

# OPTIMIZATION AND THE GENETIC CODE

A. FIGUREAU

*Institut de Physique Nucléaire (et IN2P3). Université Lyon-I  
43, boulevard du 11 novembre 1918 – 69622 Villeurbanne Cedex, France*

(Received 14 October, 1986)

**Abstract.** The present paper will focus on the relation between the structure of the table of the genetic code and the evolution of primitive organisms: it will be shown that the organization of the code table according to an optimization principle based on the notion of resistance to errors can provide a criterium for selection. The ordered aspect of the genetic code table makes this result a plausible starting point for studies of the origin and evolution of the genetic code: these could include, besides a more refined optimization principle at the logical level, some effects more directly related to the physico-chemical context, and the construction of realistic models incorporating both aspects.

## Introduction

Related amino-acids correspond to related codons. This property was noted very early in the study of the genetic code, as soon in fact as the pioneering experiments determined the codon catalogue (Nirenberg *et al.*, 1965; Söll *et al.*, 1965). It is most apparent in the grouping of synonymous triplets, and can furthermore be seen in the relative positions of these groups in the table of the code (Sonneborn, 1965; Woese, 1965; Volkenstein, 1965; Pelc, 1965; Fitch, 1966). There exists now a variety of approaches which analyze correlations between coded amino-acids and codons (or anticodons); some are based on physico-chemical properties (Goldberg and Wittes, 1966; Epstein, 1966; Alff-Steinberger, 1969; Jungck, 1978; Wolfenden *et al.*, 1979), others are deduced from molecular models (Dunnill, 1966; Lacey and Pruitt, 1969; Tze-Fei-Wong, 1975; Jukes, 1982; Root-Bernstein, 1984), coding theory (MacKay, 1967; Swanson, 1984) or group theory (Danckwerts and Neubert, 1975; Karasev, 1976; Bertman and Jungck, 1979, Findley and McGlynn, 1979).

Although definite correlations have been exhibited and studied in some detail, it has not been possible to decide whether the assignment of each amino acid to subsets of triplets was due to some stereochemical fit between these molecules, in some earlier stage of life evolution, or if it was the result of a selection, in which the number of mutations, or more generally the severity of errors in the protein-building process, would have been determinant. This question is of great importance for the understanding of the origin and evolution of the genetic code, and has been the subject of much speculation. It is largely unsolved because our knowledge of such an ancient and complex event is still very unprecise. In our view, however, the ideas of a physico-chemical determination and of a selection against errors can be compatible: if one admits that some interaction exists which could exhibit a preference for certain amino acid – triplet pairs in the earliest

stages of coding, it must be quite weak, since no such bonding preference has been detected directly. Then it is likely that the genetic code was not determined on a purely chemical basis; on the contrary a slight preference might have been amplified through a selection mechanism such as the one proposed in this article: this possibility will not be discussed in this very primitive attempt, but it should be considered in a next step, if the purpose of our model is to be met.

In the present work, we construct a simple model in which we assume that evolution proceeded through errors even in the most primitive stages of the emergence of life, and we study a selection mechanism that is based on differences in the genetic codes of various organisms. We will show quite generally that certain types of coding constitute an advantage as compared to others, in the sense that the population of organisms with different genetic codes will grow differently with time.

The central point is the following: any given set of mutations would not lead to the same errors of translation, a priori, in two organisms possessing different codon catalogues. A consequence is that some (theoretical) codes would lead to a better fidelity of translation of the genetic message than others. (The argument was first put forward by T.M. Sonneborn (1965) in the case of simple configurations; a more systematic demonstration is included in a forthcoming paper (Figureau and Pouzet, in preparation). The next step, which will be developed in this work, relates this fidelity to the aptitude of the organism to reproduce, so that a population with a 'better' code will overwhelm those with a less fit codon catalogue; our model provides thus a description of the selection mechanism through which the genetic code may have passed.

