Code Biology

Marcello Barbieri

Code Biology

A New Science of Life

Springer

Marcello Barbieri
University of Ferrara
Dipartimento di Morfologia ed Embriologia
Ferrara, Italy

ISBN 978-3-319-14534-1 ISBN 978-3-319-14535-8 (eBook)
DOI 10.1007/978-3-319-14535-8

Library of Congress Control Number: 2014959281

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

To my wife
Leopoldine Margarethe
the girl from Klagenfurt
who fell in love with Italy
and saved my life

Foreword

You approach a set of red traffic lights and you stop. Why? You have read the previous sentence and understood its meaning. How? It is because you have learnt how to associate symbols with their conventional meanings—you understand a particular encoded relationship. These are examples of cultural codes, but since the 1960s we know that a true molecular code exists in every cell: the molecular code called the genetic code, which translates the four-letter nucleotide sequence of DNA into the 20-letter amino acid sequence in proteins. What Marcello Barbieri will tell you in this ground-breaking book is that between the development of the genetic code at the beginning of life on earth and the development of cultural codes and languages some four billion years later many other organic codes were established. This is in itself already a major discovery, but its implications are perhaps even more important, going right to the heart of biology. The establishment of each of these organic codes opened up a new world of possibilities for natural selection to explore, possibilities that did not exist before. We can link each of the major transitions in evolutionary history to an organic code. In short, Barbieri will give you a new lens through which to view life and its evolution. Whether in the end you accept this new code view of life will prove to be immaterial—this book will have challenged your current view of biology and may have even changed it irrevocably. As did the discovery of Barbieri's work a decade ago change mine.

I am a biochemist morphed into a systems biologist, or, as I prefer it, a molecular cell physiologist. Since the early 1980s I have been involved in a quest to understand the functional organisation and regulation of cellular processes, especially of metabolism. In 2004, I was on a year-long sabbatical, attempting to integrate the approaches of Robert Rosen, John von Neumann and Howard Pattee into a coherent picture of the cell as a self-fabricating system. I had come to this point both through my research on metabolic control and regulation and by reading Rosen's *Life Itself*, a book that changed my view of biology profoundly. But I felt that a piece of the puzzle was missing. What bothered me as a biochemist was that the link between genotype and phenotype was not made clear enough, whether in Rosen's metabolism-repair systems, or in Von Neumann's distinction between a self-reproducing system and its description, or in Pattee's matter-symbol complementarity. Of course all three of these brilliant scientists were perfectly aware of the connection, but did not make it explicit and this set me off on a search for the "missing link". I found the answer during a search for definitions of life. *The Organic Codes: An Introduction to Semantic Biology* by Marcello Barbieri contains a compendium of 63 such definitions. However, my interest in this set of definitions was eclipsed by the excitement of reading the book itself. Here was exactly what I was looking for. The key was Barbieri's concept of the ribotype as the necessary link between the two independent worlds of genotype and phenotype. As Michael Ghiselin said in his foreword to the book, "Previous scenarios treated proteins or DNA as coming first. Both of these alternatives ran into difficulties because the one cannot exist without the other. For that very reason there must have been something additional to genotype and phenotype, which he calls the ribotype. It is RNA that bridges the gap between genotype and phenotype, and it does so by endowing the system with meaning." Meaning? What was "meaning" doing in a biological text? In the biological paradigm I grew up in such a question could not even be asked, and if it was, would be laughed out of the room. Nevertheless, I was hooked, and this question and his answer to it set me off on an epistemological adventure that eventually led to my meeting Marcello, becoming a close friend and collaborator, being welcomed into the biosemiotics community, and, finally, being part in 2013 of the creation of the International Society of Code Biology and attending its first conference in 2014. I have therefore been privy to the creation of the book that you are holding in your hands, a book that will be the standard text of this new discipline for years to come. In it you will find the answer to "What is organic meaning?" and, paraphrasing Woody Allen, to many other questions you were afraid to ask.

