THE EVOLUTION OF AMINOACYL-tRNA SYNTHETASES, THE BIOSYNTHETIC PATHWAYS OF AMINO ACIDS AND THE GENETIC CODE

MASSIMO DI GIULIO

International Institute of Genetics and Biophysics, CNR, Via G. Marconi 10, 80125 Naples, Napoli, Italy

(Received December 2, 1991)

Abstract. In this paper the partition metric is used to compare binary trees deriving from (i) the study of the evolutionary relationships between aminoacyl-tRNA synthetases, (ii) the physicochemical properties of amino acids and (iii) the biosynthetic relationships between amino acids. If the tree defining the evolutionary relationships between aminoacyl-tRNA synthetases is assumed to be a manifestation of the mechanism that originated the organization of the genetic code, then the results appear to indicate the following: the hypothesis that regards the genetic code as a map of the biosynthetic relationships between amino acids seems to explain the organization of the genetic code, at least as plausibly as the hypotheses that consider the physicochemical properties of amino acids as the main adaptive theme that lead to the structuring of the code.

1. Introduction

It can be assumed that a study of the evolutionary relationships between aminoacyl-tRNA synthetase molecules (ARSs) could shed light on how the genetic code is organized. This could be due, for instance, to the consideration that a group of amino acids with similar physicochemical properties might have been loaded on a primordial tRNA (or on a group of similar primordial tRNAs) by a rudimentary primitive ARS (the recognition was unspecific and the results of this were tolerated by the primitive protein synthesis system). The successive diversification of this synthetase may have determined evolutionary relationships between actual ARSs which are specific for similar amino acids. This view is nothing other than one way of presenting what is commonly known as the physicochemical hypothesis of the genetic code (Sonneborn, 1965; Woese *et al.*, 1966). This hypothesis suggests that the organization of the genetic code was mainly determined by a selective pressure tending to reduce the deleterious effects of mutations (Sonneborn, 1965) or translational errors (Woese *et al.*, 1966). Another hypothesis has been made, which suggests that the organization of the genetic code was primarily determined by the mechanism of codon concession by precursor to product amino acids (Wong, 1975). According to this hypothesis, the genetic code represents a map of the biosynthetic relationships between amino acids. Moreover, this hypothesis identifies in the tRNA-like molecule the device on which the transformation of the precursor into the product amino acid took place (Wong, 1975). In this case also, it is easy to hypothesize that amino acids in a precursor-product relationship must have ARSs in an evolutionary relationship. The primordial ARS of a precursor amino acid

Origins of Life and Evolution of the Biosphere **22**: 309–319, 1992.

may certainly have recognized a primordial tRNA, conceded by the precursor to its product, better than a primordial tRNA of an amino acid whose biosynthesis is in relation with another precursor amino acid.

Analysis of the ARSs has shown that the specific molecules for different amino acids are highly heterogeneous in both the primary and the quaternary structure (Schimmel, 1987; Burbaum *et al.*, 1990). This has hindered the study of the evolutionary relationships between these proteins. Nevertheless, analysis of the primary structure of the ARSs has identified two different classes (Eriani *et al.*, 1990) which correlate with the level of aminoacylation of 2'OH or 3'OH of the ribose of the last nucleotide of tRNA. However, this finding does not seem to have immediate consequences for the understanding of the origin of the genetic code (Schimmel, 1991). Recently, Nagel and Doolittle (1991) have built trees showing evolutionary relationships between ARSs. According to the authors (Nagel and Doolittle, 1991) the data argues that the radiation of ARSs took place through the adaptation of binding sites to amino acids with similar physicochemical properties. However, the authors specify that in some cases evolutionarily correlated ARSs recognize dissimilar amino acids (Nagel and Doolittle, 1991).

In this paper I compare the evolutionary trees built by Nagel and Doolittle (1991) with the ones that can be built using the physicochemical properties of amino acids and with the ones that can be obtained from the biosynthetic relationships between amino acids. The goal of this comparison is to find out whether the trees obtained from the ARSs contain traces that help us to better understand how the genetic code originated. Specifically, the comparison should establish whether the evolutionary relationships between ARSs are the result of ARS interaction with the physicochemical properties of amino acids or, rather, the result of ARS interaction with primordial tRNAs (of the precursor amino acids), or are the result of both interactions. This will lead to a discussion of the various hypotheses proposed to explain how the genetic code was organized.

