

Organic codes: Metaphors or realities?

Marcello Barbieri

Department of Morphology and Embryology, University of Ferrara
Via Fossato di Mortara 64, 44100 Ferrara, Italy
e-mail: brr@unife.it

Abstract. Coding characteristics have been discovered not only in protein synthesis, but also in various other natural processes, thus showing that the genetic code is not an isolated case in the organic world. Other examples are the sequence codes, the adhesion code, the signal transduction codes, the splicing codes, the sugar code, the histone code, and probably more. These discoveries however have not had a significant impact because of the widespread belief that organic codes are not real but metaphorical entities. They are supposed to lack arbitrariness and codemakers, the two qualifying features of real codes. Here it is shown that the arbitrariness issue can be solved on an experimental basis, while the codemaker issue is dependent on our theoretical description of the cell and can only be solved by a new concept. In order to appreciate the reality of the organic codes, in short, it is necessary to have not only a more critical evaluation of the experimental data but also a new theory of the living system.

Introduction

From time immemorial it has been thought that codes, or conventions, exist only in the world of culture. The discovery of the genetic code, in the 1960s, came therefore as a bolt from the blue, but the reaction was rather strange. The discovery of one organic code should have suggested that there could be more in nature, but what happened was the exact opposite. The genetic code was immediately declared a *frozen accident*, and any mention of other organic codes was ignored. Edward Trifonov, for example, has shown since the 1980s that there are at least three sequence codes in addition to the classic triplet code, but in vain. The situation started to change only in the late 1990s. In

1996, Redies and Takeichi described an *adhesive code* in the development of the nervous system, and in the year 2000, Gabius provided evidence for a *sugar code*, while Strahl, Allis, Turner and colleagues discovered a *histone code* (Table 1).

1	THE GENETIC CODE	(1954–1966)	<i>Gamow, Nirenberg, Khorana</i>
2	THE SEQUENCE CODES	(1988–1999)	<i>Trifonov</i>
3	THE ADHESION CODE	(1996)	<i>Redies, Takeichi</i>
4	THE SIGNAL TRANSDUCTION CODES	(1998)	<i>Barbieri</i>
5	THE SPLICING CODES	(1998)	<i>Barbieri</i>
6	THE SUGAR CODE	(2000)	<i>Gabius</i>
7	THE HISTONE CODE	(2000)	<i>Strahl, Allis, Turner</i>

Table 1. Organic codes, and their describers.

These announcements, however, have barely raised an interest. Today, the existence of other organic codes is no longer ignored as it was in the past, but it is not seen as anything special. This response may appear surprising, but is not unfounded. It is the natural consequence of a widespread and deep-seated belief that all organic codes, including the genetic code, are only useful *metaphors*, not real entities. Molecular biology has borrowed many words from ordinary language, because they have an intuitive appeal and avoid long periphrases, but they are not meant to be literally true. The genetic code itself is given the name “code” only because this term is metaphorically appropriate, but deep down most biologists are convinced that it is nothing more than a good metaphor. And this for two basic reasons. The real codes that we are familiar with have two outstanding features: they are *arbitrary* rules, and they are made by a *codemaker*. These are the key entities: *arbitrariness* and *codemaking*. No code can be a real code without these qualifying features, and most biologists are convinced that organic codes simply do not have them. This is the crucial point: why do people believe that organic codes do not have those two qualifying features?

The codes' fingerprints

A code is a set of rules that establish a correspondence between two independent worlds. The Morse code, for example, is a correspondence between combinations of dots and dashes with the letters of the alphabet, and — in the same way — the genetic code is a correspondence between combinations of nucleotides and amino acids. From the point of view of the definition, there is no difference between them. Why then do people believe that the Morse code is real and the genetic code is not? One reason, as we have seen, is *arbitrariness*. We know that the Morse code is arbitrary because we have built it ourselves, and we are certain that there is no necessary link between dots and dashes and the letters of the alphabet. But ask a biologist if the same arbitrariness exists between nucleotides and amino acids, and you are likely to get a very different response. Many would deny it out of hand, others would say that the two codes are not comparable, and some would reply that we still need more data.

