# Symmetry Breaking in the Genetic Code

Michael Forger *

Departamento de Matemática Aplicada,
Instituto de Matemática e Estatística,
Universidade de São Paulo,
Cx. Postal 66281, 05389-970 São Paulo, SP

**Abstract**

We discuss a recent proposal of J.E.M. Hornos and Y.M.M. Hornos [1] that has found great repercussion in the international scientific literature [2, 3]. According to this proposal, the degeneracy of the universal genetic code for protein synthesis is not (as many molecular biologists used to and some continue to believe) purely accidental, but can be understood as resulting from an evolutionary process which involves symmetry breaking: evolution from a highly symmetric initial state to a final state in which this symmetry is strongly broken. This evolution must have occurred, in several consecutive steps, far back in earth's early history, and so is not accessible to direct observation. For the time being, the scheme proposed by Hornos and Hornos is purely group-theoretical, its main virtue being that the initial symmetry and all intermediate steps in the sequence can be *uniquely* reconstructed from presently available data.

The great challenge for the immediate future is to identify a dynamical system modelling the underlying evolutionary process, so that the sequence of symmetry breakings which has been found can be interpreted as associated with a sequence of (generic) bifurcations. I conjecture that, once again, it should be possible to determine such a dynamical system *uniquely*, within a given class, and to identify its biological foundations. This would undoubtedly lead to a major breakthrough in our understanding of the evolution and the origin of the genetic code and hence the origin of life on earth.

*Partially supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico)

# 1 The Evolution of Matter

To begin with, let me briefly quote a few fundamental questions over which, I am sure, all of us have already pondered at some point or other during our lives:

- What is the origin of the universe?
- What is the origin of life on earth?
- What is the origin of man?

These (and similar) questions, I guess, have been with us since the dawn of mankind, and so have been the innumerous attempts to answer them (at least partially): they can be found in all cultures and in all religions, where they often play an outstanding role, as evidenced, for example, in the Book of Genesis.

The notion of *evolution*, on the other hand, is certainly much more recent. In fact, all early cultures had a clear preference for a static view of the world that surrounds us, together with a strong tendency to attribute unexplained natural phenomena to some direct interference of God, or of gods and goddesses. This is also true for the ancient Greek culture – despite the fact that Greek philosophers such as Heracleitus (and Plato in his wake) were fully aware that "one never steps twice into the same river" and "everything flows": it seems that time was simply not ripe for taking that point of view to its ultimate consequences. This remained so for more than 2.000 years, until our occidental culture began to free itself from the normative force of religious beliefs and entered the "scientific age", whose beginning is marked by names such as Copernicus, Kepler, Galilei and Newton, setting out to analyze the world as it is and not as we believe it or wish it to be.

The turning point in our view of the world, from static to dynamic, is marked by the work of Charles Darwin on "The Origin of Species", which revolutionized biology by introducing the concept of evolution – a concept that has since then been carried over, with very little modifications, into practically all other areas of science. One prominent example is cosmology, where the transition from a static to a dynamic picture of the universe as a whole is due to Einstein's theory of general relativity and Hubble's subsequent discovery of the red-shift in the spectra of distant galaxies as evidence for the expansion of our universe.

Meanwhile, I think it is safe to say, the notion of evolution has turned into one of the most important and universal paradigms of modern science, appearing in practically every area in connection with possible changes in the schemes and patterns into which matter organizes itself. In fact, the universe as a whole has undergone, and continues

to undergo, a steady process of evolution, which comprises all forms of matter, at all levels. We may distinguish, for example, physical evolution, chemical evolution and biological evolution.