Materials and Methods

The distances of Miyata *et al.* (1979) have been chosen as a measure of the physicochemical similarities existing between amino acids. This distance index is based on two physicochemical properties of amino acids: polarity and molecular volume. The choice of this index is determined by the fact that these two amino acid properties seem to have been important in structuring the genetic code (Jungck, 1978; Di Giulio, 1989a). Moreover, these two properties also have another characteristic, the ability to correlate with the relative substitution frequencies of amino acids (Grantham, 1974; Miyata *et al.*, 1979). This represents a further reason for choosing these properties as the ones that best identify the physicochemical similarities between amino acids (Wong, 1980). [Results equivalent to the ones that will be shown below are obtained using the Grantham's index (1974) instead of the index of Miyata *et al.* (1979)].

Starting from the distances matrix (Miyata *et al.*, 1979, Table 1) the tree topologies has been built using three different algorithms: the unweighted pair-group method (UPG method) as described by Nei (1975), the Li method (1981) (MUPG method) and the neighbor-joining method (NJ method) (Saitou and Nei, 1987). The UPG method is generally used in cluster analysis (even if this has been applied to distances matrices coming from molecular data) and was originally proposed to build phenograms (Sokal and Michener, 1958). The other two methods (Li, 1981; Saitou and Nei, 1987) are typically applied to data coming from the comparison of sequences of proteins or DNAs. Therefore, the choice of the three methods is justified by the fact that they should guarantee a diversity between the topologies identified. The NJ method (Saitou and Nei, 1987) has been shown to be efficient in identifying the true topology of a phylogenetic tree (Sourdis and Nei, 1988). While, as the MUPG method (Li, 1981) is a modification of the UPG method adapted to data coming from molecular evolution, it may represent an intermediate clustering procedure between the UPG method and the NJ method.

After preassigning the amino acids to either group I or group II, which are defined according to the ARS sequences (Nagel and Doolittle, 1991), these three algorithms identified topologies using Miyata *et al.*'s distances (1979, Table 1) as input data. For instance, for group I of the ARSs, the input data provided consists of the 36 values of the distances (Miyata *et al.*, 1979) which refer to all the possible distances between the 9 amino acids belonging to group I (Figure 2a). The above algorithms are used only to build the tree topologies reflecting the amino acids physicochemical properties (see Results). Whereas, with the amino acids still preassigned to groups I or II of Nagel and Doolittle (1991), the tree topologies reflecting the biosynthetic relationships between amino acids were obtained by transforming Wong's Figure 1 (1975, p. 1910) into a binary tree. These topologies are shown in Figure 3. The comparison between these topologies (Figure 3a,b)

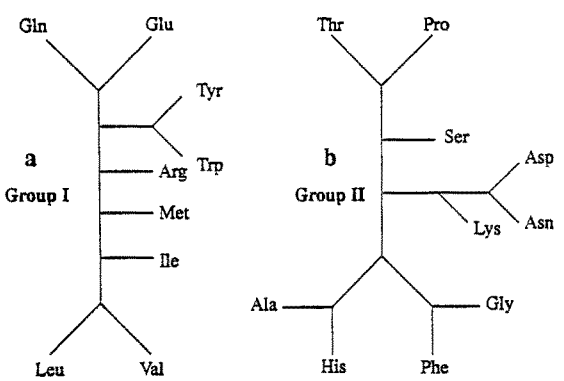

Fig. 1. This shows the (re-adapted) evolutionary relationships between the ARSs identified by Nagel and Doolittle (1991, Figure 5, p. 8124). The specific ARS for a given amino acid is indicated using the three-letter code to identify the corresponding amino acid. Groups I and II refer to the synthetase groups identified by Nagel and Doolittle (1991).

