transform the mechanical stimuli into tactile sensations are the same in all animals, and this does suggest that there is a universal mechanism at work (Nicolelis and Ribeiro 2006). As a matter of fact, the evidence in question comes from animals with three germ layers (the triploblasts), but they represent the vast majority of all animal taxa, so let us concentrate our attention on them. How can we generalize the experimental data and conclude that virtually all triploblastic animals have the same neural code?

We do know that the starting point of all neural processing is the electrical signals produced by sense organs, but we also know that the sense organs arise from the basic histological tissues of the body, and that these tissues (epithelial, connective, muscular and nervous tissues) are the same in all triploblastic animals. All signals that are sent to the brain, in other words, come from organs produced by a limited number of universal tissues, and that does make it plausible that they represent a limited number of universal inputs. But do we also have a limited number of universal outputs?

The neural correlates of the sense organs (feelings and perceptions) can be recognized by the *actions* that they produce, and there is ample evidence that all triploblastic animals have the same basic *instincts*. They all have the imperative to *survive* and to *reproduce*. They all seem to experience hunger and thirst, fear and aggression, and they are all capable of reacting to stimuli such as light, sound and smells. The neural correlates of the basic histological tissues, in short, are associated with the basic animal instincts and these appear to be virtually the same in all triploblastic animals.

What we observe, in conclusion, is a universal set of basic histological tissues on one side, a universal set of basic animal instincts on the other side, and a set of neural transformation processes in between. The most parsimonious explanation is that the neural processes in between are also a universal set of operations. And since there is no necessary physical link between sense organs and feelings, we can conclude that the bridge between them can only be the result of a virtually universal *neural code*.

The Tree and the Web of Life

Darwin (1859) described evolution as a process that started “from simple beginnings” and gave origin to increasingly diverse “most beautiful forms”, much like a tree that grows and divides into countless branches. At the basis of the tree there are the first “primordial forms” – the first cells – and out of them grows a trunk, which splits again and again to create an ever expanding tree. Each branch represents a species and the branching points are where one species gives origin to two. Most branches come to a dead end, signifying extinction, but some go all the way up to the top and are today’s living creatures. This is the “Tree of Life”, the graphic description of how every species is related to all the others that have appeared on Earth since the origin of life.

The reconstruction of the tree of life has been the holy grail of biology for generations of naturalists, but then the unexpected happened and today the project has changed almost beyond recognition. The unexpected came in the 1990s as a result of the sequencing of the genes, a technique which revealed that bacteria and archaea are routinely swapping genetic material with many other organisms in a process called horizontal gene transfer (HGT). The pattern of a tree is realized when genes are passed down from one generation to the next, i.e., when descent is “vertical”. When genes are swapped “horizontally” between co-existing species, instead, they become part of many branches simultaneously and the resulting pattern is no longer a tree but a web (Doolittle 1999). It turned out, furthermore, that HGT is by no means a secondary process. In archaea and bacteria it can account for as much as 80 % of the genes (Dagan et al. 2008), whereas in unicellular eukaryotes (protista) it is only slightly lower than that (Andersson 2005), and in animals and plants – where genes are transferred horizontally by viruses – it can make up for 40 to 50 % of the whole genome (Doolittle and Bapteste 2007).

The message that come from the experimental data is clear enough: for the first three thousand million years of evolution there has been no proper tree of life but a vastly interconnected web. A tree-like pattern began to emerge only in eukaryotes, where the effects of vertical descent became increasingly evident over the levelling drive of horizontal gene transfer. The web of life, on the other hand, has always been, and continues to be, the predominant pattern in the living world, and it makes us realize that there are three distinct forms of life on our planet.

The molecules responsible for HGT are nucleic acids (RNAs and DNAs) that are produced by copying and that survive by invading any system that is capable of copying them. They represent a form of life based on “*copying alone*” and since this is the logic of the viruses we can give them the collective name of ‘viroidea’. Archea and bacteria are cellular systems that are based on copying and coding, but have adopted a streamlining strategy that prevented them from developing new organic codes. They represent a second form of life, known as ‘prokarya’, that is based on “copying and *limited* coding”. Other cells, the ‘eukarya’, have maintained the potential of the ancestral precellular systems to explore the coding space and represent a third form of life based on “copying and *unlimited* coding”.