Our goal here is to focus on the effects specifically related to the 'geometry' of the code table, leaving deliberately aside all other factors which might govern the efficiency of reproduction at some other level. In particular, we do not attempt to include into our description the meaning of the genetic message in terms of proteins. This notion of geometry should in principle take into account all informations, even of a continuous nature, which might be a measure of similarity between two amino-acids or two triplets. So far, however, only a rudimentary characterization of what is a good code table has been considered: it consists in studying partitions of the 64 triplets into 2 sets, one of which possesses a given property  $P$  (while the other does not), and then in defining as better those which are more stable when mutations are allowed. The geometry is here reduced simply to a measure of compactness of the two sets of triplets; for illustration, Figure (1) describes this maximization criterium for the example that will be discussed in the next sections. All possible optimal sets have been studied (Figureau and Pouzet, 1984 and in preparation), and many of them can be found in the genetic code (Figureau and Labouygues, 1981; Labouygues and Figureau, 1984).

Various generalizations of this optimization procedure are possible: firstly, our assumptions (to be detailed below) can be modified to a more elaborate form (Luo and Li, 1986); secondly, various properties  $P$  can be considered simultaneously, among which a hierarchy can be defined by a more general optimization criterium (Figureau, 1986); finally, notions such as accuracy, speed of translation, or energy requirements, could be introduced as well into our model: more work is needed in these directions. For the time

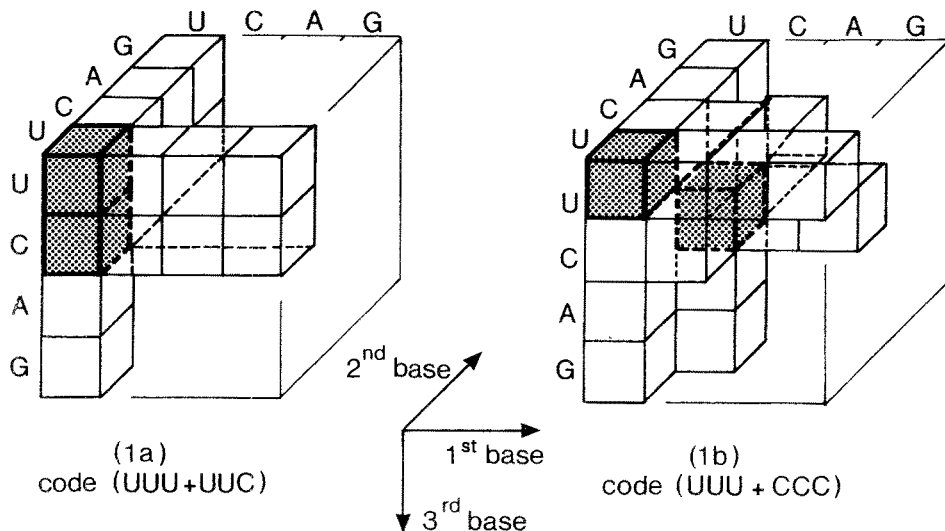


Fig. 1. Two examples of simplified genetic codes. The representation is 3-dimensional to take better account of the triplet nature of codons. In both cases, the code consists in a partition of the 64 triplets into 2 sets: 2 triplets (heavy lines: UUU and UUC in (1a), UUU and CCC in (1b)) correspond to a given property  $P$ , the 62 others do not. By definition, a code is better than the other if  $P$  is better conserved when mutations occur. The figures exhibit in both cases all point mutants of the 2 codons associated with  $P$ . This gives a visual idea of the resistance of the codes, measured by the number of mutants which still belong to the code (2 in (1a), 0 in (1b)).

being, however, we will restrict our argument to the most elementary conceivable situation, as an illustration of the mechanism which we intend to exhibit.

## 2. Assumptions

Let us be more specific: we will consider the possibility that two different codon catalogues have been able to sustain the emergence of life in the form of some primitive organisms. Here we do not try to analyze the occurrence of this situation in an evolutionary context, but simply choose it as our initial condition. We will further assume that the difference between their codon catalogues is their only difference: in particular, the reproduction ability of both organisms (characterized by a common population growth parameter  $A$ ) will be the same in the absence of mutations, whatever the structure of their codes, the requirements of their metabolism and the conditions in which they develop. As unrealistic as it may appear, this assumption could nevertheless be justified if the two forms of life are sufficiently similar, which is most likely. In any case, it symbolizes the separation which we intend to stress here between the informational aspect of the catalogue structure and all other characteristics of a living organism (this separation is clear in this model, but might be quite fuzzy if the optimization criterium included variables related to functional aspects of the organisms).

