Journal of Theoretical Biology 399 (2016) 134–140

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

The lack of foundation in the mechanism on which are based the physico-chemical theories for the origin of the genetic code is counterposed to the credible and natural mechanism suggested by the coevolution theory☆

Massimo Di Giulio

Early Evolution of Life Laboratory, Institute of Biosciences and Bioresources, CNR, Via P. Castellino, 111, 80131 Naples, Italy

HIGHLIGHTS

- Suggestion of a hypothesis on how domains of codons of the precursor amino acids evolved is suggested.
- Elongation factors constrained the evolution of all amino acids of a given biosynthetic family.
- The elongation factors recognized two characteristics of the progenitor pre-tRNAs of precursor amino acids.
- This defined the evolution of the codon domain formation.

ARTICLE INFO

Article history:
Received 4 February 2016
Received in revised form
29 March 2016
Accepted 1 April 2016
Available online 8 April 2016

Keywords:
Precursor amino acid codon domain formation
Amino acid biosynthetic classes
Polar requirement classes
Coevolution theory of code origin
Four-column theory of code origin

ABSTRACT

I analyze the mechanism on which are based the majority of theories that put to the center of the origin of the genetic code the physico-chemical properties of amino acids. As this mechanism is based on excessive mutational steps, I conclude that it could not have been operative or if operative it would not have allowed a full realization of predictions of these theories, because this mechanism contained, evidently, a high indeterminacy. I make that disapproving the four-column theory of the origin of the genetic code (Higgs, 2009) and reply to the criticism that was directed towards the coevolution theory of the origin of the genetic code. In this context, I suggest a new hypothesis that clarifies the mechanism by which the domains of codons of the precursor amino acids would have evolved, as predicted by the coevolution theory. This mechanism would have used particular elongation factors that would have constrained the evolution of all amino acids belonging to a given biosynthetic family to the progenitor pre-tRNA, that for first recognized, the first codons that evolved in a certain codon domain of a determined precursor amino acid. This happened because the elongation factors recognized two characteristics of the progenitor pre-tRNAs of precursor amino acids, which prevented the elongation factors from recognizing the pre-tRNAs belonging to biosynthetic families of different precursor amino acids. Finally, I analyze by means of Fisher's exact test, the distribution, within the genetic code, of the biosynthetic classes of amino acids and the ones of polarity values of amino acids. This analysis would seem to support the biosynthetic classes of amino acids over the ones of polarity values, as the main factor that led to the structuring of the genetic code, with the physico-chemical properties of amino acids playing only a subsidiary role in this evolution. As a whole, the full analysis brings to the conclusion that the coevolution theory of the origin of the genetic code would be a theory highly corroborated.

1. Introduction

The majority of theories advanced for explaining the origin of the genetic code maintain that the physico-chemical properties of amino acids had a fundamental role to organize the structuring of the genetic code (Woese, 1965, 1967; Woese et al., 1966a, 1966b; Epstein 1966; Sonneborn, 1965; Goldberg and Wittes, 1966; Fitch and Upper, 1987; Lacey Jr. et al., 1992; Szathmary 1993). In other words, according to these theories the physico-chemical properties of amino acids were the crucial and not subsidiary element that permitted to the genetic code to be organized. Therefore,

---

☆ In memory of Giancarlo Cammarota.

E-mail address: massimo.digiulio@ibbr.cnr.it

http://dx.doi.org/10.1016/j.jtbi.2016.04.005
0022-5193/

according to the physico-chemical postulates, studies tending to ascertain the optimization level reached from the physico-chemical properties of amino acids, during the evolution of the genetic code, would have due to show that this level would have been very high, but this does not seem to have been the case (Wong, 1980; Di Giulio, 1989a, 2001, 1989b, 1996, 1997, 1998; Massey, 2008, 2010, 2015; Stoltzfus and Yampolsky, 2007; Di Giulio and Medugno, 1999, 2008, 2001; Di Giulio et al., 1994). Effectively, we are in presence, for example, of a level of minimization of the polarity distances of amino acids that would seem to be compatible with a subsidiary and not basic role played by these distances, as instead the physico-chemical postulates should impose (Di Giulio, 1996, 1989a, 1989b; Di Giulio and Medugno, 1999, 2001; Di Giulio et al., 1994). As an extreme example tending to show that the physico-chemical properties of amino acids would have played only a subsidiary role in organizing the genetic code, we can quote the Massey's works who has shown that a completely neutral process might explain the allocations of amino acids in the genetic code (Massey, 2008, 2010, 2015). In other words, that the optimization of the organization of the genetic code is a emergent property, in which the natural selection would have played only a trivial role (Massey, 2008, 2010, 2015). Therefore, these studies (Wong, 1980; Massey, 2008, 2010, 2015; Stoltzfus and Yampolsky, 2007; Di Giulio, 1989a, 1989b, 1996, 1997, 1998, 2001; Di Giulio and Medugno, 1998, 2001, 2008; Di Giulio et al., 1994) do not favor theories that place the physico-chemical properties of amino acids to the center of the evolution of the genetic code, but are perfectly compatible with the coevolution theory of the origin of the genetic code (Wong, 1980; Di Giulio, 1989a, 1989b, 1996, 1997, 1998, 2001, 2008a; Di Giulio and Medugno, 2001; Di Giulio et al., 1994). This theory sustains, instead, that the physico-chemical properties of amino acids played only a subsidiary role in organizing the code (Wong, 1980) – and important only if understood as manifestation of the catalysis performed by proteins (Di Giulio, 1996, 2003, 2008b) – and that were, above all, the biosynthetic relationships between amino acids to structure really the genetic code (Wong, 1975, 1976, 2005; Di Giulio, 2008a).