One of the most common arguments against the arbitrariness of the genetic code is the determinism of protein synthesis. Every single step of the translation process is perfectly deterministic, in the sense that a chain of nucleotides is translated into a chain of amino acids with a precise sequence of reactions. This is the most popular argument, probably because it has a strong intuitive appeal, and yet it is not a valid one. The same determinism, in fact, is present even when cultural codes are implemented. When the mental image of an apple is formed in the visual cortex and we pronounce the word “apple”, there is a precise chain of neurological reactions between the two mental images. A neurologist would say with no hesitation that the neural connection between the visual area and the speech area of the brain is perfectly deterministic, and yet the connection was established by a linguistic code that is perfectly arbitrary. The implementation of the rules of a code, in short, is deterministic in all codes, even in the cultural ones. The arbitrariness comes in only when a code is created or modified, not when it is implemented.

We need therefore positive evidence in order to obtain reliable conclusions, and it is the very definition of the codes that tells us what to look for. Since a code is a bridge between two independent worlds, an organic code necessarily requires organic molecules that perform two independent recognition steps. These are the “adaptors”, the name that Francis Crick proposed for the molecules that today we call

transfer RNAs. All codes need molecules that perform equivalent functions, and so all these molecules can be called adaptors. The adaptors are catalysts that have two different recognition sites, and what qualifies them as adaptors is the fact that there is no necessary connection between the two sites. The site which recognises the objects of one world can be associated with any of the sites that recognise the objects of the other world, and this means that a connection can only be established by an arbitrary choice, by a “natural convention”. The adaptors, in short, are the “fingerprints” that reveal the presence of an organic code.

In the case of the genetic code, it has been possible to prove that the nucleotide site is independent from the amino acid site by actually changing the rules of the code *in vitro*, and a similar experiment has been performed *in vivo* by some micro-organisms. This should have settled the arbitrariness issue for good, but ingrained opinions are hard to die, and so we still hear the claim that the association between nucleotides and amino acids is not arbitrary, because some regularities have been discovered in the genetic code. This is true, but it has nothing to do with arbitrariness, and in fact regularities also exist in cultural codes. In the Morse code, for example, the most frequent letters of the alphabet are associated with the simplest combinations of dots and dashes, but nobody would dream to conclude that the Morse code is not a true code for that.

In the case of the genetic code, furthermore, there are also other factors in favour of its arbitrariness. The number and the types of the amino acids, for example, could have been different, because many other amino acids exist in nature, and the same is true for the nucleotides. In the genetic code, in short, we find arbitrariness not only in the rules of the code, but also in the choice of the objects which are coded by those rules. And this is perfectly equivalent to what happens in the linguistic codes, where arbitrariness exists not only in the rules of grammar, but also in the number and in the type of letters which are chosen to make up an alphabet.

The arbitrariness of the organic codes, in conclusion, can be demonstrated by a variety of experimental facts, and above all by the existence of adaptors (it was the presence of adaptors in signal transduction and in splicing, that allowed me to conclude, in 1998, that these processes are based on organic codes). Arbitrariness alone, however, is not enough, because it could be the result of an extraordinary number of coincidences. A real code requires arbitrariness and codemakers, and the existence of a codemaker is an issue where

theory plays an even greater role than experiments. It is also to theory, therefore, that we need to turn our attention.