Code Biology: A New Science of Life is a remarkable record of one man's quest to understand life from the viewpoint of codes. In the modern era of big science, research consortia, and the race for funding it is rare to find someone who, over solitary decades, pieces together strands of evidence and avenues of thought into a grand synthesis such as the one presented here. One can follow the thread: it started in the 1970s with Barbieri's proposal of a primitive ribosome model developed while at the Max Planck Institute for Molecular Genetics in Berlin (where Ada Yonath, who won the Nobel prize for chemistry in 2009 for her structural studies of the ribosome, worked alongside him). This model led to his ribotype theory on the origin of life published in 1981 and his 1985 book *The Semantic Theory of Evolution*. However, biology was in the thick of the recombinant DNA era and took no notice. Two more decades passed before Barbieri came to the attention of the biosemiotics community when he sent a first version of *The Organic Codes* to Thomas Sebeok, one of the fathers of modern biosemiotics. How the story unfolds from there on is told in Chap. 9. The ensuing relationship was very productive and lasted a number of years, but there was always an underlying discomfort that ultimately led to a parting of ways. Barbieri insisted from the start that organic meaning results from a mechanistic process of decoding, whereas the Peircian biosemiotic view was that all meaning is produced by interpretation. Barbieri refused to relinquish his strict adherence to a scientific approach, to relinquish the belief that the understanding of organic codes and organic meaning must be in terms of scientific models that can be tested.

I expect that a viewpoint as strong as Barbieri's is bound to be controversial in some circles. If you ascribe to the physicalist's "life is just chemistry", the molecular biologist's "life is chemistry plus information", or the Peircian biosemiotician's "the units of life are signs"—views that are described in Chap. 1—you may be rankled by the suggestion that these views are inadequate for explaining life. But they *are* inadequate, as Barbieri argues convincingly. The key lies in his realisation that material life as we know it is "artifact-making", that organisms are fabricators, ultimately of themselves. This was also Robert Rosen's main conclusion; his analysis in terms of Aristotelean causes led him to the statement that organisms are "closed to efficient causation", by which he meant that organisms autonomously manufacture all the molecular machinery necessary for life. The closure follows because these machines can produce the polynucleotides and proteins from which they themselves are formed. As explained by Barbieri, each of these two classes of highly specific biopolymers requires its own template-directed manufacturing process: copying for polynucleotides and coding for proteins. These two processes in turn form the basis for two types of evolutionary process: evolution by natural selection through copying and evolution by natural conventions through coding. We of course know and accept the mechanism of natural selection through copying as the basis for Darwinian evolution and the creation of relative novelties, as Barbieri calls them, but the coming into existence of absolute novelties has been a problem from the start. That we no longer have to resort to hand-waving arguments to explain absolute novelties is one of the main contributions of the new science of code biology: evolution by natural conventions through new organic codes provides the required mechanism.

Code biology unearths many other riches besides a novel mechanism of evolution. In the journey from life before the cell through all the major stages of evolution to the origins of human language, Barbieri touches on a wide range of subjects, among which each reader will find something to marvel over or to argue with. I found his treatment of organic information and organic meaning in terms of nominable entities a revelation. As Barbieri explains in Chap. 2, nominable entities are a new class of physical observables, objective but not measurable. The simplest example is a linear sequence, which can be described only by naming the order and identity of its constituents. Chemical molecules are also nominable entities: one has to both list the constituent atoms and describe how they are bound to each other. What is crucial is that Barbieri shows that nominable entities, and therefore organic information and organic meaning, are new *fundamental* physical observables, the discovery of which has always led to major advances in physics.

I should prepare you for something, though. Unlike many scientists who write mostly in the disembodied passive and like to hedge their bets, Barbieri writes in the classic style, a style beautifully described by Francis-Noël Thomas and Mark

Turner in their *Clear and Simple as the Truth*. In *The Sense of Style* Steven Pinker writes: “The guiding metaphor of classic style is seeing the world. The writer can see something that the reader has not yet noticed, and he orients the reader’s gaze so that she can see it for herself. The purpose of writing is presentation, and its motive is disinterested truth. It succeeds when it aligns language with the truth, the proof of success being clarity and simplicity. The truth can be known, and is not the same as the language that reveals it; prose is a window onto the world. The writer knows the truth before putting it into words; he is not using the occasion of writing to sort out what he thinks. Nor does the writer of classic prose have to argue for the truth; he just needs to present it. That is because the reader is competent and can recognize the truth when she sees it, as long as she is given an unobstructed view. The writer and the reader are equals, and the process of directing the reader’s gaze takes the form of a conversation.” I can think of no better way of describing Barbieri’s prose. Enjoy it.