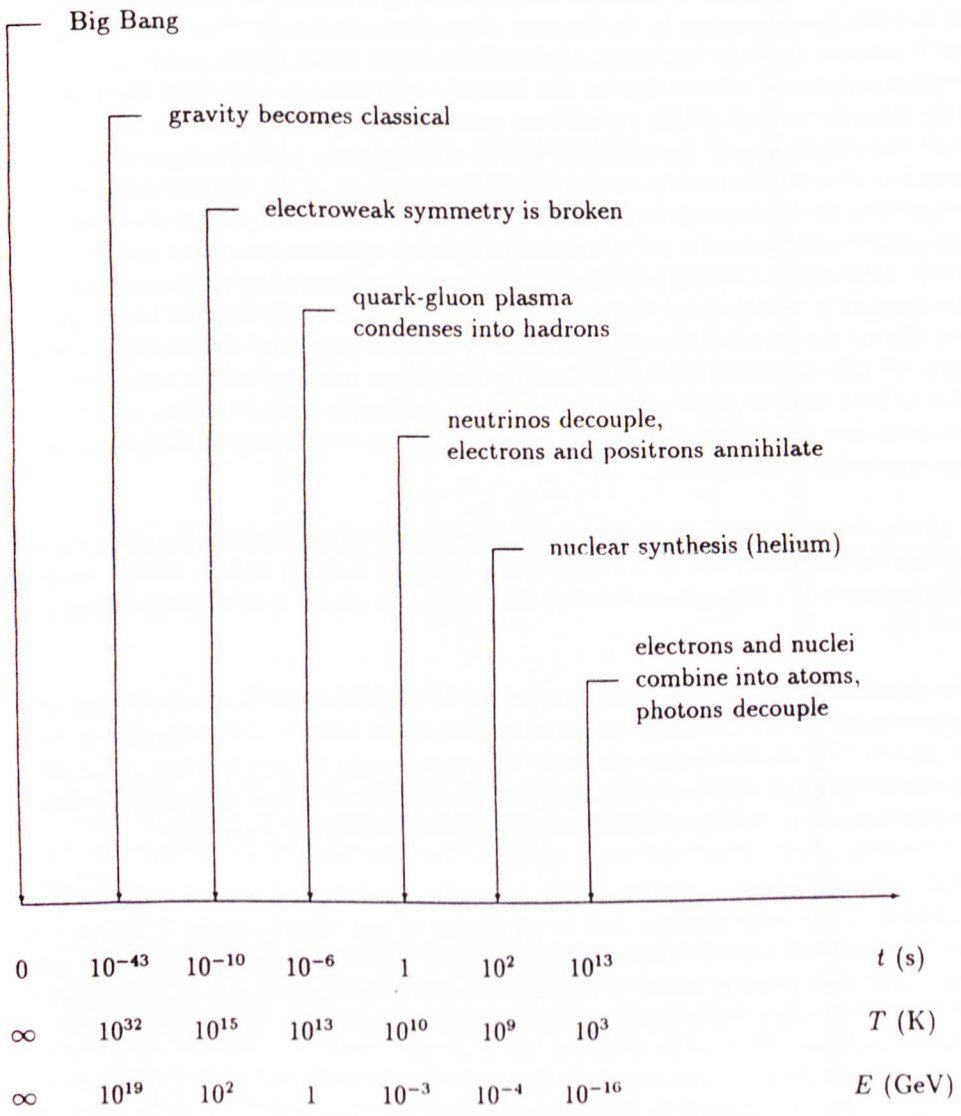



Figure 1: Physical evolution of the early universe

Let me begin with a few comments about the physical evolution of the universe: According to the standard (hot) big-bang model, this evolution begins with the big bang itself. In the very first phase (during the first $10^{-44}$ or $10^{-43}$ seconds), not only all matter, but even the gravitational field was governed by laws of quantum physics, about which practically nothing is known – except the fact that the usual concept of space-time as a smooth Lorentz manifold has no meaning and that we have no definite idea as to what should replace it. In the next phase (from about $10^{-44}$ or $10^{-43}$ seconds to $10^{-10}$ seconds after the big bang), the situation is not much better, since we have very little experimental information on the behavior of matter at such high energies. It is only towards the end of this period that we enter firm ground, since we may begin to trust the predictions of the standard model of elementary particle physics. The most prominent events afterwards are first of all the breaking of the electroweak symmetry, followed by the condensation of nuclear matter, from a state of a plasma made of quarks and gluons to a state of a gas composed of hadrons (protons, neutrons and other, less stable, strongly interacting particles). Next comes the decoupling of the neutrinos from the remaining particles and the annihilation of electrons and positrons into radiation, followed by the primordial nuclear synthesis, which has produced almost exclusively helium: all this happened during the famous "first three minutes" of the universe. Much later (about 680.000 years after the big bang), we finally record the recombination of electrons and nuclei into atoms and, as a result, the decoupling of the photons: the universe became transparent.

A schematic representation of this evolution is given in Fig. 1, where the time scale (in seconds) is supplemented by a temperature scale (in Kelvin) and an energy scale (in gigaelectronvolts). For a more detailed discussion, the reader is referred to the popular book [4].

The chemical evolution of the universe starts with nucleosynthesis in stars, providing the raw material for the organization of matter at the atomic and molecular level. In our galaxy, this process began about 10 billion years ago (this is the age of the oldest stars in the galaxy), and is continuing up to this day. Other stages of chemical evolution are the formation of molecules and radicals of ever increasing complexity.

Finally, the biological evolution of life on earth must have begun shortly after the formation of the solar system, and in particular of our planet, about 5 billion years ago. The first cell seems to have appeared about 4 billion years ago, since the oldest fossils that we presently know of (the cyanobacteria recently found in Australia) are 3,8 billion years old, while evidence for the first animals goes back to only about 1,5 billion years ago.

A schematic representation of this evolution is given in Fig. 2.