Eukarya, Prokarya and Viroidea, in conclusion, represent distinct forms of life, and there have been two parallel drives in evolution: one that created a web-like pattern and one that generated the classical branching pattern of the tree of life. The global web of viruses and bacteria accounts for more than 90 % of the mass of organic matter that exists on Earth, and is by far the dominant form of life on our planet. The fact remains, however, that life managed to go beyond the level of viruses and bacteria and to give origin to increasingly complex forms that appeared together with new organic codes. Copying alone was not enough, and that again makes us realize the extraordinary importance that coding, and more precisely *unlimited* coding, had for the evolution of life on our planet.

A New Definition of the Cell

Systems Biology describes the cell as a system that is capable of 'autopoiesis', a "system that fabricates itself". Autopoiesis is indeed a fundamental process but if we take a close look we realize that it captures only one important aspect of life and leaves out many others. It does not explain, for example, the fact that during embryonic development the cells become *different* from their progenitors. It does not tell us why cells can do something that is very distant from autopoiesis: they can suicide themselves.

One may say that these are collateral properties that evolution *added* to original autopoietic systems, so it is autopoiesis that we must start with because it was the beginning of everything. But this is precisely what we *cannot* say because autopoiesis, or self-fabrication, was itself the result of evolution not the starting point. The starting point were ancestral systems that had not yet developed a genetic code and this means that they were producing systems that were inevitably *different* from themselves.

We know virtually nothing of the common ancestor, but the molecular data tell us that it had at least three types of RNAs one of which was able to join peptides together into polypeptides. Before the origin of the genetic code, that ancestral ribonucleoprotein system, or ribotype, was still in the process of evolving its coding rules and was therefore a *code exploring system*. After the origin of the genetic code, however, the situation changed dramatically. No other modification in coding rules was tolerated and the cell became a *code conservation system*. From that moment onwards, the quintessential point, the true constant in the history of life, was the imperative to conserve the rules of the genetic code.

Some cells, furthermore, maintained the potential to develop *other* coding rules and gave origin to a new *code exploring system*. The ancestral *Eukarya*, for example, had *a code conservation part* for the genetic code, but also a *code exploring part* for the splicing code. This tells us that the eukaryotic cells had to solve two distinct problems:

(1) the first is a *code conservation problem*: the cells managed to conserve their organic codes for billions of years despite the fact that all their genes were subject to continuous mutations,
(2) the second is a *code exploring problem*: some cells managed to maintain the potential to explore the coding space and to give origin to new organic codes.

We realize in this way that what is really important in the cell is two things: one is the ability to conserve its organic codes and the other is the potential to evolve new ones. This gives us an entirely new definition of the cell that can be expressed in this way: "*the cell is a system that is capable of creating and conserving its own codes*".

This definition accounts for the two most important events that took place in evolution. The ability to create coding rules accounts for the origin of the genetic code and of all the other codes that followed. The ability of the cell to conserve its own codes accounts for the fact that the organic codes are the great invariants of life, the entities that must be conserved while everything else is changing.

Code biology, the study of the organic codes, is indeed the new frontier, a new science of life.

Conclusion

The basic mechanism of the Modern Synthesis is the *copying* of the genes, the process that leads in the short run to heredity and in the long run to natural selection. The basic mechanism of Systems Biology is *self-assembly*, the process that spontaneously generates biological structures from their components without additional genetic information (the *order-for-free* effect). Molecular copying and self-assembly are well documented realities and are rightly regarded as fundamental mechanisms of life. But molecular *coding* is also an outstanding fact and has all the characteristics of a third fundamental mechanism. Such a conclusion can be avoided only if coding can be reduced to copying or to self-assembly, so let us see if this is possible.