The third party

The extraordinary thing about codes is that they require a new entity. In addition to energy and information they require *meaning*. For centuries, meaning has been regarded as a spiritual or a transcendental entity, but in reality it is a perfectly natural entity because we can define it with an operative procedure just as we do with all physical quantities. *Meaning is an object which is related to another object by a code*. The meaning of the word *apple*, for example, is the mental object of the fruit which is associated to the mental object of that word by the code of the English language. More in general, a cultural meaning is always a mental object which is associated to another mental object by a convention. But the operative definition of meaning need not be restricted to the mental world because it applies equally well to the organic world. The meaning of a combination of dots and dashes is a letter of the alphabet, in the Morse code. And in the same way, the meaning of a combination of three nucleotides is usually an amino acid, in the genetic code (from which it follows that the meaning of a gene is usually a protein).

We are well aware that it is man who gives meaning to mental objects — in the realm of the mind he is the *codemaker* — but this does not mean that a code of correspondence between two independent worlds must be produced by a conscious activity. The only logical necessity is that the codemaker is *an agent which is ontologically different* from those worlds, because if it belonged to one of them the two worlds would no longer be independent. A code, in other words, requires three entities: two independent worlds and a code-maker which belongs to a third world (from a philosophical point of view this is equivalent to the triadic system proposed in semiotics by Charles Peirce).

The problem is that the cell is described as a dualistic system of genes and proteins, genotype and phenotype, software and hardware, and in a dualistic system there is no third party that can act as a codemaker. This is why I proposed, in 1981, that the cell is not a duality of genotype and phenotype but a trinity made of genotype, phenotype and *ribotype*. The ribotype was defined as the ribonucleo-

protein system of the cell, and it was underlined that it represents *a new cell category*. As phenotype is the seat of metabolism and genotype the seat of heredity, so ribotype is the seat of genetic coding.

It is an experimental fact that the genetic code is implemented by ribonucleoproteins, and this strongly suggests that the ribotype is the codemaker of the genetic code, but it does not prove it. Only a theory can establish the ontological status of the ribotype as an independent cell category. We have therefore before us two very different concepts: the cell as a duality (*the genotype–phenotype theory*) or the cell as a trinity (*the ribotype theory*). The problem is how to choose between them.

The origin-of-life metaphors

The evaluation of theories is a complex affair, in general, but there are theories which can be illustrated by metaphors, and in these cases the metaphors should be discussed first, because their intuitive appeal often takes priority over rational thinking. In our case, a theory of the cell can be illustrated by a metaphor on the origin of life, because the nature and the origin of a system are two faces of the same problem. If the cell is a duality of genotype and phenotype, for example, the problem of the origins is understanding whether it was the genes or the proteins which came first. The genotype–phenotype theory, in other words, corresponds to *the-chicken-and-the-egg* metaphor on the origin of life. In this framework, it doesn't even make sense to speak of three categories, and so the ribotype theory had to be illustrated by a totally different metaphor. More precisely, by *the-cell-as-a-city* metaphor, where the proteins of the cell are compared to the houses of a city, and the genes to their blueprints (Barbieri 1981; 1985). In this framework, it is the chicken-and-the-egg problem that makes no sense, because it would be equivalent to asking if it was the houses or the blueprints which came first, and either answer would be wrong. What came first was a third party, the inhabitants, i.e. the intermediaries between houses and blueprints in a city which correspond to the intermediaries between proteins and genes in a cell.

Our theories of the cell are illustrated therefore by different metaphors on the origin of life, and it may be worthwhile to examine them in some detail. As a matter of fact, as soon as we take a closer look at *the-chicken-and-the-egg* metaphor, we realise that there is

something wrong with it. The egg and the chicken are not the two faces of one duality. They are two dualistic systems in different stages of development. Each one of them is a complete genotype–phenotype entity, and it is pure fiction to say that one represents the genotype and the other stands for the phenotype.

We do indeed need a better metaphor, and *the-cell-as-a-city* does have a certain intuitive appeal. But this metaphor has not become anything like as popular as *the-chicken-and-the-egg*, and it is highly instructive to understand why. The crucial point is that in a city only the inhabitants are alive, whereas houses and blueprints are not. The city metaphor, in other words, implies that *genes and proteins are molecular artifacts, just as blueprints and houses are human artifacts*. And this seems a preposterous idea. How can one accept that genes and proteins, the very molecules of life, are inanimate manufactured objects? That probably explains why the ribotype theory has not attracted the attention of the origin-of-life people. And yet it has never been proved that the preposterous idea is false. It may be interesting therefore to take a look at it.