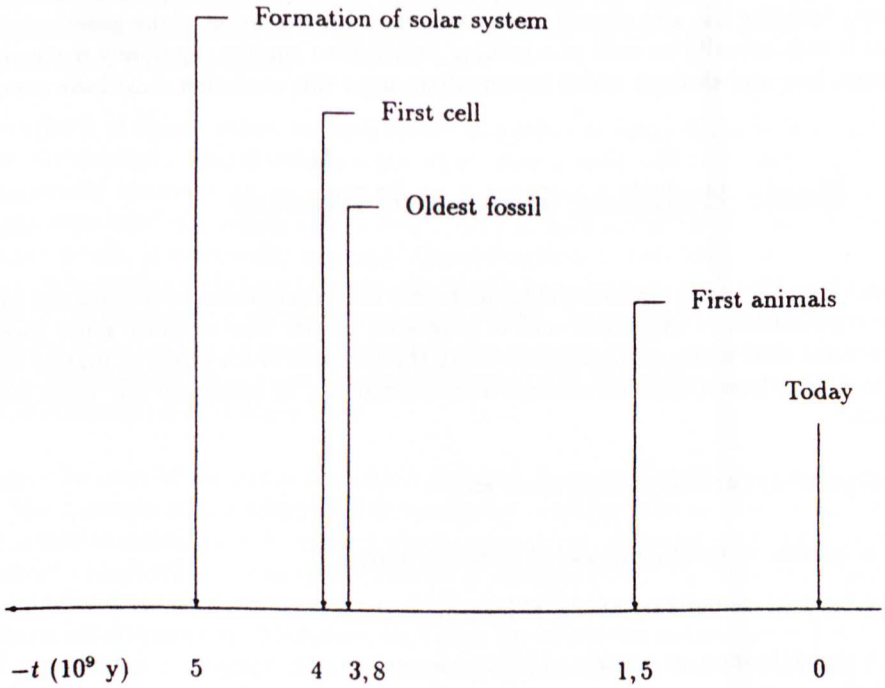



Figure 2: Early biological evolution on earth

In this scheme, the subject of interest here, namely the evolution of the genetic code for protein synthesis, must have occurred during the earliest phase of the formation of life on earth, somewhere between 5 and 3 billion years ago (more exact dates are not available).

An important aspect of many evolutionary processes, and often one of their most prominent features, is the phenomenon of symmetry breaking: it occurs when an initial state of high symmetry evolves to a later state of lower symmetry. One typical example in cosmology, already mentioned above, is the breaking of the electroweak symmetry, that is, the symmetry between the electromagnetic forces and the weak nuclear forces: this is one of the salient features of the standard model of elementary particle physics, also known as the Glashow-Salam-Weinberg model. Another example from the same area is the process of formation of galaxies (or galaxy clusters) from the "primordial soup" (a hot gas composed mainly of photons, electrons, protons, helium nuclei and

neutrinos): during this process, spatial homogeneity of the universe was lost, i.e., the translational symmetry present in the previous stage of evolution was broken.

The main point of the work to be reported here is that the same phenomenon of symmetry breaking has also played a decisive role in the evolution of the genetic code and that it may actually be used as a guiding principle to analyze, by purely mathematical means, how and through which intermediate steps this evolution must have occurred.

## 2 Basic Building Blocks of Matter

Apart from the many obvious and fundamental differences between physics and biology, the organization of inanimate and of animated matter also exhibits some surprising structural similarities, the main one being the fact that both forms of matter fall into several big classes, each of which is constructed out of relatively few basic building blocks.

In elementary particle physics, these are

- quarks, as building blocks of hadronic matter,
- leptons, as building blocks of leptonic matter,
- gauge bosons, as carriers of interactions.

In biology, we encounter

- sugars, as energy sources,
- lipid acids, as building blocks of membranes,
- amino acids, as building blocks of proteins,
- nucleic acids, as building blocks of the information carriers DNA or RNA.

It is this structural analogy which lends support to the idea that basic ideas which have proved fruitful in one area may very well be guiding principles for the other one. We have already commented on the notion of evolution, which was transferred from biology to physics and led to great progress in the understanding of complex open systems. Conversely, it is to be expected that the ideas of symmetry and symmetry breaking, which have been so enormously successful in physics, will prove to be useful in biology as well.

To avoid misunderstandings, it should perhaps been pointed out that when I speak of symmetries, one should not necessarily think of standard "spatial" symmetries, realized through translations and/or rotations and/or reflections in the usual 3-dimensional "physical" space, but should admit the option of considering "internal" symmetries, realized through transformations in an abstract "internal" space that has nothing to do with the usual physical space.

In physics, both of these options are well known and occur in many different instances. Consider, for example, spatial rotations (in which case a body will be "symmetric" if it is rotationally invariant or, in other words, is spherically symmetric), as opposed to "charge rotations" (in which case a body will be "symmetric" if it is uncharged or, in other words, is electrically neutral). Generalizations of this idea (based on the concept of "converting global symmetries into local symmetries by gauging") form the backbone of the most fundamental modern physical theories on the structure of matter and of space-time, leading to general relativity (by gauging spatial or, more precisely, space-time symmetries) on the one hand and to Yang-Mills gauge theories (by gauging internal symmetries) on the other hand.

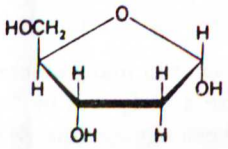
In biology, the state of the art is somewhat different, because biologists normally encounter the concepts of symmetry and of symmetry breaking only in disguise and in just one single situation, namely through the famous *chiral symmetry* under which one distinguishes non-chiral molecules (as "symmetric" bodies) from – left-handed or right-handed – chiral molecules (as "non-symmetric" bodies): this is a prominent example of a global spatial symmetry. Therefore, biologists are in general unfamiliar with other kinds of symmetries – internal rather than spatial, or local rather than global. But the symmetry that appears in the evolution of the genetic code is an internal one. This implies, among other things, that it is not a symmetry of the constituents of the genetic code (nucleic acids) in any physical sense, but rather is to be viewed as an abstract symmetry in the scheme according to which these constituents are organized in the genetic code. To my knowlegdge, this is the first time that an internal symmetry appears in biology, and the difficulties encountered among biologists to understand and appreciate this idea should therefore come as no surprise.

