

The Origin of the Genetic Code: Matter of Metabolism or Physicochemical Determinism?

Massimo Di Giulio

Received: 14 October 2013 / Accepted: 18 October 2013 / Published online: 26 October 2013
© Springer Science+Business Media New York 2013

In the research field on the origin of the genetic code two lines of thought are evident. The first puts at the core of the origin of the genetic code the physicochemical properties of amino acids (1967; Woese et al. 1966; Jungck 1978; Weber and Lacey 1978; Lacey and Mullins 1983; Di Giulio 1989a, b; Taylor and Coates 1989; Di Giulio 1996; Freeland and Hurst 1998; Lacey et al. 1992). Indeed, the physicochemical theory of the origin of the genetic code suggests that these properties have been the main selective pressure that determined the organisation of the genetic code (Sonneborn 1965; Woese et al. 1966; Fitch and Upper 1987; Di Giulio 1997). A prediction in part equivalent is also made by the stereochemical theory of the origin of the genetic code that maintains that interactions between codons or anticodons and amino acids have been the main forces that organised the genetic code (Woese 1967; Shimizu 1982; Szathmáry 1993; Yarus 1998; Yarus et al. 2009). Therefore, also for the stereochemical hypothesis the physicochemical properties of amino acids were essential in organising the code. On the other hand, the second line of thought considers the metabolism and the biosynthetic relationships between amino acids as the main force that defined the evolutionary organisation of the genetic code (Wong 1975, 2005; Di Giulio 2008a), with the

properties of amino acids having only a subsidiary role in this organisation (Wong 1980; Di Giulio 1989a).

The current special issue of the *Journal of Molecular Evolution*, dedicated to the Evolution of the Genetic Code, presents four papers that summarise this field of study at the moment and bring some insight into these two lines of thought. Naturally, whether the physicochemical properties of amino acids would reflect the structure of proteins, that is to say their function, then this would not be incompatible with the coevolution theory of the genetic code, because the protein structure, by means of the physicochemical properties of amino acids, might have been the main selective pressure that determined the genetic code organisation. Obviously, this would also imply a correlation between the biosynthetic relationships of amino acids and their physicochemical properties, in addition to that with the structure of the genetic code, as some works seem to indicate (Di Giulio 1991, 1996). One of the works presented in this special issue by Francis (2013) is inclined to maintain a similar position. Indeed, the title of this paper is: ‘Evolution of the genetic code by incorporation of amino acids that improved or changed protein function’. Furthermore, Francis (2013) considers the hypothesis that very early on the genetic code used only codons of the GNC kind (Ikehara 2002; Trifonov 2004; Francis 2011), a code completely compatible also with the coevolution theory (Di Giulio 2008a).

However, the observation that the properties of amino acids are mirrored in the genetic code organisation does not necessarily imply their determining role in the genetic code organisation, unless one does not validate that the level reached by the minimisation of physicochemical distances between amino acids in the genetic code structure is very elevated. This, at the moment, does not seem to be the case (Wong 1980; Di Giulio 1989a; Di Giulio et al. 1994; Di

M. Di Giulio (✉)
Laboratory for Molecular Evolution, Institute of Genetics and Biophysics ‘Adriano Buzzati Traverso’, CNR, Via P. Castellino, 111, 80131 Naples, Napoli, Italy
e-mail: massimo.digiulio@igb.cnr.it; digiulio@igb.cnr.it