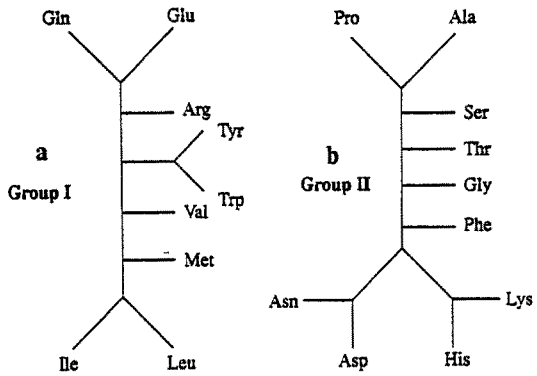

Fig. 2. Topologies of the trees built by starting from the physicochemical properties of amino acids (see Materials and Methods; Results). The amino acids are shown using the standard abbreviations. Groups I and II refer to the synthetase groups identified by Nagel and Doolittle (1991).

and the biosynthetic relationships between the amino acids (Wong 1975, Figure 1, p. 1910) should clarify the straight forward criterion used to build these binary trees, since Wong's Figure 1 is almost in tree form. Moreover, the considerations Wong makes on the biosynthetic relationships between amino acids were also used to build these trees (Figure 3). The reader is invited to read the comments Wong (1975) makes on his Figure 1 so as to fully understand the straight forward criterion adopted to build the topologies of Figure 3. Some details for the construction of these trees using Wong's comments are given in the Results. Finally, with the amino acids still preassigned to groups I or II (Nagel and Doolittle, 1991), the tree topologies that can be associated to the ambiguity reduction hypothesis (Fitch and Upper, 1987) have been built using the criterion introduced by Fitch and Upper (1987, Figure 2; p. 762). In short, starting from the undifferential codon NNN

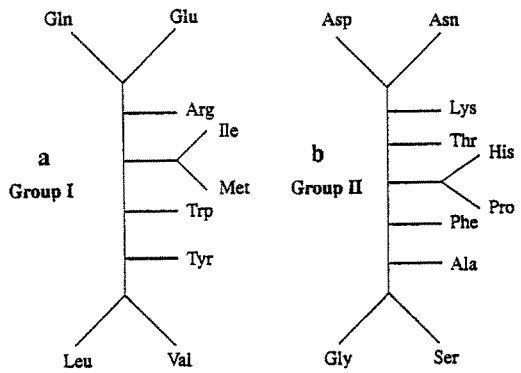

Fig. 3. Biosynthetic relationships between amino acids expressed in the form of a binary tree. See the text (Materials and Methods; Results) for further information. The amino acids are shown using the standard abbreviations. Groups I and II refer to the synthetase groups identified by Nagel and Doolittle (1991).

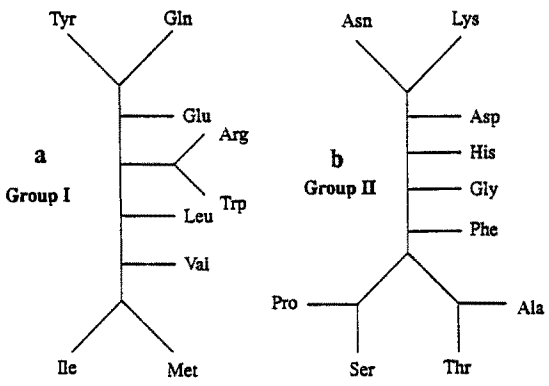

Fig. 4. Trees showing some of the evolutionary relationships between the amino acids as suggested in the ambiguity reduction hypothesis (Fitch and Upper, 1987). See the text (Materials and Methods; Results) for further information. The amino acids are shown using the standard abbreviations. Groups I and II refer to the synthetase groups identified by Nagel and Doolittle (1991).

(N = any base), exact tree topologies can be built by differentiating codons, for instance, first in the second position thus producing the codons NRN and NYN (R = purine and Y = pyrimidine) and then in the first position producing the codons YRN, RRN, YYN, and RYN. Every codon differentiation step produces a bifurcation on the tree. The further reduction of codon ambiguity makes it possible to identify exact topologies. In the Results, further information is provided for building these trees, two of which are shown in Figure 4. However, the reader is referred to the work of Fitch and Upper (1987, Figure 2, p. 762) for a full understanding of the codon differentiation patterns (see Results) used to build the topologies of these trees.