Self-assembly is entirely accounted for by the laws of physics, but in the genetic code there is something else that comes into play. As Monod (1970) has underlined, the links between codons and amino acids are chemically *arbitrary*, in the sense that they are not the result of physical necessity, and this means that the genetic code was the result of *natural conventions*, not of deterministic processes. All the components of the genetic code, furthermore, are specified by genes that are continuously subject to mutations and the code has been strongly conserved in evolution only because all its modifications have been systematically suppressed in every cell and in every generation ever since the origin of life. The genetic code, in short, is made of conventional and potentially changeable rules, and there is no way that it can be reduced to the universal and immutable processes of self-assembly.

The possibility that coding is reducible to copying may look more plausible, at first, because the rules of the organic codes are implemented by adaptors and the genes of these molecules are passed on from one generation to the next by molecular copying. But the fact that a system is made of molecules does not exclude the existence of rules that are generated by the system and not by the molecules. Copying and coding represent the two levels of organization that exist in every living organism – the collective level of the whole system and the individual level of its components – and neither is reducible to the other because they are *complementary* mechanisms. They evolved in parallel in the history of life just as individual words and rules of grammar evolved in parallel in the history of language.

There are, furthermore, two other important differences between copying and coding.

One is the fact that copying produces either exact copies or slightly different versions of the copied molecules, which means that natural selection produces new objects by gradually modifying preexisting ones. Natural selection, in other words, creates only *relative* novelties, not absolute ones. In the case of coding, instead, the situation is totally different. The rules of a code are not dictated by physical necessity, and this means that they can establish relationships that have never existed before in the Universe. Natural conventions, in short, have the potential to create *absolute* novelties. Another distinction between copying and coding is that they involve two different entities. A variation in the copying of a gene changes the linear sequence, i.e., the *information* of that gene. A variation in a coding rule changes instead the *meaning* of that rule. The great difference that exists between copying and coding, and therefore between natural selection and natural conventions, comes from the difference that exists between 'information' and 'meaning'.

There are, in short, three major differences between copying and coding: (1) copying acts on individual objects whereas coding acts at the collective level, (2) copying modifies existing objects whereas coding brings new objects into existence, and (3) copying is about biological information whereas coding is about biological meaning. All of which means that there are two distinct types of evolutionary change in life: evolution by natural selection, based on copying, and evolution by natural conventions, based on coding.

We arrive in this way at the conclusion that coding is not reducible to copying or to self-assembly, and the close relationship that exists between organic codes and macroevolution tells us that coding is indeed a fundamental mechanism. Code biology is truly a new science in the making, the exploration of a vast and still largely unexplored dimension of the living world, the real new frontier of biology.

Acknowledgments I am most grateful to all members of the Biosemiotics Board who have sent comments on this paper. I thank in particular Stanley Salthe, Anna Aragno and Joachim De Beule for their very useful suggestions.

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2007). *Molecular biology of the cell* (5th ed.). New York: Garland.

Andersson, J. O. (2005). Lateral gene transfer in eukaryotes. *Cellular and Molecular Life Sciences, 62*, 1182–1197.

Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B. J., & Frey, B. J. (2010). Deciphering the splicing code. *Nature, 465*, 53–59.

Barbieri, M. (1981). The ribotype theory on the origin of life. *Journal of Theoretical Biology, 91*, 545–601.

Barbieri, M. (1985). *The semantic theory of evolution*. London & New York: Harwood Academic Publishers.

Barbieri, M. (1998). The organic codes. The basic mechanism of macroevolution. *Rivista di Biologia-Biology Forum, 91*, 481–514.

Barbieri, M. (2003). *The organic codes. An introduction to semantic biology.* Cambridge: Cambridge University Press.

Barbieri, M. (2008). Biosemiotics: a new understanding of life. *Naturwissenschaften, 95*, 577–599.

Barbieri, M. (2011). Origin and evolution of the brain. *Biosemiotics, 4*(3), 369–399.

Burks, A. W. (1970). *Essays on cellular automata*. Urbana: University of Illinois Press.

Cech, T. R. (1983). RNA splicing: three themes with variations. *Cell, 34*, 713–716.