Copymakers and codemakers

There was a time when atoms did not exist. They came into being within giant stars, and were scattered all over the place when those stars exploded. There was a time when molecules did not exist. They originated from the combination of atoms on a variety of different places such as comets and planets. There was a time when polymers did not exist. They were produced when molecules joined together at random and formed chains of subunits. There was a time when all the polymers of our planet were random molecules, but that period did not last forever. At a certain point, new types of polymers appeared. Some molecules started making copies of polymers, and for this reason I call them *copymakers*. Other molecules made coded versions of the copies, and I refer to them as *codemakers*. On the primitive Earth, the copy-makers could have been RNA-replicas and the codemakers could have been transfer-RNAs, but other possibilities exist, and so here we will use the generic terms of copymakers and codemakers. All that matters, for our purposes, is the historical fact that copymakers and codemakers came into being and started producing copied molecules and coded molecules.

Now let us take a look at these new polymers. The formation of a random chain of subunits is accounted for by the laws of thermodynamics and does not require any new physical quantity. But when a copymaker makes a copy of that chain, something new appears: the sequence of subunits becomes *information* for the copymaker. In a similar way, when a codemaker takes a chain of monomers of one kind to produce a chain of monomers of a different kind, something new appears: the second chain becomes the *meaning* of the first one. *It is only the act of copying that creates information, and it is only the act of coding which creates meaning.* Information and meaning, in other words, appeared in the world when copymakers and codemakers came into existence and started functioning.

The appearance of copied polymers and coded polymers was a major event also for another reason. Up to that point, all molecules formed on the primitive Earth had one thing in common: their structure was entirely determined by the assembly properties of their atoms, i.e., *from within*. In the case of copied and coded polymers, in contrast, the order of the subunits was determined by external templates, i.e., *from without*. In everyday language, we distinguish between *natural* and *artificial* products in a straightforward way: the objects which are formed spontaneously are natural, while those which are shaped by external agents are artificial. And that is precisely the distinction that exists between random polymers on one hand and copied or coded polymers on the other. I conclude therefore that copied molecules (genes) and coded molecules (proteins) are indeed, in a very deep sense, *artificial* molecules. They are artificial because they are produced by external agents, because their primary structure is determined from without and not from within, because their production involves outside processes based on information and meaning.

There was a time when the world was inhabited only by *natural* molecules, but that period did not last forever. At a certain point copied and coded molecules appeared, and the world became also inhabited by *artificial* molecules — by *artifacts made by nature*. And that was not just another step toward life. It was the appearance of the very logic of life because, from copymakers and codemakers onward, all living creatures have been artifact-makers. In a very fundamental sense, we can define life itself as *artifact-making*.

The handicapped replicator

The *cell-as-a-city* metaphor suggests that proteins and genes are artificial molecules, and we have just seen that, deep down, that is precisely what they are. The metaphor also suggests that modern cells are to primitive cells what large cities are to small villages, and this is not an unreasonable analogy. Modern eukaryotic cells, for example, contain millions of ribosomes, like the inhabitants of large cities, while prokaryotic cells have only hundreds or thousands of ribosomes, like the inhabitants of villages.

The metaphor can also be extended to earlier stages of evolution. If the origin of the first cells is likened to the origin of the first villages, we can compare the age of precellular evolution to the period of history in which villages did not exist. The interesting point is that this metaphor allows us to take a closer look at today's most popular model on precellular evolution: the model of the naked gene as the first replicator (Dawkins 1976).