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

Giulio and Medugno 1998, 2001). A deepened understanding of the physicochemical similarity between amino acids, as investigated by Stephenson and Freeland (2013) in this special issue, could help to better clarify the role played by the physicochemical properties of amino acids during the evolution of the genetic code. For this reason, another work presented in this special issue by Buhrman et al. (2013) shows that the genetic code is optimal only when we consider in addition to the polarity distances between amino acids, and some stereochemical constraints, also their biosynthetic relationships. This result would seem to imply that these three forces—the polarity distances, some stereochemical constraints and the biosynthetic relationships between amino acids—were important to organise the genetic code. In general, although there are models that make compatible the coevolution theory and the stereochemical one (Di Giulio 1998), nevertheless I find that the strong physicochemical determinism imposed by the stereochemical theory to the origin of the genetic code makes these two theories incompatible. This is because one pertains to a strong physicochemical determinism, while the other concerns a determinism of historical nature linked to the origin of the biosynthetic relationships between amino acids. Furthermore, the stereochemical constraints used by Buhrman et al. (2013) concern only the Phe, Tyr, Trp, His, Leu, Ile and Arg amino acids that are considered to be product amino acids (that is to say, derived biosynthetically from other amino acids), and therefore introduced very late in the genetic code, not only by the coevolution theory (Wong 1975; Di Giulio 2008a) but also by other points of view that consider these amino acids as late addition to the genetic code (Trifonov 2000; Higgs and Pudritz 2009). This would seem to imply that the stereochemical constraints were important for amino acids that entered into the genetic code only late, and this seems to be a paradox. Indeed, one expects that stereochemical constraints were most important in the early phases of the origin of the genetic code and not in the late ones given that in the early phases the entire ‘system’ might have been more susceptible to stereochemical interactions than in the late phases, in which the system might already have assumed greater biological features. In addition, if the stereochemical constraint was really important to organise the genetic code then this should have been important above all for the precursor amino acids for which the historical (biosynthetic) constraint—according to the coevolution theory (Wong 1975)—should have been weak or even absent, and therefore the precursor amino acids should have been more susceptible to the stereochemical constraint. This does not turn out from the model used by Buhrman et al. (2013), and therefore their observations may require further clarification in the future.

Therefore, I continue to maintain that more time should be spent in trying to understand how and why the

biosynthetic relationships between amino acids are reflected in the organisation of the genetic code. Additionally, it will be important to try to find other molecular fossils that extend the biosynthetic pathways of amino acids that occur on tRNAs—and that are a very strong corroboration of the coevolution theory (Wong 1988; Di Giulio 2002)—also to biosynthetic transformations involving other amino acids, as for instance, might be the Ser-tRNA^{Trp} → Trp-tRNA^{Trp} pathway that I think might be found in some organism.

The last article presented in the special issue is the one by Morgens (2013) that makes a review of the origin of translation. I like very much to think that the evolution of the genetic code has been guided by the selective pressure of the enzymatic catalysis, resulting in the origin of protein synthesis first and after in that of the genetic code. Therefore, the genetic code would be nothing other than the evolution of the coded catalysis (Di Giulio 2003, 2008b). Morgens (2013), in a sense, follows a similar route, using in the title of his work the phrase: ‘the protein invasion’. Morgens bases his analysis above all on the complexity of protein synthesis considered notoriously to be a strong obstacle to its evolution (Crick et al. 1976). Yet, models that predict that the ancestral catalysis was mediated by RNA in some way peptidated—that is to say, catalytic forms also able to explain the origin of peptidyl-tRNA, the key intermediary of the protein synthesis otherwise barely under the action of natural selection (Wong 1991; Di Giulio 2003, 2008b)—seem to me to be able to remove completely the problem of the non-reducibility of the origin of protein synthesis and with it also that of the genetic code origin. This seems to generate a natural view of the origin both of protein synthesis and of the genetic code (Di Giulio 2008b).

Finally, I hope that the reading of articles presented here encourages and stimulates in some way the study of these very fascinating set of problems.

References

- Buhrman H, van der Gulik PTS, Klau GW, Schaffner C, Speijer D, Stougie L (2013) Realistic model under which the genetic code is optimal. *J Mol Evol* 77:168–182
- Crick FHC, Brenner S, Klug A, Piecznik G (1976) A speculation on the origin of protein synthesis. *Orig Life* 7:389–397
- Di Giulio M (1989a) 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 (1989b) Some aspects of the organization and evolution of the genetic code. *J Mol Evol* 29:191–201
- Di Giulio M (1991) On the relationships between the genetic code coevolution hypothesis and the physicochemical hypothesis. *Z Naturforsch C* 46:305–312
- Di Giulio M (1996) The β -sheets of proteins, the biosynthetic relationships between amino acids, and the origin of the genetic code. *Orig Life Evol Biosph* 26:589–609