For our purposes, it will be sufficient to consider ultrasimplified genetic messages, one

triplet long, based on 2-word vocabularies; as a concrete example, we will compare the code in which UUU and UUC have a translation (at this level, they look like synonyms, whereas the 62 other triplets are considered mis-sense) to that consisting in UUU and CCC (Figure 1). How such a simple example can exhibit some realistic feature will be discussed later.

Now we can introduce mutations in our model: we will treat only the case of point mutations, which are the most probable if the error rate is not too large (simultaneous errors on 2 or 3 letters in a triplet could be included in our treatment with no conceptual difficulty). Mistakes of a different type, such as deletions or insertions, will not be considered explicitly: their occurrence, which would result in a reading frameshift of the genetic message, would generally lead to disastrous effects for the reproduction of the organism. Their description can thus be considered as included implicitly in the unknown growth parameter, at least at the level of this simple model.

We will then suppose that the probabilities of occurrence of point mutations are the same for the two populations considered. This amounts to ignoring any mechanism through which the genetic message could control the error frequencies (or at least to assuming that such a feedback is the same in both cases): if one thinks in particular of error repairs, these are most likely to appear only later in evolution and could rightly be neglected here: in the opposite case of mutations affecting components of the translation apparatus their effect can be accounted for implicitly by the growth parameter, since the next generations of these mutant organisms would be more easily prone to an error catastrophe and would not contribute much to the total population (implicitly, we suppose here that our initial organisms have the best possible genetic apparatus, in the sense that any modification makes it less efficient; were this assumption unsatisfied, it would be possible to adapt our argument with a simple redefinition of our initial states). An interesting example was discovered recently (Ya-Ming Hou and Paul Schimmel, 1988) in which errors of transcription are induced by particular modifications of tRNA molecules; this case could easily be included into our description but our model is too crude to usefully incorporate such a result bearing only on one type of mutations. Finally, we will attribute the same numerical value  $p$  ( $p \ll 1$ ) to all point mutations in our concrete example, for the transparency of our results; this amounts in particular to treating all three positions in a triplet on an equal basis.

The effects of mutations are of two types; first the information content of the genes may be modified with time, leading to variations of the genetic message and eventually of its content (if the mutated triplet has a different meaning through the codon catalogue). In our examples, there are respectively 14 and 18 mutant messages, which correspond to the triplets that are the nearest mutational neighbours of the 2 codons constituting each of our 2 codes. They are represented in Figure 1, where the different graphical aspect of these two sets of mutant triplets reveals the difference between the two chosen codes. Secondly, a mutation can lead to a different metabolism, and in particular to a different reproduction rate; with no loss of generality, we shall assume that a mutated organism develops more slowly than the original one. The difference in the growth rates of the initial and mutated populations represents the selective advantage of a faithful translation:

the existence of this difference is our main assumption, without which our conclusions could not be derived. In the same spirit as our choice of identical growth parameters for the two populations with distinct codes, we will consider a common value  $B$  for the growth rates of all mutated organisms in the two cases ( $B < A$ ). The main purpose of this assumption is again to separate the informational aspects that we intend to study from the functional ones. Furthermore this hypothesis, which can be applied also in the case of more complex (and more realistic) genetic messages, simplifies enormously the problem by defining the selective advantage of each initial type of organism as an average over all its mutants: the type of mutation, but also its position in the genetic message becomes then unimportant. The dynamics of the population evolution is simply made more transparent, without being changed qualitatively, by this essentially technical argument. A subsequent benefit is the suppression of an explicit reference to the frequency of use for each triplet, which is then indirectly included in our model through the averaging process over the selective advantage of all mutant organisms.