The comparison between two binary trees T_1 and T_2 is made using the partition metric (Robinson and Foulds, 1979, 1981). The distance between the topology of the T_1 and T_2 trees, indicated with $d(T_1, T_2) = m$, represents the number of classes derived from T_1 or T_2 but not both. For a fast method to calculate m (= d (T_1, T_2)), see Penny and Hendy (1985). The probability [P (m, n)] associated to a given value of m, with n indicating the number of operational taxonomic units has been evaluated using Table 4 in Hendy *et al.* (1984) (see also Table 1 in Penny *et al.*, 1982).

Finally, another method is used in this paper. Fisher (1950) showed that the quantity $-2 \ln P_i$ *(where* P_i is the probability deriving from the i-th significance test) follows an χ^2 distribution with 2 degrees of freedom (*df*). Thus, the quantity $-2\sum_{i=1}^{K} \ln P_i$ has an χ^2 distribution with 2K degrees of freedom. This method makes it possible to join the probabilities deriving from independent significance tests in a single value.

Results

Figure 1 shows the evolutionary relationships between the aminoacyl-tRNA synthetase molecules (ARSs) as identified by Nagel and Doolittle (1991). In group I of the synthetases, Val and Leu (Figure 1a) form a pair of terminal vertices linked to the same internal vertex, even if in Nagel and Doolittle's trees Ile is also shown to be associated to Val. This assumption is justified both by the smaller value of the distance scores (between Val and Leu), as can be seen in Figure 5 of Nagel and Doolittle (1991), and by the fact that the ARS of Val of *S. cerev.* and the ARS of Leu of *E. coli* show the maximum percentage identity of all those analyzed by Nagel and Doolittle (1991).

Figure 2a shows the topology identified regardless of the clustering algorithm adopted, by using the distances of Miyata *et al.* (1979) (see Materials and Methods). Whereas Figure 2b shows the topology identified by the UPG method. This topology is different from the ones generated by the NJ method and the MUPG method. However, all three methods identify the same pairs (Pro, Ala) (Asp, Asn) and (Lys, His) of terminal vertices linked to the same internal vertex and, for the purposes of tree comparison, this is sufficient to univocally define a probability value.

Figure 3 shows the topologies that can be obtained by transforming the biosynthetic relationships between amino acids into a binary tree (Wong, 1975, Figure 1) (see Materials and Methods). In this transformation, the considerations Wong (1975) makes on the precursor-product amino acid relationships have also been used. For instance, in Figure 3b Asp and Asn are terminal vertices linked to the same internal vertex. As Lys originated from Asp, Lys could therefore occupy the position of Asn. However, the latter topology is less likely in the light of the coevolution hypothesis as, in order to remove all the non-contiguities, Wong (1975) postulated that Asn (like Gln) was derived from Asp (from Glu) at a later stage compared to the distribution of codons from precursor to product amino acids. Analogous considerations made for the other amino acids seem to justify the choice of the topology in Figure 3b.

Table I shows the probabilities that are obtained by making all the possible comparisons between the tree topologies shown in Figures 1, 2, 3.

I then made a more general comparison. I joined the topologies of groups I and II, assuming that these had a common origin even if Nagel and Doolittle (1991) did not find a significant similarity in the sequences between the two groups and therefore suggested the possibility that each group once supported an independent protein synthesis system. This assumption creates an unrooted tree with 19 terminal vertices. [This tree was obtained by joining the interior branch between Arg and the pair (Tyr, Trp) (Figure 1a) with the interior branch between Ser and the cluster (Lys, Asp, Asn) (Figure 1b). This is suggested by Figure 5 in Nagel and Doolittle (1991)]. By comparing this tree with the tree (not shown) of the biosynthetic relationships between amino acids, the following common pairs or terns are seen: (Asp, Asn), (Asp, Asn, Lys), (Glu, Gln) and (Val, Leu). These simi-

larities between the two trees determine a value $m = 24$ ($n = 19$) and a probability value of $P < 8.4 \times 10^{-5}$ (value extrapolated from Table 4 in Hendy *et al.*, 1984).