## 3 DNA, RNA, Protein Synthesis and the Genetic Code

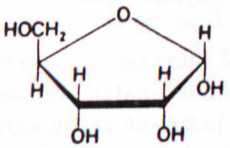
In this section, I shall briefly recall the basic structure of DNA and RNA and the way in which they code the synthesis of proteins.

In all forms of life on earth, the genetic information is stored in two polymers called DNA (deoxyribonucleic acid) and RNA (ribonucleic acid), which are made of

- sugar (desoxyribose and ribose, respectively),



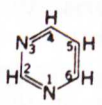
2-Deoxy-α-D-ribofuranose

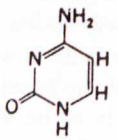


α-D-Ribofuranose

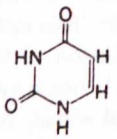
- phosphate,
- four different nucleic bases:
  A (adenine), C (cytosine), G (guanine) and
  T (thymine) in DNA, U (uracil) in RNA.

Chemically, cytosine, thymine and uracil are pyrimidines, while adenine and guanine are purines.

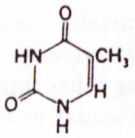

Pyrimidine



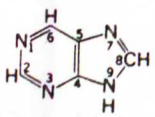
Cytosine
(2-Oxy-4-aminopyrimidine)

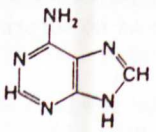


Uracil
(2,4-Dioxopyrimidine)

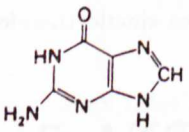


Thymine
(5-Methyl-2,4-dioxopyrimidine)


Purine



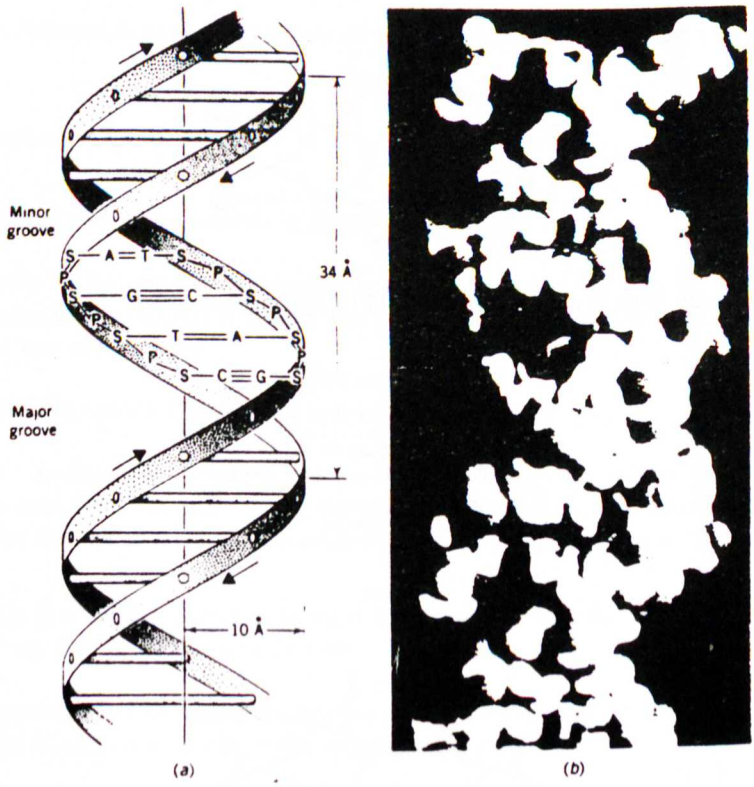
Adenine
(6-Aminopurine)



Guanine
(2-Amino-6-oxopurine)

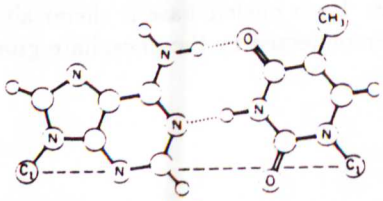
## 3.1 DNA

DNA – the primary genetic material – forms the famous *double helix*, consisting of two strands of nucleic bases. Each nucleic base is chemically bound to a deoxyribose molecule, and these are interconnected by the phosphate groups to form a strand.