### 3. Results and Discussion

In principle, then, the spread with time of the genetic message for each of the two initial states is simply calculable in terms of the growth rates (one for the initial organisms, one for their mutants), given their codon catalogues and the probabilities of errors. Since we are interested in qualitative features, and in particular in this first step we do not study the competition between the two populations, but only compare their growth with time, it is sufficient to consider the global populations, without reference to their spatial dependence. If  $x_i = x_i(t)$  represents the number of organisms bearing as its genetic message the triplet  $i$  ( $i = 1, 2, \dots, 64$  so that  $UUU = 1, UUC = 2 \dots CCU = 21, CCC = 22 \dots$  etc. in some ordered manner), it is easy to write the evolution equations for each of our two cases as:

$$dx_i/dt = A_i (1 - \sum_{j \neq i} M_{ij}) x_i + \sum_{j \neq i} M_{ij} A_j x_j + \Phi_i \quad (1)$$

The coefficients  $A_i$  represent the growth parameters for each type of organisms: we have already assumed that  $A_1 = A_{UUU} = A_2 = A_{UUC} = A$  for one case,

$$A_1 = A_{UUU} = A_{22} = A_{CCC} = A \text{ (the same } A)$$

for the other, and that all others are equal to some value  $B, B < A$ .

In the absence of errors ( $M_{ij} = 0 \forall i$  and  $j$ ) and of external constraint ( $\Phi_i = 0$ ), the system (1) writes;

$$dx_i/dt = A_i x_i \quad (2)$$

and the obvious solution is  $x_i(t) = x_i(0) e^{A_i t}$ , so that only the initial organisms proliferate: the two codes give identical results if the initial populations are the same in the two cases.

The coefficients  $M_{ij}$  in Equation (1) describe the mutations from an organism  $i$  to

another,  $j$ , or in our example from the triplet  $i$  to the triplet  $j$ ; they are defined so that  $M_{ii} = 1 - \sum_{j \neq i} M_{ij}$  and  $\sum_j M_{ij} = \sum_j M_{ji} = 1$ . (With our assumptions, only  $9 \times 64 = 576$  values  $M_{ij}$  are equal to  $p$ , which correspond to all pairs of triplets mutant of each other, while all other non-diagonal values are zero).

Finally the flux terms  $\phi_i$  allow the introduction into our model of the notion of competition:  $\phi_i$  can then represent its lethal effects among the organisms of type  $i$ . For instance, one could assume that limited resources will maintain the total population to a constant level  $\sigma = \sum_i x_i$  after some initial growing phase.

Even in our simple example, the system (1) of differential equations introduces a coupling (generally non linear via the flux terms) between all 64 variables  $x_i$ . A general solution is known (Eigen and Schuster, 1977) which we will refer to in our discussion. Here, however, we prefer to exhibit an approximate solution, exact only to first order in  $p$ , and for simplicity we choose to discuss essentially its asymptotic behaviour in a particular case, so as to pinpoint more clearly the dynamical features of interest.

Then we consider the system (1) when  $\phi_i = 0$  and with initial conditions  $x_1(0) = x_2(0) = X$  in one case,  $x_1(0) = x_{22}(0) = X$  in the other: from symmetry arguments, the number of variables reduces strongly and the system can be solved to first order in  $p$ , with respectively

$$\begin{aligned} x_1 = x_2 &= X e^{A(1-8p)t} + O(p^2) \\ \sigma = \sum_i x_i &\sim_{t \rightarrow \infty} 2X e^{A(1-8p)t} (1 + 8pA/(A-B)) + O(p^2) \end{aligned} \quad (3a)$$

and

$$\begin{aligned} x_1 = x_{22} &= X e^{A(1-9p)t} + O(p^2) \\ \sigma = \sum_i x_i &\sim_{t \rightarrow \infty} 2X e^{A(1-9p)t} (1 + 9pA/(A-B)) + O(p^2). \end{aligned} \quad (3b)$$