The topology of the tree deriving from the physicochemical properties of the 19 amino acids shares only the pair (Asp, Asn) and the cluster (Leu, Ile, Met, Val) (using the UPG method) with the tree obtained by joining groups I and II. These similarities give a value of $m = 28$ ($n = 19$) and a value of $P < 0.0147$ (value extrapolated from Table 4 in Hendy *et al.*, 1984). Of course, the three clustering algorithms produce three different topologies, and the topology that best seems to reflect the similarities between amino acids is the one derived from the UPG method. None of the three topologies show the pairs (Glu, Gln) (in all three topologies Glu forms a cluster with Asp and Asn) and (Tyr, Trp) (the pairs (Phe, Tyr), (Phe, Trp) and (Phe, Trp) are seen) which would lower the value of m and, therefore, the probability. While, in all the topologies the pairs (Pro, Ala) and (Lys, Arg) are observed. The invariance of the (Pro, Ala) pair precludes the possibility for the clusters (Thr, Pro) or (Thr, Pro, Ser) on the part of the tree shown in Figure 1b to be shared by the tree deriving from the physicochemical properties. Therefore, the value of m does not decrease. The invariance of the (Lys, Arg) pair on the physicochemical tree precludes (together with the behavior of Glu on these topologies) the possibility for the groups (Asn, Asp, Lys) on the part of the tree shown in Figure 1b to be shared by the physicochemical tree. Therefore, the value of m does not decrease. Overall, this and other observations lead to the conclusion that the most likely value is $m = 28$, even if other clustering algorithms or distance indices could lower this value to 26 (for instance, by forming the pair (Tyr, Trp)).

Fitch and Upper (1987) proposed the ambiguity reduction hypothesis to explain the origin of the genetic code. In order to test this hypothesis I have built binary trees (analogous to the ones in Figures 2, 3) using the criterion introduced by Fitch and Upper (1987, Figure 2, p. 762) (see Materials and Methods). There are several tree topologies that can thus be associated to the structure of the genetic code. First, the code could be decomposed so as to obtain trees formed of four main clusters of amino acids belonging either to the columns (second codon base) or to the rows (first codon base) of the code. As far as we know about the distribution of the physicochemical properties in the genetic code, it seems that amino acids belonging to the same column of the code have similar physicochemical properties (Wolfenden *et al.*, 1979; Sjostrom and Wold, 1985; Di Giulio, 1989a). I have thus built four trees, two of which are shown in Figure 4 and which refer to a codon differentiation made according to the columns of the genetic code. None of the four trees thus built showed significant values of m ($P \geq 0.176$) when compared with the trees in Figure 1. Second, the tree topologies that can be associated to the genetic code derive from considering the code as being formed of four squares with codons of type YYN, RYN, YRN and RRN (Y = pyrimidine, R = purine and N = any base) inside each square. The successive codon differentiation by row or column makes it possible to accurately define tree topologies. In these codon differentiation patterns, only one case in which m assumes a significant value is

observed. This case is observed in group I of the ARSs (Figure 1a) with the formation of the pair (Tyr, Trp) and the cluster (Ile, Met, Val, Leu) ($m = 8$, $n = 9$, P (m, n) = 0.0314). This is obtained by differentiating the codons inside each of the four squares, first in the second codon position and then on the first codon base. By following the same codon differentiation pattern I have built a tree that has been compared to group II of the ARSs (Figure 1b), obtaining a non-significant value of m ($m = 14$, $n = 10$, $P = 0.793$). If the data from the two comparisons is joined to form a single probability value (see Materials and Methods), a non-significant χ^2 value is obtained ($\chi^2 = 7.38, df = 4, P > 0.10$). Moreover, this codon differentiation pattern cannot be justified.

