The Coevolution Theory of the Origin of the Genetic Code

In order to establish the statistical significance between the biosynthetic pathways of amino acids and the organization of the genetic code, Amirnovin (1997) generates random codes of amino acid permutation (Di Giulio 1989), i.e., codes that maintain the relative position of the blocks of synonymous codons constant, as in the genetic code, but also allow amino acids to be randomly permuted in the various positions of the genetic code. Amirnovin (1997) associates a value to every pair of amino acids which is simply given by the number of times that the codons of a certain amino acid transform (by single base substitution alone) into those of another amino acid on the basis of the genetic code structure. The sum of these numbers for the various pairs of amino acids considered represents the ‘codon correlation score’ (CCS) for a given amino acid (Amirnovin 1997).

In the tests reported in his paper, Amirnovin (1997) generates 32,000 random codes which allow him to build a distribution, go and see where in this distribution the set of pairs of amino acids in biosynthetic relationships considered is located, and then ascribe a probability to that set.

The first set of pairs of amino acids in biosynthetic relationships used by Amirnovin (1997, Table 1) is the same one used by Wong (1975) to support the coevolution theory. Amirnovin (1997) finds a probability of 0.001, against the probability of 0.0002 found by Wong (1975). As Wong’s probability is only five times smaller than Amirnovin’s; this suggests that the two methods are, on the whole equivalent, and that, therefore, it remains confirmed that for these eight pairs of amino acids in precursor–product relationship, there is a strong correlation between the biosynthetic pathways of amino acids and the organization of the genetic code.

Amirnovin (1997) criticises use of these pairs of amino acids as he maintains that this set contains two pairs, Val–Leu and Gln–His, which contribute more than the other pairs to making the probability of 0.001 so small. This is debatable. Although the Val–Leu pair does contribute to the CCS value with 6 units, the Ser–Trp pair contributes with 1 unit, i.e., the lowest possible value given the structure of the genetic code and the precursor–product relationship, whereas the value of 4 units attributed to the Gln–His pair is, in this framework, counterbalanced by the pairs Glu–Gln, Asp–Asn, and Phe–Tyr, which are attributed 2 units, which turn out to be among the lowest attributable in view of the genetic code structure. Finally, it is worthwhile pointing out that in Wong’s (1975) method, only the Val–Leu pair has a low probability compared to the other pairs (Wong 1975, Table 2), but this hypothetical irregularity seems to be compensated by the use of two pairs, Ser–Trp and Ser–Cys, for which the probabilities are not significant (Wong 1975, Table 2). Therefore, the pairs used by Wong (1975) have nothing special about them, as is maintained by Amirnovin (1997), and there is thus no reason why these two pairs (Val–Leu and Gln–His) should be eliminated and the probabilities recalculated, as Amirnovin (1997) does.

Finally, Amirnovin (1997) uses another set of 12 pairs of amino acids in biosynthetic relationships (Lehninger et al. 1973) and finds this set not to be significant with a probability of 0.34. This set contains some pairs, Asp–Thr, Asp–Met, Asp–Lys, Glu–Arg, and Glu–Pro, which contribute to the CCS value by 0 unit. Use of these pairs in this way is incorrect in respect of the coevolution theory as it does not take into account some of this theory’s assumptions. Indeed, in order to remove some of the noncontiguities between amino acids in biosynthetic relationships Wong (1975) postulated that the codons of Asn codified for Asp for a long period of time during the evolution of the genetic code and that only at the end of this evolution were they conceded to Asn. An analogous consideration was made for the pair Glu–Gln, and therefore, the codons of Gln likewise belonged to Glu for a long period of time (Wong 1975). These two simple postulates are able to remove all the noncontiguities existing in the genetic code between amino acids in biosynthetic relationships (Wong 1975).
If we therefore take these postulates into account in the calculation of the CCS, we obtain a value of at least 26 units, against the 16 units calculated by Amirnovin (1997) and, therefore, this CCS value of 26 units implies (Amirnovin 1997, Fig. 2) a highly significant probability.