The comparison between these two solutions shows immediately that the mutant and total populations grow faster in the first case (3a): in fact they get much larger than those in (3b) within a length of time measured characteristically by  $3T/p$ , where  $T$  ( $AT = \log 2$ ) is the doubling period of the populations (in Figure 2 we have represented the ratio  $x_{1(a)}/x_{1(b)}$  as a function of time). The difference comes essentially from the factors  $(1-9p)$  and  $(1-8p)$ , respectively, in the exponentials; these are directly related to the structure of the code table since they measure in each case how often mutations from a sense triplet lead to another sense triplet (in Figure 1, it is easy to count that there are respectively 8 and 9 possible errors for each of the codons constituting the codes). It is then clear that these factors could be generalized to include more than so simple a measure of relatedness, along the lines of recent work (Figureau and Labouygues, 1981; Cullmann and Labouygues, 1983; Figureau and Pouzet, 1984; Labouygues and Figureau, 1984; Figureau, 1985, 1986). Whatever the precise result of such investigations, however, the point is that, in our model, the exponential growth would be governed by factors of relatedness possessing the same structure  $1-rp$  and  $1-r'p$  (Figureau and Pouzet, in preparation). Obviously the population originating from a code with a lower  $r$  will grow faster than others, whence the idea of some optimization principle at the level of the

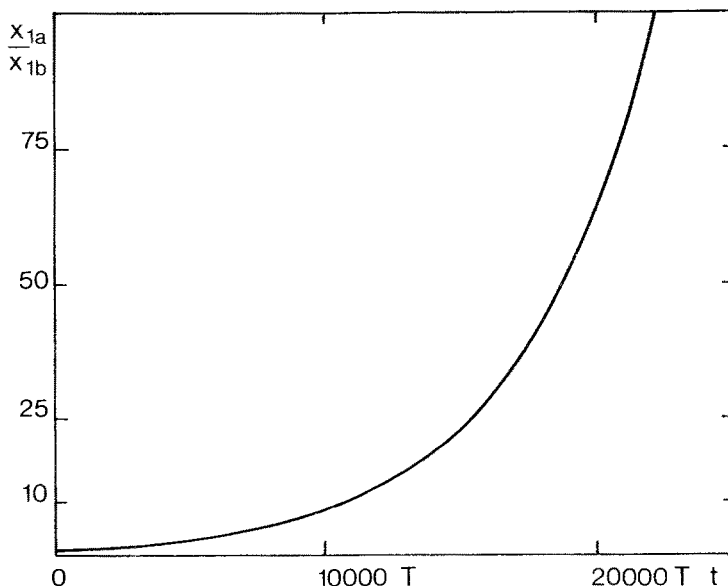


Fig. 2. The time evolution of the ratio  $x_{1a}/x_{1b}$ , where  $x_{1a}$  and  $x_{1b}$  are the populations of organisms with the message UUU for each of the two codes represented in Figure 1. The parameters in our model have been chosen as  $X = 1$ ,  $p = 3 \times 10^{-4}$ ,  $A = (\text{Log } 2)/T$ . The ratio of the total populations  $\sigma_{1a}$  and  $\sigma_{1b}$  varies approximately in the same proportion.

genetic code table. (Note that a lower  $r$  corresponds to a higher resistance to errors as defined in preceding works.)

Our second point in this work is that the time necessary for a population with a 'better' code to overcome all others is certainly very small as compared to geologic times: if we choose  $p = 3 \times 10^{-4}$ , corresponding to missense error frequencies in the  $10^{-3}$  range, an estimate as realistic as possible for such an uncertain and controversial probability (Kurland and Ehrenberg, 1984), we find that the populations built with the code UUU + CCC become negligible as compared to those built with the code UUU + UUC, in less than 20,000 generations (see Figure 2).

This number should not be taken at face value, however: our model is much too crude to give other than qualitative statements. In particular, because of the absence of non-linear terms  $\phi_i$ , which represent the effects of competition, the exponentially growing populations are unrealistically large after such a length of time. Notwithstanding this restriction, we conclude that there is plenty of time for the dominance to be effective.

A more realistic model should take into account the effects of competition: the equations describing the evolution with time of both sets of populations would then be coupled through the terms  $\phi_i$ . The solution of this system of equations introduces the notion of quasi-species, which was studied recently in great detail (Eigen and Schuster, 1977 and 1978), though in a different context. The above populations  $\sigma$ , which include mutant organisms, constitute particularly simple quasi-species.