Cech, T. R. (1986). RNA as an enzyme. *Scientific American, 255*, 64–75.

Dagan, T., Artzy-Randrup, Y., & Martin, W. (2008). Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proceedings of the National Academy of Science USA, 105*, 10039–10044.

Darwin, C. (1859). *On the origin of species by means of natural selection*. London: John Murray.

Dhir, A., Emanuele Buratti, E., van Santen, M. A., Lührmann, R., & Baralle, F. E. (2010). The intronic splicing code: multiple factors involved in ATM pseudoexon definition. *The EMBO Journal, 29*, 749–760.

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science, 284*, 2124–2129.

Doolittle, W. F., & Bapteste, E. (2007). Pattern pluralism and the Tree of Life hypothesis. *Proceedings of the National Academy of Sciences USA, 104*, 2043–2049.

Eigen, M. (1971). Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften, 58*, 465–523.

Eigen, M., & Schuster, P. (1977). The hypercycle. A principle of natural self-organization. *Naturwissenschaften, 64*, 541–565.

Flames, N., Pla, R., Gelman, D. M., Rubenstein, J. L. R., Puelles, L., & Marin, O. (2007). Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. *The Journal of Neuroscience, 27*(36), 9682–9695.

Gabius, H.-J. (2000). Biological information transfer beyond the genetic code: the sugar code. *Naturwissenschaften, 87*, 108–121.

Gabius, H.-J. (2009). *The sugar code. Fundamentals of glycosciences*. Wiley-Blackwell.

Gamble, M. J., & Freedman, L. P. (2002). A coactivator code for transcription. *Trends in Biochemical Sciences, 27*(4), 165–167.

Gilbert, W. (1986). The RNA world. *Nature, 319*, 618.

Gilbert, S. F. (2006). *Developmental biology* (8th ed.). Sunderland: Sinauer.

Guerrier-Takada, C., & Altman, S. (1984). Catalytic activity of an RNA molecule prepared by transcription in vitro. *Science, 223*, 285–286.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., & Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell, 35*, 849–857.

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology, 3*, 318–356.

Jessell, T. M. (2000). Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nature Genetics, 1*, 20–29.

Johannsen, W. (1909). *Elemente der exacten Erblichkeitslehre*. Jena: Gustav Fisher.

Knoll, A. H. (2003). *Life on a young planet. The first three billion years of evolution on Earth*. Princeton: Princeton University Press.

Linde Medina, M. (2010). Two “EvoDevos”. *Biological Theory, 5*(1), 7–11.

Luisi, P. L., & Varela, F. J. (1989). Self-replicating micelles – a chemical version of a minimal autopoietic system. *Origins of Life and Evolution of the Biosphere, 19*, 633–643.

Marquardt, T., & Pfaff, S. L. (2001). Cracking the transcriptional code for cell specification in the neural tube. *Cell, 106*, 651–654.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: the realisation of the living*. Dordrecht, Holland: D. Reidel Publishing Company. 1980.

Maynard Smith, J., and Szathmáry, E. (1995). *The major transitions in evolution.* Oxford University Press.

Mesarovic, M. D. (1968). *Systems theory and biology.* Berlin: Springer.

Miller, S. L. (1953). A production of amino acids under possible primitive earth conditions. *Science, 117*, 528–529.

Monod, J. (1970). *Chance and necessity.* New York: A. Knopf.

Morowitz, H. J. (1992). *Beginnings of cellular life.* Yale University Press.

Nicolelis, M., & Ribeiro, S. (2006). Seeking the neural code. *Scientific American, 295*, 70–77.

Niesert, U., Harnasch, D., & Bresch, C. (1981). Origin of Life between Scylla and Charybdis. *Journal of Molecular Evolution, 17*, 348–353.

Nitta, I., Kamada, Y., Noda, H., Ueda, T., & Watanabe, K. (1998). Reconstitution of peptide bond formation. *Science, 281*, 666–669.

Perissi, V., & Rosenfeld, M. G. (2005). Controlling nuclear receptors: the circular logic of cofactor cycles. *Nature Molecular Cell Biology, 6*, 542–554.