Wissenschaftskolleg zu Berlin, Berlin, Germany Jannie Hofmeyr
November 2014

on sabbatical leave from
the Department of Biochemistry
and the Centre for Studies in Complexity
of the University of Stellenbosch, South Africa

Introduction

The genetic code appeared on Earth at the origin of life, and the codes of culture arrived almost four billion years later. For a long time it has been assumed that these are the only codes that exist in Nature, and if that were true we would have to conclude that codes are *extraordinary exceptions* that appeared only at the beginning and at the end of the history of life. In reality, various other organic codes (codes between organic molecules) have been discovered in the past few decades.

In 1975, the American biochemist Gordon Tomkins published a paper entitled '*The Metabolic Code. Biological symbolism and the origin of intercellular communication*' (Tomkins 1975). That was the first announcement of a new organic code after the discovery of the genetic code, but tragically Tomkins died that very year and his new world of biological symbolism remained unexplored.

In 1979, David Elder pointed out that the formation of body segments in annelid worms is described by combinatorial rules that in electronics are known as *Gray code*, and proposed that the annelid body plan is based on a biological version of those rules that he referred to as *epigenetic code* (Elder 1979).

At the end of the 1980s, Edward Trifonov started a life-long campaign in favour of the idea that genomes simultaneously carry several overlapping codes, not just the genetic code, and gave them the collective name of *sequence codes* (Trifonov 1987, 1989, 1996, 1999).

Finally, at the end of the 1990s and in the early 2000s, a wide variety of new organic codes came to light. Among them: the *adhesive code* (Redies and Takeichi 1996; Shapiro and Colman 1999), the *splicing codes* (Barbieri 1998, 2003; Fu 2004; Matlin et al. 2005; Pertea et al. 2007; Wang and Burge 2008; Barash et al. 2010; Dhir et al. 2010), the *signal transduction codes* (Barbieri 1998, 2003), the *histone code* (Strahl and Allis 2000; Jenuwein and Allis 2001; Turner 2000, 2002, 2007; Kühn and Hofmeyr 2014), the *sugar code* (Gabius 2000, 2009), the *compartment codes* (Barbieri 2003), the *cytoskeleton codes* (Barbieri 2003; Gimona 2008), the *tubulin code* (Verhey and Gaertig 2007), the *nuclear signalling code* (Maraldi 2008), the *apoptosis code* (Basañez and Hardwick 2008; Füllgrabe et al. 2010), the *ubiquitin* *code* (Komander and Rape 2012), the *bioelectric code* (Tseng and Levin 2013; Levin 2014), the *glycomic code* (Buckeridge and De Souza 2014) and the *acoustic codes* (Farina and Pieretti 2014).

It must be underlined that codes have been defined in different ways, a problem that is not uncommon in biology, but in our case this is not an insurmountable obstacle because there is an operative definition that can be applied to all organic codes. This definition, furthermore, has been instrumental in the development of mathematical models that can estimate the presence of organic codes in natural systems (De Beule et al. 2011; Görlich et al. 2011; Görlich and Dittrich 2013).

An Operative Definition

An operative definition is one that allows us to make experimental tests which prove whether or not organic codes exist in Nature. The starting point is the idea that a code is always *a set of rules that establish a correspondence* (or a *mapping*) *between two independent worlds* (Barbieri 2003). The Morse code, for example, is a mapping between the letters of the alphabet and groups of dots and dashes. The highway code is a correspondence between street signals and driving behaviours (a red light means 'stop', a green light means 'go', and so on).

What is essential in all codes is that the coding rules, although completely compatible with the laws of physics and chemistry, are not dictated by these laws. In this sense they are *arbitrary*, and the number of arbitrary relationships between two independent worlds is potentially unlimited. In the Morse code, for example, any letter of the alphabet could be associated with countless combinations of dots and dashes, which means that a specific link between them can be realized only by selecting a small number of rules. And this is precisely what a code is*: a small set of arbitrary rules selected from a potentially unlimited number in order to ensure a specific correspondence between two independent worlds*.