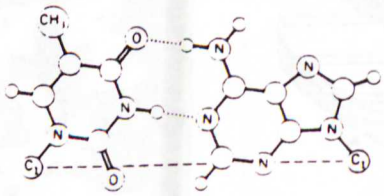


The sequence of nucleic bases in any one of the two strands completely determines the other one because

- A pairs only with T and T only with A (2 hydrogen bonds):

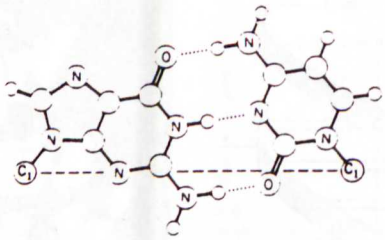


A : T

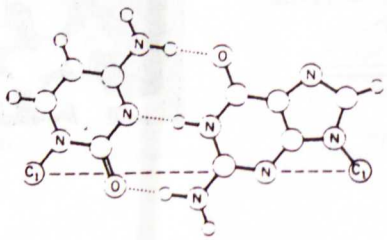


T : A

- G pairs only with C and C only with G (3 hydrogen bonds):



G : C



C : G

Note that mathematically, this prescription can be viewed as a kind of dual pairing.

## 3.2 RNA

RNA – the secondary genetic material – consists of only one strand of nucleic bases, each chemically bound to a ribose molecule, and these are interconnected by the phosphate groups. There are various types of RNA:

- mRNA – messenger or matrix RNA,
- tRNA – transfer RNA,
- rRNA – ribosomal RNA.

These have different functions, but they all play an important role in protein synthesis.

## 3.3 Protein Synthesis

Protein synthesis in cells proceeds in basically two stages:

*Transcription:* The genetic information is copied from the DNA to the mRNA: the basis sequence of the mRNA is simply the mirror image of that part of that strand of the DNA it was copied from.

The mRNA then carries this information to the ribosomes.

*Translation*: In the ribosomes (the protein synthesis factories inside the cell), the basis sequence is read from the mRNA and translated into a sequence of amino acids which will form the desired protein. The essential mediator of this process is the tRNA.

Thus we face the basic question as to what is the genetic code for protein synthesis (or as we shall say for short, the genetic code).

The basic answers to this question have been known for almost 30 years, as a result of genetic experiments carried out in the late 50's and early 60's.

- One unit of information of the code, commonly called a *codon*, is a sequence of 3 bases, each representing uniquely one amino acid.
- The code is *non-overlapping*, that is, no nucleic acid can belong to two different codons simultaneously, and is *gap-free*, that is, every nucleic acid in a sector of a chromosome which codes for protein synthesis (each such sector is marked by a starting signal and a termination signal) belongs to some codon.

- The code is *degenerate*, that is, different codons may represent one and the same amino acid.

Indeed, since there are 4 bases, we have $4 \times 4 \times 4 = 64$ possible codons, but there are only 20 amino acids which play an active part in biological processes, plus the termination signal. (The codon for the starting signal is identical with the codon for methionine, one of the 20 amino acids.)

The complete correspondence between codons (at the level of the mRNA) and amino acids is shown in Table 1:

| first base | second base U | C | A | G | third base |
|---|---|---|---|---|---|
| U | Phe | Ser | Tyr | Cys | U |
| | Phe | Ser | Tyr | Cys | C |
| | Leu | Ser | TERM | TERM | A |
| | Leu | Ser | TERM | Try | G |
| C | Leu | Pro | His | Arg | U |
| | Leu | Pro | His | Arg | C |
| | Leu | Pro | Gln | Arg | A |
| | Leu | Pro | Gln | Arg | G |
| A | Ile | Thr | Asn | Ser | U |
| | Ile | Thr | Asn | Ser | C |
| | Ile | Thr | Lys | Arg | A |
| | Met | Thr | Lys | Arg | G |
| G | Val | Ala | Asp | Gly | U |
| | Val | Ala | Asp | Gly | C |
| | Val | Ala | Glu | Gly | A |
| | Val | Ala | Glu | Gly | G |

Table 1: The genetic code of mRNA (fully deciphered by June 1966)

This table raises a number of interesting questions:

- What are the relations, if any, between the physico-chemical properties of the amino acids and the respective codons?
- Did the code have its present form since the very beginning, or is it the result of an evolutionary process, and if so, what form did it have in intermediate stages?
- Did the first forms of life already use all 20 amino acids, or did they use less, and if so, which?
- What is the role of chirality – that is, of chiral symmetry and chiral symmetry breaking – of the amino acids (and more generally, of most biologically active molecules)?
- Does the code itself exhibit some symmetry, which was broken during the evolutionary process alluded to before?

It is this last question that we shall address next.

To this end, we have to say a few words about symmetry and symmetry breaking.

# 4 Symmetry and Symmetry Breaking

Symmetry principles are a part of human culture since the very beginning of mankind, pervading arts and sciences.

> Examples are abundant in architecture (Egyptian or Mayan pyramids, Greek temples, Arab ornaments), in painting (Escher) or music (Bach's fugues).
>
> In fact, it seems to be a general feature of human nature that an intuitive notion of beauty of an object (or even a person) is associated with some inherent symmetry.

Symmetries can be *exact* or *broken*.

> One speaks of a broken symmetry when a symmetry is only approximate, that is, the deviation from the exact symmetry is sufficiently small for it to remain clearly perceptible.

One example of a broken symmetry which has already been mentioned before and which is equally important in all natural sciences – physics, chemistry and biology – is *chiral* symmetry. But there are other examples.

Another important observation is that symmetry breaking does not necessarily occur in one stroke, but often in a sequence of steps.