In this perspective, our conclusions are validated for the solutions of the general system (1) of equations and even of the larger ones that should be studied for longer

genetic messages: under certain conditions (Eigen and Schuster, 1977) on the quantities of information involved and the relative values of our parameters ( $A_i, p$ ), a quasi-species consisting of an error-free population and its mutants can compete successfully with others in a relatively short time. In our case, the determinant quantity, which gives to the organism its selective advantage, is the 'geometry' of the code table, characterized by its resistance to errors, or more generally its degree of relatedness  $r$  (for more realistic code tables). This quantity appears in the Equations (1) together with  $A_i$  and  $p$ , so that the evolution of the populations  $x_i$  is governed by parameters of the type  $A_i pr$ , the leading population  $x_i$  being characterized by the largest value of this parameter. In the light of the results obtained by Eigen and Schuster (1978) and in many later works, it will be most interesting to study the time dependance of solutions for the system (1) in some detail; this is being undertaken, together with a generalization of our model to include the competition between different codes. It is also possible to extend this work to more complex situations (Demetrius, 1983), in order to allow a more realistic description of evolutionary processes.

As an example, we present in Figure 3 a solution of (1) obtained numerically: here we represent the time evolution of all 4 possible messages corresponding to each of two

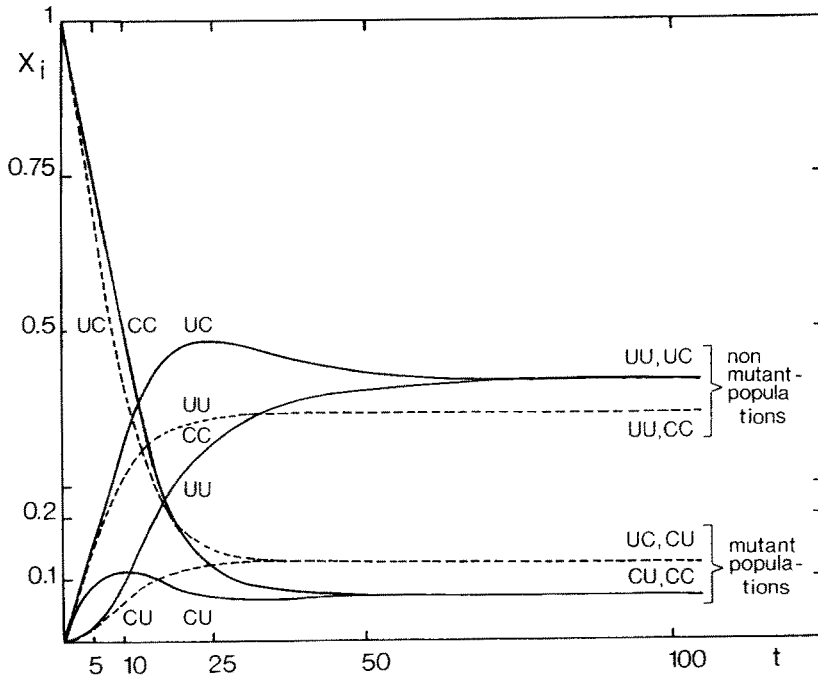


Fig. 3. The time evolution of the populations for 1-word messages built on a doublet code, obtained by solving Equations (1), with  $p = 5 \times 10^{-2}$  and  $A = 1.40B$ . The total populations are constrained to  $\sum_{i=1,64} x_i = 1$ . Full curves are for the code (UU + UC), analogous to our example (UUU + UUC) in the text; dotted curves are for the code (UU + CC), analogous to (UUU + CCC). The initial conditions were respectively

$$x_{CC}(0) = 1, x_{UU}(0) = x_{UC}(0) = x_{CU}(0) = 0,$$

and

$$x_{UC}(0) = 1, x_{UU}(0) = 0, x_{CU}(0) = x_{CC}(0) = 0.$$

Note that equilibrium is attained in few generations, the values of our parameters being larger than in the text.