Dawkins has readily admitted that genes are not doing any replication, but since they code for the molecules that replicate them, he finds it legitimate to call them "replicators" in order to avoid long periphrases. Michael Ghiselin (1997) has pointed out that this is confusing the "object" with the "agent" of replication, but Dawkins' use of the word has stuck, and today most biologists seem to be taking for granted that genes are replicators. This is why I have avoided that word altogether and I have used the term *copymakers*. The distinction between copymakers and copies is still alive and well, and so there is no danger of confusing what is copied with what does the copying. Whatever one's choice of words, however, the real point is the substance, not the terminology.

The substance of the replicator model is that *all that matters in life is information, and all that matters in evolution is the replication of information with occasional mistakes*. But at the heart of life there are two fundamental entities, not one. Information and meaning are two independent entities, copying and coding are two independent processes, and the codemaker between genes and proteins must be a third party because otherwise there would be no real code. The replicator model is not wrong, but incomplete (or handicapped), because what matters in life is *replication and coding*, not replication alone (I prefer to speak of *copying and coding*, but the message is the same). The replicator model would be right if the cell were a von Neumann

automaton where the hardware is completely described by the software, and information is really everything, but nature has not taken that path. And probably for very good reasons, because that path was seriously undermined by the error catastrophes.

One could still argue, however, that a “naked gene” phase should have preceded a phase of “copying-and-coding”, and this is where the cell-as-a-city metaphor can help us. The metaphor suggests that before cities there were villages, that before villages there were humans living in the open, that before humans there were ancestral hominids, and so on. The point is that in all stages there were “agents” not just “objects”. There has never been a time in precellular evolution in which copied molecules (genes) could exist without copymakers, or coded molecules (proteins) without codemakers. It was copymakers and codemakers which came first, because they were the first “agents” in the history of life. The first molecules of the ribotype world were produced by random processes and the chances of getting copymakers or codemakers (for example, RNA-replicases or transfer-RNAs) were not substantially different. Any one could have appeared before the other, without making much difference. What did make a difference was the appearance of both of them because only their combination created a renewable link between genes and proteins. It was a ribotypic system containing copymakers and codemakers that started life, because that was the simplest possible *lifemaker*, i.e., the simplest *agent*. Admittedly, a naked gene would have been a simpler system but it would not have been an *agent*, and that makes all the difference. As Einstein once remarked, “*things should be made as simple as possible, but not simpler*”.

Conclusion

There are experimental facts (the adaptors) and theoretical concepts (the ribotype) which show that organic codes have the two qualifying features of all real codes (arbitrariness and codemakers). But adaptors and ribotype are still largely ignored, and so it is not surprising that most biologists continue to believe — in perfect good faith — that organic codes do not really exist out there. Which is rather reassuring, in a way, because it shows that even in this age of high technology what we see in nature is what our theories allow us to see.