The mathematical theory of symmetries is *group theory*, initiated by *Galois* (discrete groups) and *Lie* (continuous groups).

> This still holds, even though some modern variants such as supersymmetry or quantum group symmetries no longer involve groups strictu sensu.

Let us introduce some group-theoretical terminology:

- Exact symmetry:

  An exact symmetry is described abstractly by a *group* $G$ and realized concretely by a set of matrices which, taken together, form a *representation* of this group on some finite-dimensional vector space $V$.

  To simplify, we shall assume in the following that this representation is *irreducible*: this means, in the terminology used in many areas of science other than mathematics (such as physics or chemistry) that the vectors of some (arbitrarily chosen) basis of $V$ form a single *multiplet* under $G$. (Otherwise, the representation must be split into its irreducible constituents, which leads to an entire set of multiplets.)

- Broken symmetry:

  A broken symmetry is described by fixing, in addition, a subgroup $H$ of $G$ representing the "residual symmetry", i.e., that part of the symmetry which has remained intact during the breaking. Then the irreducible representation of $G$ on $V$ breaks into several irreducible subrepresentations of $H$ on subspaces of $V$, that is, the multiplet under $G$ breaks up into several multiplets under $H$.

  More generally, $G$ comes with a sequence of subgroups $G_1, \ldots, G_k$ which form a chain

  $$G \supset G_1 \supset \ldots \supset G_k \ ,$$

  leading to a whole sequence of such splittings, where at each step an irreducible representation of the previous group in the chain breaks into several irreducible subrepresentations of the next group in the chain.

Briefly, we may say that symmetries are associated with *degeneracies*, while symmetry breaking leads to the *lifting of degeneracies*.

Next, we inquire about the converse problem, which is the following. Given just a set of multiplets, find a group $G$ and a chain of subgroups

$$G \supset G_1 \supset \ldots \supset G_k$$

such that the given set of multiplets can be arranged into an irreducible representation of that group and be reproduced by reduction through that chain of subgroups.

Examples of this "spectroscopic approach" towards the identification of symmetries are abundant throughout physics. In fact, in some areas such as quantum theory, whose objects are not directly observable to us, this may be the only available method to experimentally verify the presence of symmetries: through their spectroscopic consequences.

> Just to mention a single example from atomic physics, consider the degeneracy of energy levels in a central potential due to rotational symmetry and its lifting by application of an external field, which results in a splitting of spectral lines, such as in the Stark effect and the Zeeman effect.

In the following, we shall apply the same strategy to analyze the degeneracy of the genetic code.

# 5 Broken Symmetry in the Genetic Code

The framework in which the problem has been investigated so far is that of simple compact Lie groups.

These groups have been completely classified by *Cartan* in 1914; they fall into four series (the so-called classical groups) and five exceptional groups; see Table 2 below.

Note that any compact Lie group is locally isomorphic to a product of simple compact Lie groups and the circle group $U(1)$, so the restriction to simple groups is, in this context, merely a matter of convenience.

An alternative would be to study the problem using simple finite groups: this has not been carried out yet.

Another alternative would be to use quantum group symmetries: this is also still a completely untilled field.

| Cartan Label | Complex Group | Compact Group | Dimension |
|---|---|---|---|
| $A_r$ | $SL(r+1,\mathbf{C})$ | $SU(r+1)$ | $r(r+2)$ |
| $B_r$ | $SO(2r+1,\mathbf{C})$ | $SO(2r+1)$ | $r(2r+1)$ |
| $C_r$ | $Sp(2r,\mathbf{C})$ | $Sp(2r)$ | $r(2r+1)$ |
| $D_r$ | $SO(2r,\mathbf{C})$ | $SO(2r)$ | $r(2r-1)$ |
| $E_6$ | | | 78 |
| $E_7$ | | | 133 |
| $E_8$ | | | 248 |
| $F_4$ | | | 52 |
| $G_2$ | | | 14 |

Table 2: Cartan's list of simple Lie groups

The strategy for analyzing the degeneracy of the genetic code is now the following.

1. Look for all simple compact Lie groups which have a 64-dimensional irreducible representation.
2. For each of the possibilities obtained, analyze all chains of closed subgroups and try to find at least one – even better, precisely one – chain of groups which reproduces the degeneracy of the genetic code, that is, which yields
   - 3 sextuplets,
   - 5 quadruplets,
   - 2 triplets,
   - 9 doublets,
   - 2 singlets.

In the first step, we must find all simple compact Lie groups which have a 64-dimensional irreducible representation. In Table 3, we list all such groups up to rank 8, together with the highest weight of that representation, according to the tables of McKay and Patera [5].

| Cartan Label | Compact Group | Highest Weight |
|---|---|---|
| $A_1$ | $SU(2)$ | 63 |
| $A_2$ | $SU(3)$ | (3,3) |
| $C_2$ | $Sp(4)$ | (3,1) |
| $G_2$ | | (1,1) |
| $A_3$ | $SU(4)$ | (1,1,1) |
| $C_3$ | $Sp(6)$ | (1,1,0) |
| $B_6$ | $SO(13)$ | (0,0,0,0,0,1) |
| $D_7$ | $SO(14)$ | (0,0,0,0,0,0,1) |

Table 3: Simple compact Lie groups with a 64-dimensional irreducible representation

In the second step, we have to investigate all chains of subgroups of all these groups to see whether they correctly reproduce the degeneracy of the genetic code. Generically, each subgroup of the chain will be a maximal subgroup of the previous one, so to begin with, we need to know all maximal connected closed subgroups of compact simple Lie groups – a problem completely solved by *Dynkin*. In Table 4, we list all such subgroups of the groups from Table 3, again according to the tables of McKay and Patera [5].