Pertea, M., Mount, S. M., & Salzberg, S. L. (2007). A computational survey of candidate exonic splicing enhancer motifs in the model plant *Arabidopsis thaliana*. *BMC Bioinformatics, 8*, 159.

Readies, C., & Takeichi, M. (1996). Cadherine in the developing central nervous system: an adhesive code for segmental and functional subdivisions. *Developmental Biology, 180*, 413–423.

Rosen, R. (1958). A relational theory of biological systems. *Bulletin of Mathematical Biophysics, 20*, 245–260.

Rosen, R. (1991). *life itself: a comprehensive inquiry into the nature, origin, and fabrication of life*. New York: Columbia University Press.

Schopf, J. W. (1999). *Cradle of life: the discovery of earth's earliest fossils*. Princeton University Press.

Segré, D., Ben Eli, D., Deamer, D. W., & Lancet, D. (2001). The lipid world. *Origins of Life and Evolution of the Biosphere, 31*, 119–145.

Shapiro, L., & Colman, D. R. (1999). The diversity of Cadherins and implications for a synaptic adhesive code in the CNS. *Neuron, 23*, 427–430.

Spemann, H. (1901). Entwicklungphysiologische Studien am Tritonei I. *Wilhelm Roux' Archiv für Entwicklungsmechanik., 12*, 224–264.

Spiegelman, S. (1967). An in vitro analysis of a replicating molecule. *American Scientist, 55*, 3–68.

Strahl, B. D., & Allis, D. (2000). The language of covalent histone modifications. *Nature, 403*, 41–45.

Szathmáry, E. (1999). Chemes, genes, memes: a revised classification of replicators. *Lectures on Mathematics in the Life Sciences., 26*, 1–10.

Tomkins, M. G. (1975). The metabolic code. *Science, 189*, 760–763.

Trifonov, E. N. (1987). Translation framing code and frame-monitoring mechanism as suggested by the analysis of mRNA and 16s rRNA nucleotide sequence. *Journal of Molecular Biology, 194*, 643–652.

Trifonov, E. N. (1989). The multiple codes of nucleotide sequences. *Bulletin of Mathematical Biology, 51*, 417–432.

Trifonov, E. N. (1996). Interfering contexts of regulatory sequence elements. *Cabios, 12*, 423–429.

Trifonov, E. N. (1999). Elucidating sequence codes: three codes for evolution. *Annals of the New York Academy of Sciences, 870*, 330–338.

Tudge, C. (2000). *The variety of life. A survey and a celebration of all the creatures that have ever lived*. Oxford and New York: Oxford University Press.

Turner, B. M. (2000). Histone acetylation and an epigenetic code. *BioEssay, 22*, 836–845.

Turner, B. M. (2002). Cellular memory and the Histone code. *Cell, 111*, 285–291.

Verhey, K. J., & Gaertig, J. (2007). The Tubulin code. *Cell Cycle, 6*(17), 2152–2160.

von Bertalanffy, L. (1969). *General system theory.* New York: George Braziller.

von Neumann, J. (1966). *Theory of self-reproducing automata*. Urbana: University of Illinois Press.

Wiener, N. (1948). *Cybernetics: or control and communication in the animal and the machine*. Paris: Hermann.

Woese, C. R. (1987). Bacterial evolution. *Microbiology Reviews, 51*, 221–271.

Woese, C. R. (2000). Interpreting the universal phylogenetic tree. *Proceedings of the National Academy of Sciences USA, 97*, 8392–8396.

Woese, C. R. (2002). On the evolution of cells. *Proceedings of the National Academy of Sciences USA, 99*, 8742–8747.

Wong, J. T. F. (1975). A co-evolution theory of the genetic code. *Proceedings of the National Academy of Sciences USA, 72*, 1909–1912.

Wong, J. T. F. (1980). Role of minimization of chemical distances between amino acids in the evolution of the genetic code. *Proceedings of the National Academy of Sciences USA, 77*, 1083–1108.