Although there are several authors that consider the physico-chemical properties of amino acids, that is to say the optimization of these, as the crucial factor that determined the organization of the code (Woese, 1965, 1967; Woese et al., 1966a, 1966b; Epstein 1966; Sonneborn, 1965; Goldberg and Wittes, 1966; Fitch and Upper, 1987; Lacey Jr. et al., 1992; Szathmary, 1993; Freeland and Hurst, 1998; Freeland et al., 2000), I feel on the contrary that this is not the case. In particular, I have now reached the certainty that the mechanism on which lie on the majority of theories based on the physico-chemical properties of amino acids is not credible or at least not satisfactory. Therefore, I here attempt to analize its evolutionary plausibility. I conclude that this mechanism is at least forced if not completely unacceptable. On the contrary the coevolution theory turns out to be extremely simple and natural. In addition, I here suggest a new hypothesis regarding the codon domain formation of precursor amino acids that it is not included in the original formulation of the coevolution theory. I make that disapproving severely the four-column theory of the origin of the genetic code (Higgs, 2009) and reply to the criticism that was directed towards the coevolution theory. Furthermore, I analize, by means of the Fisher's exact test (Di Giulio, 2008a), the distribution in the genetic code of the physico-chemical properties of amino acids and their biosynthetic relationships, because this analysis might show – at least in the comparative sense – which of these two characteristics of amino acids has mainly contributed to the structuring of the genetic code. I conclude that the biosynthetic relationships between amino acids have played a crucial role in organizing the genetic code, since the physico-chemical properties of amino acids show an ambiguous distribution in the code compared to the clear one of their biosynthetic relationships, and for several other reasons that I try to discuss as follows.

2. Materials and methods

The amino acids have been grouped in four classes on the base of their polarity values (Woese et al., 1966a; Di Giulio and Medugno, 1998), following the logic of cluster analysis. The four classes are

Class I: Cys (4.8), Leu (4.9), Ile (4.9), Phe (5.0), Trp (5.2), Met (5.3), Tyr (5.4), Val (5.6).

Class II: Thr (6.6), Pro (6.6), Ala (7.0), Ser (7.5), Gly (7.9).

Class III: His (8.4), Gln(8.6), Arg (9.1), Asn (10.0), Lys (10.1).

Class IV: Asp (12.5), Glu (13.0).

During the analysis I have also considered – in the case in which the significance in the statistical test was favored – that Val is present also in the class II and Gly in the class III.

The biosynthetic families have been identified following Wong (1975), Taylor and Coates (1989), Di Giulio (2008a), Di Giulio and Amato (2009), and Di Giulio and Medugno (1999). They are as follows:

Serine family: Ser, Gly, Cys, Trp.

Glutamate family: Glu, Gln, His, Pro, Arg.

Phosphoenolpyruvate family: Phe, Tyr.

Aspartate family: Asp, Asn, Lys, Thr, Ile, Met.

Pyruvate family: Ala Val, Leu.

It has also been considered the family composed by Ala, Val, Asp, Glu, Gly, and Ser on the base of observation that these amino acids do not have amino acid precursors (Di Giulio, 2008a).

The use of Fisher's exact test to calculate the probability values on how the amino acids are distributed in the genetic code, has been suggested for the first time by Rob Knight (Di Giulio, 2008a). For instance, considering the biosynthetic class of the amino acids that first evolved along the path of the energetic metabolism (Ala, Val, Asp, Glu, Gly, and Ser) (Di Giulio, 2008a), that is to say, for the amino acids of the last row of the genetic code (GNN codons), we obtain the following. It is necessary to calculate the probability with which the amino acids – Ser, Gly, Ala, Val, Asp and Glu – can be observed in the GNN codons of the genetic code while also taking into account the distribution of the amino acids in the non-GNN codons. Fisher's exact test seems to be able to calculate this probability (Di Giulio, 2008a, 2008b). If we consider that, of these 6 amino acids, only Ser is not codified by GNN type codons, we obtain for these amino acids with non-amino acid precursors: (i) 5 of these are codified by GNN codons (=a), while (ii) only 1 (Ser) is codified by non-GNN codons (=b). For amino acids with amino acid precursors, we have: (i) 0 of these are codified by GNN codons (=c), and (ii) 14 of these are codified by non-GNN codons (=d). By applying Fisher's exact test we obtain a probability $P = 3.9 \times 10^{-4}$ (a=5, b=1, c=0, d=14) which is highly significant (Di Giulio, 2008a).

In order to join together several P values, we can calculate the $-2\ln P$ quantities and to use another Fisher's method [38] that establishes that the – 2lnP quantity is distributed according to the χ^2 law with two degrees of freedom [38]. We then can sum these quantities ($-2\ln P$), and from the total χ^2 to establish the probability of the aggregate (Fisher, 1950; Di Giulio, 1991).

2.1. The distribution in the genetic code of classes of biosynthetic relationships between amino acids and of classes of polarity values

Higgs (2009) states that: " the layout in Fig. 6 seems pretty random with respect to the GNN code, and there seems no particular reason why the earliest amino acids should have expanded