- Di Giulio M (1997) On the origin of the genetic code. *J Theor Biol* 187:573–581
- Di Giulio M (1998) Reflections on the origin of the genetic code: a hypothesis. *J Theor Biol* 191:191–196
- Di Giulio M (2002) Genetic code origin: are the pathways of type Glu-tRNA(Gln) \rightarrow Gln-tRNA(Gln) molecular fossils or not? *J Mol Evol* 55:616–622
- Di Giulio M (2003) The early phases of genetic code origin: conjectures on the evolution of coded catalysis. *Orig Life Evol Biosph* 33:479–489
- Di Giulio M (2008a) An extension of the coevolution theory of the origin of the genetic code. *Biol Direct* 3:37
- Di Giulio M (2008b) Why the genetic code originated. Implications for the origin of protein synthesis. The codes of life: the rules of macroevolution (Chap. 4). Springer, Dordrecht, pp 59–67
- Di Giulio M, Medugno M (1998) The historical factor: the biosynthetic relationships between amino acids and their physicochemical properties in the origin of the genetic code. *J Mol Evol* 46:615–621
- Di Giulio M, Medugno M (2001) The level and landscape of optimization in the origin of the genetic code. *J Mol Evol* 52:372–382
- Di Giulio M, Capobianco MR, Medugno M (1994) On the optimization of the physicochemical distances between amino acids in the evolution of the genetic code. *J Theor Biol* 186:43–51
- Fitch W, Upper K (1987) The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. *Cold Spring Harbor Symp Quant Biol* 52:759–767
- Francis BR (2011) An alternative to the RNA world hypothesis. *Trends Evol Biol* 3:e2
- Francis BR (2013) Evolution of the genetic code by incorporation that improved or changed protein function. *J Mol Evol* 77:132–156
- Freeland SJ, Hurst LD (1998) The genetic code is one in a million. *J Mol Evol* 47:238–248
- Higgs PG, Pudritz RE (2009) A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. *Astrobiology* 9:483–490
- Ikehara K (2002) Origins of gene, genetic code, protein and life: comprehensive view of life systems from a GNC–SNS primitive genetic code hypothesis. *J Biosci* 27:165–186
- Jungck JR (1978) The genetic code as a periodic table. *J Mol Evol* 11:211–224
- Lacey JC Jr, Mullins DW Jr (1983) Experimental studies related to the origin of the genetic code and the process of protein synthesis—a review. *Orig Life* 13:3–42
- Lacey JC Jr, Wickramasinghe NSMD, Cook GW (1992) Experimental studies on the origin of the genetic code and the process of protein synthesis: a review update. *Orig Life Evol Biosph* 22:243–275
- Morgens DW (2013) The protein invasion: a broad review on the origin of the translational system. *J Mol Evol* 77:183–194
- Shimizu M (1982) Molecular basis for the genetic code. *J Mol Evol* 18:297–303
- Sonneborn TM (1965) Degeneracy of the genetic code: extent, nature, and genetic implications. In: Bryson V, Vogel HJ (eds) *Evolving genes and proteins*. Academic Press, New York, pp 377–397
- Stephenson JD, Freeland SJ (2013) Unearthing the root of amino acid similarity. *J Mol Evol* 77:157–167
- Szathmáry E (1993) Coding coenzyme handles: a hypothesis for the origin of the genetic code. *Proc Natl Acad Sci USA* 90:9916–9920
- Taylor FJR, Coates D (1989) The code within the codons. *BioSystems* 22:177–187
- Trifonov EN (2000) Consensus temporal order of amino acids and evolution of the triplet code. *Gene* 261:139–151
- Trifonov EN (2004) The triplet code from first principles. *J Biomol Struct Dyn* 22:1–11
- Weber AL, Lacey JC Jr (1978) Genetic code correlations: amino acids and their anticodon nucleotides. *J Mol Evol* 11:199–210
- Woese CR (1967) *The genetic code*. Harper & Row, New York
- Woese CR, Dugre DH, Dugre SA, Kondo M, Saxinger WC (1966) On the fundamental nature and evolution of the genetic code. *Cold Spring Harbor Symp Quant Biol* 31:723–736
- Wong JT (1975) A co-evolution theory of the genetic code. *Proc Natl Acad Sci USA* 72:1909–1912
- Wong JT (1980) Role of minimization of chemical distances between amino acids in the evolution of the genetic code. *Proc Natl Acad Sci USA* 77:1083–1086
- Wong JT (1988) Evolution of the genetic code. *Microbiol Sci* 5:174–182
- Wong JT (1991) Origin of genetically encoded protein synthesis a model based on selection for RNA peptidation. *Orig Life Evol Biosph* 21:165–176
- Wong JT (2005) Coevolution theory of the genetic code at age thirty. *BioEssays* 27:416–425
- Yarus M (1998) Amino acids as RNA ligands: a direct-RNA-template theory for the code's origin. *J Mol Evol* 47:109–117
- Yarus M, Widmann JJ, Knight R (2009) RNA-amino acid binding: a stereochemical era for the genetic code. *J Mol Evol* 69:406–429