Next, we have to use *branching rules* for irreducible representations of a group into irreducible representations of a subgroup (these branching rules, for groups up to rank 8, form the main body of the tables of McKay and Patera) to decide which of the many possibilities agrees best with the experimental findings.

Here are two cases which can be easily ruled out.

- $G = SU(2)$ with the irreducible representation of highest weight 63 (i.e., of spin 63/2). After the first breaking to $U(1)$, the degeneracy is completely lifted: we obtain 64 singlets.
- $G = SO(13)$ or $G = SO(14)$ with the irreducible representation of highest weight $(0,0,0,0,0,1)$ or $(0,0,0,0,0,0,1)$, respectively. As in the previous example, the degeneracy is, in all chains, removed very abruptly, leading to much more amino acids than we actually find in nature.

| Compact Group | Maximal Closed Subgroups |
|---|---|
| $SU(2)$ | $U(1)$ |
| $SU(3)$ | $SU(2)$, $SU(2)$ |
| $Sp(4)$ | $SU(2) \times SU(2)$, $SU(2)$ |
| $G_2$ | $SU(3)$, $SU(2) \times SU(2)$, $SU(2)$ |
| $SU(4)$ | $SU(3)$, $Sp(4)$, $SU(2) \times SU(2)$ |
| $Sp(6)$ | $Sp(4) \times SU(2)$, $SU(3)$, $SU(2) \times SU(2)$, $SU(2)$ |
| $SO(13)$ | $SO(12)$, $SU(4) \times SO(7)$, $Sp(4) \times SO(8)$, $SU(2) \times SO(10)$, $SU(2) \times SU(2) \times SO(9)$, $SU(2)$ |
| $SO(14)$ | $SU(4) \times SO(8)$, $SU(2) \times SU(2) \times SO(10)$, $SU(7)$, $SO(13)$, $SU(2) \times SO(11)$, $SO(7) \times SO(7)$, $Sp(4) \times SO(9)$, $Sp(6)$, $Sp(4)$, $G_2$ |

Table 4: Maximal closed subgroups of the simple compact Lie groups of Table 3

A careful analysis of all possible chains has been performed by Hornos and Hornos [1], with the following result.

1. There is no solution that fits perfectly.

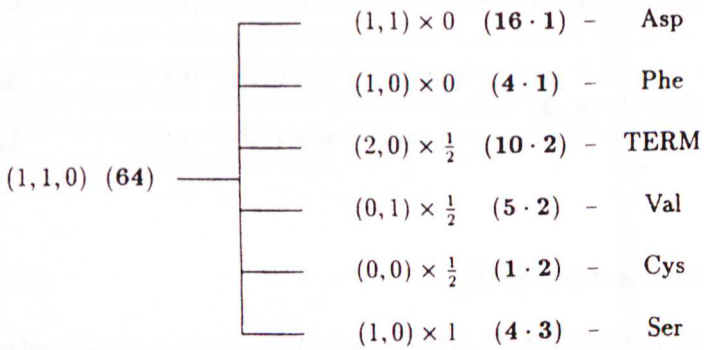
2. There is however precisely one solution that comes very close to reproducing the right number of multiplets with the right multiplicities; it is based on the symplectic group of rank 3 and on the following chain of subgroups:

$$\begin{aligned} Sp(6) &\supset Sp(4) \times SU(2) && \text{I} \\ &\supset SU(2) \times SU(2) \times SU(2) && \text{II} \\ &\supset SU(2) \times U(1) \times SU(2) && \text{III} \\ &\supset SU(2) \times U(1) \times U(1) && \text{IV/V} \end{aligned}$$

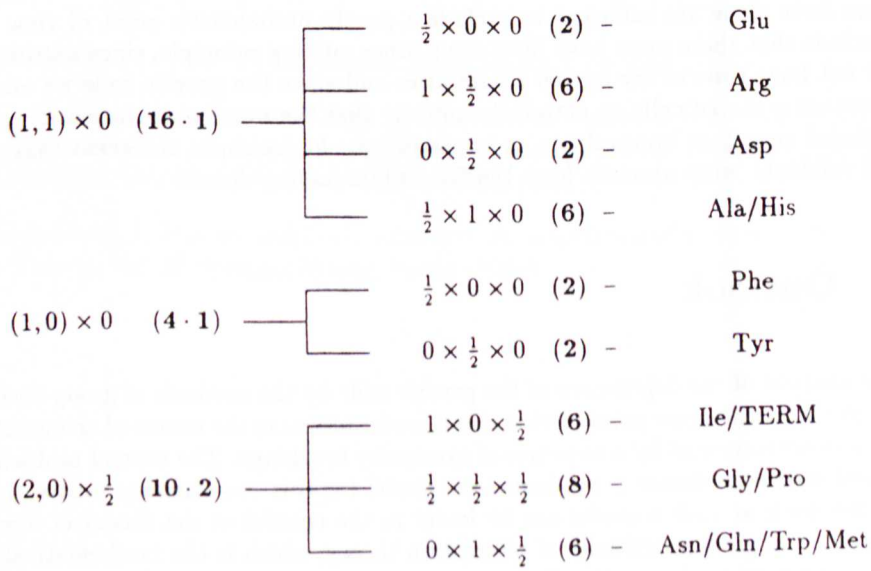
3. The fit is not perfect in that the process of symmetry breaking, or lifting of degeneracies, was "accidentally" interrupted, or "frozen", in the last step (IV $\rightsquigarrow$ V). But exactly this was foreseen by *Crick*, using completely different arguments.