References

- Barbieri, Marcello 1981. The ribotype theory on the origin of life. *Journal of Theoretical Biology* 91: 545–601.
- 1985. *The Semantic Theory of Evolution*. London: Harwood Academic Publishers.
- 1998. The organic codes: The basic mechanism of macroevolution. *Rivista di Biologia — Biology Forum* 91: 481–514.
- 2003. *The Organic Codes: An Introduction to Semantic Biology*. Cambridge: Cambridge University Press.
- Dawkins, Richard 1976. *The Selfish Gene*. Oxford: Oxford University Press.
- Gabius, Hans-Joachim 2000. Biological information transfer beyond the genetic code: The sugar code. *Naturwissenschaften* 87(3): 108–121.
- Gamble, Matthew J.; Freedman, Leonard P. 2002. A coactivator code for transcription. *Trends in Biochemical Sciences* 27(4): 165–167.
- Gamow, George 1954. Possible relation between deoxyribonucleic acid and protein structure. *Nature* 173: 318.
- Ghiselin, Michael 1997. *Metaphysics and the Origin of Species*. Albany: State University of New York Press.
- Jenuwein, Thomas; Allis, C. David 2001. Translating the histone code. *Science* 293: 1074–1080.
- Khorana, Har Gobind; Büchi, H.; Ghosh, H.; Gupta, N.; Jacob, T. M.; Kössel, H.; Morgan, R.; Narang, S. A.; Ohtsuka, E.; Wells, R. D. 1966. Polynucleotide synthesis and the genetic code. *Cold Spring Harbor Symposia on Quantitative Biology* 31: 39–49.
- Nirenberg, Marshall W.; Matthaei, J. H. 1961. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences of the USA* 47: 1588–1602.
- Nirenberg, Marshall W.; Caskey, T.; Marshall, R.; Brimacombe, R.; Kellogg, D.; Doctor, B.; Hatfield, D.; Levin, J.; Rottman, F.; Pestka, S.; Wilcos, M.; Anderson, F. 1966. The RNA code and protein synthesis. *Cold Spring Harbor Symposia on Quantitative Biology* 31: 11–24.
- Redies, Christoph; Takeichi, Masatoshi 1996. Cadherine in the developing central nervous system: An adhesive code for segmental and functional subdivisions. *Developmental Biology* 180: 413–423.
- Strahl, Brian D.; Allis, C. David 2000. The language of covalent histone modifications. *Nature* 403: 41–45.
- Trifonov, Edward N. 1988. Codes of nucleotide sequences. *Mathematical Biosciences* 90: 505–517.
- 1989. The multiple codes of nucleotide sequences. *Bulletin of Mathematical Biology* 51: 417–432.
- 1999. Elucidating sequence codes: Three codes for evolution. *Annals of the New York Academy of Sciences* 870: 330–338.
- Turner, Bryan M. 2000. Histone acetylation and an epigenetic code. *BioEssays* 22(9): 836–845.

Органические коды: метафоры или реальность?

Характеристики кодирования найдены не только у синтеза белка, но и в ряде других процессов живой природы. Это доказывает, что генетический код не является чем-то исключительным для органического мира. В качестве примеров можно привести коды последовательностей, адгезивный код, коды трансдукции сигнала, сплайсинг-коды, сахарный код, гистонный код, и еще ряд других. Эти открытия все же не имеют пока заметного влияния из-за широко распространенной предпосылки о метафорическом характере этих кодов, об отсутствии у них арбитранности и образователей кода. Последние считаются необходимыми характеристиками реальных кодов. В настоящей работе показано, что свойство арбитранности можно доказать экспериментально, а проблема образователей кода зависит от нашего теоретического представления о клетке и решается, следовательно, с помощью нового понятийного аппарата. Итак, для признания органических кодов требуется не только более критическая оценка экспериментального материала, но и новая теория живой системы.

Orgaanilised koodid: metafoorid või tõelused?

Kodeerimise tunnuseid pole leitud ainult valgusünteesis, vaid ka mitmetes teistes eluslooduse protsessides. See tõendab, et geneetiline kood pole erandlik juhtum orgaanilises maailmas. Teisteks näideteks on järjestuskoodid, adhesiooni kood, signaali transduktsiooni kood, splaissingu koodid, suhkru kood, histoonide kood ja ilmselt veelgi. Neil avastustel pole siiski olnud märkimisväärset mõju laialt levinud uskumuse tõttu, et orgaanilised koodid pole mitte tõelised, vaid on pigem metafoorsed. Arvatakse, et neil puuduvad arbitraarsus ja koodi moodustajad — kaks reaalse koodi tunnust. Käesolevas töös näidatakse, et arbitraarsuse küsimuse saab lahendada eksperimentaalsel alusel, samas kui koodi moodustaja küsimus sõltub raku teoreetilisest kirjeldusest ning on lahendatav vaid uue mõisteaparaadi kaasabil. Niisiis, orgaaniliste koodide tunnustamiseks on vaja mitte ainult eksperimentaalse andmestiku kriitilisemat hindamist, vaid ka elussüsteemi uut teooriat.