Distribution of YYN, RYN, YRN and RRN type codons on the topology of group II of the ARSs (Figure 1b) is such that only the cluster (Asn, Lys, Asp) may be observed, considering all the codon differentiation patterns so far used. In these cases the smallest value of m is non-significant ($m = 12$, $n = 10$, P (m, n) = 0.176).

Finally, I have built two trees for the 19 amino acids. The topologies of these trees refer to the clustering of amino acids according to the rows and columns of the genetic code. These trees have been compared with the tree joining groups I and II of the ARSs. These comparisons did not give significant values of m. (The smallest value was $m = 30$ with $P \simeq 0.13$ ($n = 19$). The value extrapolated from Table 4 in Hendy *et al.* (1984)). Non-significant values of m were also obtained from the comparison of the tree topologies for the 19 amino acids, obtained by differentiating the codons inside the four squares both by row and by column, with the tree joining groups I and II of the ARSs.

Discussion

The conclusion that can be drawn from this paper is that the tree topologies identified through a study of the evolutionary relationships between aminoacyl-tRNA synthetases (ARSs) (Nagel and Doolittle, 1991) are consistent (Table I) with the hypotheses that were suggested in order to explain the origin of the genetic code (Sonneborn, 1965; Woese *et al.*, 1966; Wong, 1975). (This seems to confirm the reliability of the topologies identifed by Nagel and Doolittle (1991)). This might seem obvious, at least from some points of view, but it is not really given the considerable heterogeneity characterizing ARSs (Schimmel, 1987; Burbaum *et al.*, 1990). This has led to the suggestion that as there might be many ways of transforming an RNA enzyme (which existed in the RNA world (Gilbert, 1986)) into a protein enzyme (White, 1982; Alberts, 1986), these ways should justify the observed diversity between the ARSs (Weiner and Maizels, 1987).

The quantitative analysis performed in this paper confirms, at least for group I of the ARSs, the conclusion reached by Nagel and Doolittle (1991) that one of the main themes in the evolution of synthetases involved the physicochemical properties of amino acids (Table I). However, this is not verified for group II of

TABLE I

This shows the probabilities calculated using the partition metric (see Materials and Methods) of all the possible comparisons of tree topologies. The values of m and n (m, n) are shown in brackets (see Materials and Methods). A indicates the topologies deriving from the paper of Nagel and Doolittle (1991) (Fig. 1). B indicates the tree topologies deriving from the biosynthetic relationships between amino acids (Fig. 3). C indicates the topologies deriving from the physicochemical properties of amino acids (Fig. 2). Groups I and II indicate the groups of synthetases originally indentified by Nagel and Doolittle (1991).

	B	C
Group I A	0.0314 (8, 9)	0.00480 (6, 9)
Group II	0.0265 (10, 10)	0.176 (12, 10)
Group I B		0.0314 (8, 9)
Group II		0.176 (12, 10)

the ARSs (Table I). Table I shows that the evolutionary relationships between ARSs are also explanined, in both groups, by the precursor-product amino acid relationships predicted by the coevolution hypothesis (Wong, 1975). This seems to suggest that the radiation of ARSs took place through the adaptation of these molecules to the primordial tRNA of precursor amino acids (or to the amino acid-primordial tRNA complex). The results in Table I do not make it possible to find out if the physicochemical properties of the amino acids or the precursor-product relationships (the primordial tRNA of the precursors) had a greater influence in modeling the evolution of ARSs ($\chi^2 = 14.15$, $df = 4$, $P < 0.01$, versus $\chi^2 = 14.18$, $df = 4$, $P < 0.01$, Table I; see Materials and Methods). It seems that both types of interactions were important. The situation is further complicated by the relationships between the topologies deriving from the physicochemical properties of amino acids and those deriving from the precursor-product amino acid relationships ($\chi^2 = 10.40$, $df = 4$, $P < 0.05$; Table I) (Di Giulio, 1991).