doublet codes analogous to those discussed precedently. In both cases, for large values of time an equilibrium is attained, which exhibits the structure of the corresponding quasi-species: the populations of the two triplets which constitute the code are always higher at equilibrium than those of the two other triplets. When comparing the two cases, one sees that the more compact code (equivalent to our triplet code UUU + UUC) yields a higher number for the non-mutant population than the other (equivalent to UUU + CCC). This is directly related to its higher resistance to errors. If the two codes were competing instead of being studied separately, only one, or more precisely only the quasi-species built on the most resistant code, would be present after some time: this exact result is a straightforward consequence of the mathematical structure of the equations describing the competition. It does not depend on the values of parameters, nor on the initial populations. It is valid only for particular forms of the non-linearities  $\phi_i$ , and was proven for all three examples which were considered ('constant overall organization' and constant flux: Eigen and Schuster, 1977; constant limited resources: Figureau, work in progress). However, these cases are sufficiently different of each other to describe at least approximately most plausible conditions. The essential condition is that organisms with one code do not mutate into one with another code, so that a quasi-species can be defined for each. Otherwise, the notion of a code would have probably to be redefined for the most primitive organisms.

A second condition bears on the relevance of these results: the predicted equilibrium must be attained, which means that the succession of vocabulary extensions is not too rapid: many nonoptimal codes with  $n$  codons have a better resistance than optimal ones with  $n - 1$  codons, so that the competition between them would not lead to optimal groups of triplets as observed in the present genetic code. Then the occurrence of these groups indicates that vocabulary extensions were a long process leaving time enough for equilibrium to be established. This could mean for example that coding was from the origin a phenomenon more complex than the direct affinity between triplets and amino-acids.

A difficulty seems to appear in our model, however; since many theoretical codes possess the same maximal resistance to errors, our equations lead to a solution in which all of them coexist in some equilibrium. It is then difficult to understand the 'universality' of the present genetic code. One may invoke stochastic factors to explain that only one among many equivalent quasi-species, each having a different optimal code, survived, due to local conditions or to random fluctuations. Another explanation lies in the extreme simplifications introduced in the model; a more realistic, less symmetric, definition of the quantity to be optimized would lead to a unique optimal code (or at most to very few solutions) since the resistance of the codes would all (or nearly all) be different from each other. Then a more detailed analysis would allow to compare quantitatively the development of populations with optimal and non optimal codes: an approach more directly related to the formalisms of population dynamics would be of interest. Even then, however, it would be premature to expect some precise number for the time scale of the code evolution: the determination of our parameters would be a problem in itself, since our model is based on many unknowns and various phenomena are neglected. On the

contrary, a comprehensive study of the solutions for the system (1) would allow to better apprehend the dynamics of genetic code building, and probably to render our model more precise and realistic.

#### 4. Conclusions

Our goal in this work was more modest: we have shown that a particular type of variable, namely the 'geometry' of the code table, can be structured along some optimization principle so as to constitute a selective advantage, and hence can have some pertinence for the study of the genetic code. This was possible within a simple theoretic model, elaborated from a few general ideas; when a particular choice was necessary in this construction, we tried not to alter the dynamics of evolution. Some alternatives could and should be considered, but our conclusion is expected to survive more complex calculations.

In particular, our model in its present state does not describe in any detail the physico-chemical context upon which the genetic code was built. This particularity has the advantage that our conclusions should be valid unless drastic changes in our assumptions are made. On the contrary, we cannot pretend yet to a quantitative understanding of the genetic code structure: such a level will be attained only when more realistic models are tested. Even so, however, previous works (Figureau and Labouygues, 1981; Cullmann and Labouygues, 1983; Figureau and Pouzet, 1984; Figureau, 1986) have exhibited a good convergence between the genetic code and the theoretical predictions deduced from the simplest optimization principle: as our scheme seems adequate for the understanding of the code table, it is a good starting point for further development, be it a search for a more general optimization principle, or for a more detailed characterization of the code mechanisms and of their physico-chemical parameters, or the integration of our model into a more general one. These various approaches should all be taken up in conjunction, so as to lead to a consistent picture of the genetic code in all its complexity.