This definition allows us to make experimental tests because organic codes are relationships between two worlds of organic molecules and are necessarily implemented by a third type of molecules, called *adaptors*, that build a bridge between them. The adaptors are required because there is no necessary link between the two worlds, and a fixed set of adaptors is required in order to guarantee the specificity of the correspondence. The adaptors, in short, are essential 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 code.

This gives us an *objective* criterion for discovering organic codes, and their existence in Nature is no longer a matter of speculation. It is, first and foremost, an experimental problem. More precisely, we can prove that an organic code exists, if we have three things: (1) two independents worlds of molecules, (2) a set of adaptors that create a mapping between them, and (3) the demonstration that the mapping is arbitrary because its rules can, at least in principle, be changed.

Two Outstanding Examples

The Genetic Code

In protein synthesis, a sequence of nucleotides is translated into a sequence of amino acids, and it has been shown that there is no necessary link between nucleotides and amino acids. These molecules belong to two independent worlds, and a bridge between them is realized by a third type of molecules, called *transfer-RNAs*, that act as adaptors and perform two distinct operations: at one site they recognize groups of three nucleotides, called *codons*, and at another site they receive amino acids from enzymes called *aminoacyl-tRNA-synthetases*. The key point is that any codon can, in principle, be associated with any amino acid. Hou and Schimmel (1988), for example, introduced two extra nucleotides in a tRNA and found that the resulting tRNA was recognized by a different synthetase and was therefore carrying a different amino acid. The number of possible connections between codons and amino acids, in other words, is potentially unlimited, and only the selection of a small fixed set of adaptors can ensure a specific mapping. This is *the genetic code*: a fixed set of rules of correspondence between codons and amino acids that are implemented by adaptors. In protein synthesis, in conclusion, we find all the three essential components of a code: (1) two independents worlds of molecules (nucleotides and amino acids), (2) a set of adaptors that create a mapping between them, and (3) the proof that the mapping is arbitrary because its rules can be changed.

The Signal Transduction Codes

Signal transduction is the process by which cells transform the signals from the environment, called *first messengers*, into internal signals, called *second messengers*. First and second messengers belong to two independent worlds because there are literally hundreds of first messengers (hormones, growth factors, neurotransmitters, etc.) but only four great families of second messengers (cyclic AMP, calcium ions, diacylglycerol and inositol trisphosphate) (Alberts et al. 2007). The crucial point is that the molecules that perform signal transduction are true adaptors. They consist of three subunits: a *receptor* for the first messengers, an *amplifier* for the second messengers, and a *mediator* in between (Berridge 1985). This allows the transduction complex to perform two independent recognition processes, one for the first messenger and the other for the second messenger. Laboratory experiments have proved that any first messenger can be associated with any second messenger, which means that there is a potentially unlimited number of arbitrary connections between them. In signal transduction, in short, we find all three essential components of a code: (1) two independents worlds of molecules (first messengers and second messengers), (2) a set of adaptors that create a mapping between them, and (3) the proof that the mapping is arbitrary because its rules can be experimentally changed.

Organic Codes and Macroevolution

In addition to the genetic code and the signal transduction codes, various other organic codes have come to light, and it is likely that more will be discovered in the future. The existence of many organic codes in Nature is therefore an *experimental* fact—let us never forget this—but also more than that. It is one of those facts that have extraordinary implications.

The data from molecular biology have revealed that all cells descend from three primary kingdoms, or domains, that Carl Woese (1987, 2000) called *Archaea*, *Bacteria* and *Eucarya*. In order to understand the evolution of the first cells we need to keep in mind 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 the 1.8 billion-year-old stromatolites that have been found in the fossil record (Schopf 1999; Knoll 2003). The most primitive bacteria, in other words, already had the main characteristics of their modern descendants, whereas other primordial cells went through profound evolutionary changes. This tells us that the first cells 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, including the ability to evolve new organic codes and have remained substantially the same ever since, whereas other cells conserved the potential to explore the coding space and have became increasingly complex.