However, the fact that the analysis performed in this paper has not been able to provide data in favour of the ambiguity reduction hypothesis (Fitch and Upper, 1987) would seem to suggest that the results of Fitch and Upper (1987) must be interpreted in favour of the coevolution hypothesis of the genetic code (Wong, 1975) rather than in favour of the physicochemical hypothesis (Sonneborn, 1965; Woese *et al.*, 1966). This seems to derive from the following considerations. The ambiguity reduction hypothesis (Fitch and Upper, 1987) does not seem to be substantially different from the physicochemical hypothesis (Sonneborn, 1965; Woese *et al.*, 1966) which seems to be a more general hypothesis. Moreover, the former hypothesis has the merit of making more accurate predictions in terms of binary trees than the latter. The fact that it was not possible to find data in favour of

the ambiguity reduction hypothesis would seem to suggest that the phylogeny of tRNAs, which is in relation with the pattern of the genetic code (Fitch and Upper, 1987), is therefore the result of the concession of codons from the precursor to the product amino acids (Wong, 1975) rather than an expression of the physicochemical hypothesis (Sonneborn, 1965; Woese *et al.*, 1966). It is somewhat curious that in ARS tree topologies (Figure 1) very few and non-significant cases are identified in which two terminal vertices linked to the same internal vertex bear amino acids codified by codons with the form YYN, RYN, YRN or RRN, as suggested by the ambiguity reduction hypothesis (Fitch and Upper, 1987) or more generally by the physicochemical hypothesis (Sonneborn, 1965; Woese *et al.*, 1966). The data seems to be interesting given that the properties of the amino acids have been repeatedly related with the organization of the genetic code (Woese *et al.*, 1966; Epstein, 1966; Goldberg and Wittes, 1966; Allf-Steinberger, 1969; Jungck, 1978; Weber and Lacey, 1978; Di Giulio, 1989b) and in particular with the columns of the code (Wolfenden *et al.*, 1979; Sjostrom and Wold, 1985; Di Giulio, 1989a).

Therefore, the assumption that the mechanism that gave rise to the organization of the genetic code is seen in the evolutionary relationships between the ARSs leads to the conclusion suggested by Wong (1980) that the physicochemical properties might have played a subsidiary role in structuring the genetic code and that it was the precursor-product relationships that defined its main layout. The latter conclusion is also supported by the comparison between the ARS general tree (obtained by joining groups I and II of the synthetases) and (i) the general tree derived from the biosynthetic relationships between amino acids, (ii) the tree derived from the physicochemical properties of amino acids (and with the trees that can be associated to the ambiguity reduction hypothesis). However, the result of the latter comparisons must be considered with care as there is no evidence of a common origin for the ARSs of groups I and II (Eriani *et al.*, 1990; Nagel and Doolittle, 1991).

In conclusion, the present analysis has shown that the mechanism of codon concession from the precursor to the product amino acid (Wong, 1975) might have been the main factor determining the structuring of the genetic code. However, the physicochemical properties of amino acids must have played an important role in structuring the genetic code in the sense that amino acids with similar properties originated and entered the code in response to a selective pressure tending to provide the protein structure with groups of amino acids capable of performing almost analogous functions, e.g., the aliphatic amino acid group (this might be one of the reasons why similar amino acids are codified by similar codons). Otherwise it would not be possible to explain why the physicochemical properties of amino acids are (i) reflected in the ARS trees (Table I), (ii) related to the precursor-product relationships (Table I) (Di Giulio, 1991) and (iii) more generally related to the organization of the genetic code (Woese *et al.*, 1966; Epstein, 1966; Goldberg and Wittes, 1966; Allf-Steinberger, 1969; Jungck, 1978; Weber and Lacey, 1978; Wolfenden *et al.*, 1979; Sjostrom and Wold, 1985; Di Giulio, 1989a, b).

Acknowledgements

I would like to thank Dr. Wen-Hsiung Li and Prof. Masatoshi Nei for having kindly provided me with the programs used for constructing binary trees. I also thank Marco Valenzi for his help in using the computer.