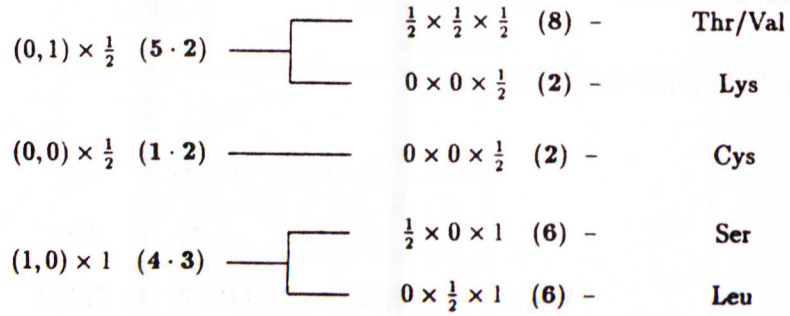
We briefly describe the first two steps, together with the (tentative) assignments of (primordial) amino acids:

- Step I: $Sp(6) \rightsquigarrow Sp(4) \times SU(2)$




- Step II: $Sp(4) \times SU(2) \rightsquigarrow SU(2) \times SU(2) \times SU(2)$






For the remaining steps, see [1].

To conclude, I should like to briefly comment on a point of view adopted by many molecular biologists, namely the hypothesis that the evolution of the genetic code was not guided by any principle, but was purely accidental, or in other words, that the genetic code was built by a simple trial and error process. However, estimates on how many genetic codes could be constructed in this way lead to a number of the order of $10^{71}$! Moreover, almost all of them would contain exclusively singlets. In my view, these facts alone are sufficient to exclude a purely probabilistic point of view and to conclude that there must have been some other guiding principle, since nature simply did not have time to try out all possibilities and since the genetic code we encounter in nature is statistically so extremely unlikely that the event of its formation must be regarded simply as impossible – as impossible as, for example, the event that a stone will suddenly jump upwards from the floor while cooling down.

## 6 Outlook

The analysis of the degeneracy of the genetic code by the methods of group theory has led to a surprising new point of view about its formation in the course of an evolutionary process accompanied by a sequence of symmetry breakings. The central problem to be solved now is to devise a mathematical model for this evolutionary process itself. I conjecture that such a model can be found in the context of the theory of dynamical systems and more specifically of bifurcation theory, which is the mathematical theory appropriate to analyzing the abrupt structural changes in the behavior of dynamical systems resulting from the continuous variation of external parameters. In fact, it is well known that in a dynamical system with symmetry, a bifurcation is always accompanied

by a (partial) breakdown of that symmetry; a classical example from fluid dynamics is the famous Bénard problem. See [6] for an extensive discussion.

The problem of constructing a realistic dynamical system for the evolution of the genetic code, exhibiting an explicit $Sp(6)$-symmetry and such that the sequence of symmetry breakings described above can be viewed as the unavoidable result of a sequence of (generic) bifurcations, is presently under investigation. One of the criteria to be observed is certainly that such a dynamical system should be constructed using the minimum possible number of degrees of freedom and, still more important, the minimum possible number of external parameters. Once this is achieved, there will appear an interesting task for molecular biologists, namely to provide a biological interpretation of these degrees of freedom and of these parameters.

## References

[1] J.E.M. Hornos and Y.M.M. Hornos: *Algebraic Model for the Evolution of the Genetic Code*, Phys. Rev. Lett. **71** (1993) 4401-4404.

[2] J. Maddox: *The Genetic Code by Numbers*, Nature **367** (1994) 111.

[3] I. Stewart: *Broken Symmetry in the Genetic Code?*, New Scientist (5 March 1994) 16.

[4] S. Weinberg: *The First Three Minutes*, Basic Books, New York (1977).

[5] W.G. McKay and J. Patera: *Tables of Dimensions, Indices and Branching Rules for Representations of Simple Lie Algebras*, Lecture Notes in Pure and Applied Mathematics, Vol. 69, Marcel Dekker, New York (1981).

[6] M. Golubitsky, I. Stewart and D.G. Schaeffer: *Singularities and Groups in Bifurcation Theory, Vol. II*, Springer-Verlag, Berlin (1988).