The cells that did not evolve new organic codes became *bacteria* (*archaebacteria* and *eubacteria*) and have never fundamentally changed their cellular organization. The cells that evolved new codes, such as splicing codes, cytoskeleton codes, compartment codes, histone code and so on, became *eukarya* and have generated increasingly complex cellular structures.

We realize in this way that there is a close relationship between the appearance of new organic codes, and the appearance of genuine novelties in evolution, and we can easily understand why. The reason is that a new code brings into existence something that has never existed before because it creates arbitrary associations and generates relationships that are not determined by physical necessity.

A New Science of Life

It is an experiment fact that the organic codes have been highly conserved in evolution, and this has revealed a totally unexpected side of life. Before the origin of the genetic code, the common ancestor was engaged in evolving coding rules and was therefore a *code exploring system*. After the origin of the code, however, no other modification in coding rules was allowed and the cell became a *code conservation system*. Another part of the ancestral cells, however, maintained the potential to evolve the rules of different codes and behaved as new *code exploring,* or *code generating, systems*. In the early *Eukarya*, for example, the cells had *a code conservation part* for the genetic code, but also a *code exploring part* for the splicing code, and this tells us something important about life.

The origin of the first cells was based on the ability of the ancestral systems to *generate* the rules of the genetic code, and the subsequent evolution of the cells was based on two complementary processes: one was the *generation* of new organic codes and the other was the *conservation* of the existing ones. Taken together, these two processes are referred to as *codepoiesis* (Barbieri 2012), a phenomenon that 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 sole entities that have been perpetuated in evolution. They are the great invariants of life, the sole entities that have been conserved while everything else has changed.

Another outstanding implication of the existence of organic codes in Nature comes from the fact that any code involves *meaning* and we need therefore to introduce in biology, *with the standard methods of science*, not only the concept of biological information but also that of biological meaning.

The study on the organic codes, in conclusion, is bringing to light new mechanisms that operated in the history of life and new fundamental concepts. It is an entirely new field of research, the exploration of a vast and still largely unexplored dimension of the living world, the real new frontier of biology.

About This Book

The first part of the book (Chaps. 1, 2 and 3) deals with the paradigms of biology, and in particular with the controversy between the chemical paradigm (the idea that *life is chemistry*) and the information paradigm (the idea that *life is chemistry+information*). Here it is shown that the copying of the genes and the coding of proteins are equally fundamental processes, and this leads to a third theoretical framework that is referred to as the 'code paradigm' (the idea that *life is chemistry+information+codes*). It is shown, furthermore, that genetic information and coding rules are neither quantities nor qualities but a new type of fundamental observables that are referred to as *nominable* entities. Finally, the experimental basis of the code paradigm is described by illustrating the major organic codes that have been discovered so far.

The second part of the book (Chaps. 4, 5, 6, 7, and 8) is about the role that the organic codes have played in the history of life. It is shown that the genetic code was a precondition for the origin of the first cells, the signal transduction codes divided the first cells into three primary kingdoms (*Archaea*, *Bacteria* and *Eukarya*), the splicing codes were instrumental to the origin of the nucleus, the histone code provided a new regulation system in eukaryotic genomes, and the cytoskeleton codes allowed the *Eukarya* to perform internal movements, including those of mitosis and meiosis. It is shown, furthermore, that organic codes had a key role in the transitions to multicellular life, in particular in the origin of animals, the origin of mind and the origin of language. The great events of macroevolution, in other words, were associated with the origin of new organic codes, and this gives us not only a new description but also a completely new understanding of the history of life.

The third part of the book (Chaps. 9 and 10) is about the attempts to build a theoretical framework that is based on both biological information (genetic sequences) and biological meaning (coding rules). More precisely, Chap. 9 is dedicated to *Biosemiotics*, the project launched by Thomas Sebeok in the 1980s and 1990s where the semiotics of Peirce is taken as a new '*paradigm for biology*' and all biological concepts are reformulated in Peircean terms. Chapter 10 is dedicated instead to Code Biology, the research field that studies all codes of life with the standard methods of science and where we learn *from experiments*, not from *ad hoc* definitions, what the semiotic properties of Nature actually are.

Contents