#### Acknowledgement

The author would like to express his gratitude to Prof. M. Ericson and his physicist colleagues for their sustained encouragement, Prof. M. Pouzet for many fruitful discussions and Prof. J. Ninio for his suggestion at the origin of this work.

#### References

- Alff-Steinberger, C.: 1969 *PNAS* **64**, 584.  
 Bertman, M. O. and Jungck, J. R.: 1979, *J. Heredity* **70**, 379.  
 Cullmann, G. and Labouygues, J. M.: 1983, *Biosystems* **16**, 9.  
 Danckwerts, H. J. and Neubert, D.: 1975, *J. Mol. Evol.* **5**, 327.  
 Demetrius, L.: 1983, *J. Theor. Biol.* **103**, 619.  
 Dunnill, P.: 1966, *Nature* **210**, 1267.  
 Eigen, M. and Schuster, P.: 1977, *Naturwissenschaften* **64**, 541.  
 Eigen, M. and Schuster, P.: 1978, *Naturwissenschaften* **65**, 7.

- Epstein, C. J.: 1966, *Nature* **210**, 25.
- Figureau, A.: 1987, *Biologie Théorique*, Solignac 1985, Ed. C.N.R.S., Paris.
- Figureau, A.: 1986, *Origins of Life* **17**, 439.
- Figureau, A. and Labouygues, J. M.: 1981, *Origin of Life*, ed. Y. Wolman, D. Reidel Publ. Co., Dordrecht, p. 431.
- Figureau, A. and Pouzet, M.: 1984, *Origins of Life* **14**, 579.
- Findley, G. C. and McGlynn, S. P.: 1979, *Int. J. Quant. Chem.: Quantum Biology Symp.* **6**, 313.
- Fitch, W. M.: 1966, *J. Mol. Biol.* **16**, 1.
- Goldberg, A. L. and Wittes, R. E.: 1966, *Science* **153**, 420.
- Hou, Y.-M. and Schimmel, P.: 1988, *Nature* **333**, 140.
- Jukes, T. H.: 1982, *Biochem. Biophys. Res. Com.* **107**, 225.
- Jungck, J. R.: 1978, *J. Mol. Evol.* **11**, 211.
- Karasev, V. A.: 1976, *Vestn. Leningrad Univ., Biol.* **3**, 93.
- Kurland, C. G. and Ehrenberg, M.: 1984, *Progr. Nucl. Acid. Res. and Mol. Biol.* **31**, 191.
- Labouygues, J. M. and Figureau, A.: 1984, *Origins of Life* **14**, 685.
- Lacey, J. C. Jr. and Pruitt, K. M.: 1969, *Nature* **223**, 799.
- Luo, L. F. and Li, Q. Z.: 1986, *Proceedings of the Eighth International Conference on the Origin of Life*, Berkeley, California, U.S.A., July 21–25.
- MacKay, A. L.: 1967, *Nature* **216**, 159.
- Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F., and O'Neal, C.: 1965, *PNAS* **53**, 1161.
- Pelc, S. R.: 1965, *Nature* **207**, 597.
- Root-Bernstein, R. S.: 1984, *J. Theor. Biol.* **94**, 895.
- Söll, D., Ohtsuka, E., Jones, D. S., Lohmann, P., Hayatsu, H., Nishimura, S., and Khorana, H. G.: 1965, *PNAS* **54**, 1378.
- Sonneborn, T. M.: 1965, *Evolving Genes and Proteins*, Rutgers Symposium 1964, Eds. V. Bryson and H. Vogel, Academic Press, London, New York.
- Swanson, R.: 1984, *Bull. Math. Biol.* **46**, 187.
- Tze-Fei-Wong, J.: 1975, *PNAS* **72**, 1909.
- Volkenstein, M. V.: 1965, *Nature*, **207**, 294.
- Woese, C.: 1965, *PNAS* **54**, 71.
- Wolfenden, R. V., Cullis P. M., and Southgate, C. C. F.: 1979, *Science* **206**, 575.