References

Alberts, B. M.: 1986, *Am. Zool.* **26**, 781–796.
Allf-Steinberger, C.: 1969, *Proc. Natl. Acad. Sci. USA* **64**, 584–591.
Burbaum, J. J., Strazyk, R. M., and Schimmel, P.: 1990, *Proteins* 7, 99–111.
Di Giulio, M.: 1989a, *J. Mol. Evol.* **29**, 191–201.
Di Giulio, M.: 1989b, *J. Mol. Evol.* **29**, 288–293.
Di Giulio, M.: 1991, *Z. Naturforsch.* **46c**, 305–312.
Epstein, C. J.: 1966, *Nature* **210**, 25–28.
Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D.: 1990, *Nature* **347**, 203–206.
Fisher, R. A.: 1950, *Statistical Methods for Research Workers*, 11-th ed., p. 99, Oliver & Boyd, Edinburgh, London.
Fitch, W. M. and Upper, K.: 1987, *Cold Spring Harbor Sym. Quant. Biol.* **52**, 759–767.
Gilbert, W.: 1986, *Nature* **319**, 618.
Goldberg, A. L. and Wittes, R. E.: 1966, *Science* **153**, 420–424.
Grantham, R.: 1974, *Science* **185**, 862–864.
Hendy, M. D., Little, C. H. C., and Penny, D.: 1984, Soc. Ind, Appl. Math. (SIAM), *J. Appl. Math.* **44**, 1054–1067.
Jungck, J. R.: 1978, *J. Mol. Evol.* **11**, 211–224.
Li, W-H.: 1981, *Proc. Natl. Acad. Sci. USA* **78**, 1085–1089.
Miyata, T., Miyazawa, S., and Yasunaga, T.: 1979, *J. Mol. Evol.* **12**, 219–236.
Nagel, G. M. and Doolittle, R. F.: 1991, *Proc. Natl. Acad. Sci. USA* **88**, 8121–8125.
Nei, M.: 1975, *Molecular Population Genetics and Evolution* , North-Holland, Amsterdam, pp. 199–201.
Penny, D., Foulds, L. R., and Hendy, M. D.: 1982, *Nature* **297**, 197–200.
Penny, D. and Hendy, M. D.: 1985, *Syst. Zool.* **34**, 75–82.
Robinson, D. F. and Foulds, L. R.: 1979, in *Lecture Notes in Mathematics*. Volume 748, Springer-Verlag, Berlin, pp. 119–126.
Robinson, D. F. and Foulds, L. R.: 1981, *Math. Biosci.* **53**, 131–147.
Saitou, N. and Nei, M.: 1987, *Mol. Biol. Evol.* **4**, 406–425.
Schimmel, P.: 1987, *Annu. Rev. Bioch.* **56**, 125–158.
Schimmel, P.: 1991, *Trends Biochem. Sci.* **16**, 1–3.
Sjostrom, M. and Wold, S.: 1985, *J. Mol. Evol.* **22**, 272–277.
Sokal, R. R. and Michener, C. D.: 1958, *Univ. Kansas Sci. Bull.* **38**, 1409–1438.
Sonneborn, T. M.: 1965, in Bryson, V. and Vogel, R. J. (eds.), *Evolving Genes and Proteins*, Academic Press, New York, pp. 377–397.
Sourdis, J. and Nei, M.: 1988, *Mol. Biol. Evol.* **5**, 298–311.
Weber, A. L. and Lacey, J. C. Jr.: 1978, *J. Mol. Evol.* **11**, 199–210.
Weiner, A. M. and Maizels, N.: 1980, *Proc. Natl. Acad. Sci. USA* 77, 1083–1086.
White, H. B. III.: 1982, in Everse, J., Anderson, B. and You, K., (eds.) *The Pyridine Nucleotide Coenzymes*, Academic, New York, pp. 1–17.
Woese, C. R., Dugre, D. H., Dugre, S. A., Kondo, M., and Saxinger., W. C.: 1966, *Cold Spring Harbor Symp. Quant. Biol.* **31**, 723–736.
Wolfenden, R. V., Cullis, P. M., and Southgate, C. C. B.: 1979, *Science* **206**, 575–577.
Wong, J. T-F.: 1975, *Proc. Natl. Acad. Sci. USA* **72**, 1909–1912.
Wong, J. T-F.: 1980, *Proc. Natl. Acad. Sci. USA* 77, 1083–1086.