However, what arguments do we have nowadays, 24 years after the formulation of the coevolution theory, to maintain that the two postulates described above are justified? The coevolution theory postulates that the mechanism for the concession of codons from the precursor amino acid to the product amino acid took place through tRNA-like molecules on which, the theory maintains, the biosynthetic transformations between amino acids occurred (Wong 1975). Therefore, this theory predicts that it is possible to identify molecular fossils that are witnesses of these ancient biosynthetic pathways. In actual fact, these fossils have both been identified, one regarding the pathway Asp-tRNAAsn (Marcker and Sanger 1964; Guillon et al. 1988) and the other regarding the pathway Glu-tRNAGln (Wilcox and Nirenberg 1968; Schon et al. 1988) and the other tRNA Sec (Bock et al. 1991) and Met-tRnAfMet (Wong 1975). The (tRNAs charged in this unusual way later take part in protein synthesis). Moreover, the existence of other pathways, Ser-tRNA Sec → Sec-tRNA Sec (Bock et al. 1991) and Met-tRNA Met → Met-tRNA Met (Marcker and Sanger 1964; Guillon et al. 1992), occurring on tRNAs seems to indicate that this phenomenon is more common than was previously thought. Indeed, the absence of cysteinytl-tRNA synthetase from the genome of Methanococcus jannaschii (Bult et al. 1996) could imply the existence of another of what we can now call Wong’s pathways. Moreover, the anomalous phylogenetic distribution of lysyl-tRNA synthetase, which in some eubacteria and archaeabacteria is class I, while in others it is class II (Ibba et al. 1997), can be easily explained with a biosynthetic pathway leading to Lys and taking place on a tRNA. In this case, even if there were a Wong’s pathway, it could have been abandoned at different evolutionary times and could have been compensated by the evolution, in some cases, of a lysyl-tRNA synthetase of class I. (This interpretation could be of a more general meaning). Moreover, in light of this curious distribution of lysyl-tRNA synthetase (Ibba et al. 1997) this pathway on the tRNA might still be present in living organisms and might therefore still be identifiable. Therefore, these molecular fossils should lead us to consider the coevolution theory as the best theory available in order to explain the origin of the genetic code (Di Giulio 1997a, b).

References

Amirnovin R (1997) An analysis of the metabolic theory of the origin of the genetic code. J Mol Evol 44:473–476
Baron C, Bock A (1995) The selenocystein-inserting tRNA species: Structure and function. In: Sool D, RajBhandary UL (eds) tRNA: structure, biosynthesis, and function. ASM Press, Washington, DC, pp 529–544
Bock A, Forchhammer K, Heider J, Leinfelder W, Sawers G, Veprek B, Zinoni F (1991) Selenocysteine: the 21st amino acid. Mol Microbiol 5:515–520
Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghegan NSM, Venter JC (1996) Complete genome sequence of the methanogenic archaean, Methanococcus jannaschii. Science 273:1058–1073
Curnow AW, Ibba M, Soll D (1996) tRNA-dependent asparagine formation. Nature 382:589–590
Di Giulio M (1989) The extension reached by the minimization of polarity distances during the evolution of the genetic code. J Mol Evol 29:288–293
Di Giulio M (1997a) On the origin of the genetic code. J Theor Biol 187:573–581
Di Giulio M (1997b) The origin of the genetic code. Trends Bioch Sci 22:49
Guillon JM, Meinnel T, Mechulam Y, Lazennec C, Blanquet S, Fayat G (1992) Nucleotides of tRNA governing the specificity of Escherichia coli methionyl-tRNA formyltransferase. J Mol Biol 224: 359–367
Ibba M, Bono JL, Rosa PA, Soll D (1997) Archaeal-type lysyl-tRNA synthetase in the Lyme disease spirochete Borrelia burgdorferi. Proc Natl Acad Sci USA 94:14383–14388
Lehninger A, Nelson D, Cox M (1993) Principles of biochemistry. Worth, New York, pp 697–709
Markert KA, Sanger F (1964) N-formylmethionyl-sRNA. J Mol Biol 8:835–840
Schon A, Kannagara CG, Gough S, Soll D (1988) Protein biosynthesis in organelles requires misaminoacylation of tRNA. Nature 331: 187–190
Wilcox M, Nirenberg M (1968) Transfer RNA as a cofactor coupling amino acid synthesis with that of protein. Proc Natl Acad Sci USA 61:229–236
Wong JT (1975) A co-evolution theory of the genetic code. Proc Natl Acad Sci USA 72:1909–1912

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

* * * * *

Response

DiGiulio has criticized the analysis of the metabolic theory of the genetic code. His discussion deals with details of the choice of biosynthetic relationships and statistical scores. Before dealing with these points, we would like to restate our conclusion. The coevolution theory of the origin of the genetic code is an attractive theory, but it cannot be proved by codon correlations. The biosynthetic pathways of amino acids are very interrelated, and so codons of almost any code will show correlations. The present genetic code shows many such correlations, but there are other genetic codes that show more. Another way of stating this is that there would be