

Codon Assignment to Amino Acids in the Genetic Code Table

Suvendra Kumar Ray*

Department of Molecular Biology and Biotechnology, Tezpur University, Tezpur, Assam, 784028, India

***Corresponding Author:** Suvendra Kumar Ray,
Department of Molecular Biology and Biotechnology,
Tezpur University, Tezpur, Assam, 784028, India; Tel:
00913712275406, Fax: 00913712267005; E-mail:
suven@tezu.ernet.in

I feel privileged to get an invitation to write an article for *Aperito Journal of Computer Science and Biology*. I believe that the question on the origin of the genetic code has come across to the mind of all scientists who understand the central dogma of molecular biology. Computational analysis of huge genomic sequences is essential to understand and to answer some of the exciting questions on molecular evolution. There are a large number of scientists carrying research in biology using the knowledge of computer science. It is good for the journal to put forward an exciting biological question that needs to be addressed by the approach of computer science. With this objective I put forward my view on the origin of the genetic code with an objective to stimulate researcher in this interdisciplinary area to work on the problem. As a member of the editorial board of this journal I also welcome researchers to publish their ideas and views in this journal.

The standard genetic code table has 64 triplet codons out of which 61 codons have been assigned to 20 amino acids [1, 2]. The codons have been assigned non-randomly that is evident from the synonymous codons, codon encoding the same amino acid, are being placed close to each other in the table. In addition, the two degeneracy codons are either ending with Y or

R always. Again in the table amino acids along the column are more similar in their physical and chemical properties than across the table.

In spite of the above known features of the genetic code, the main question that remains still unanswered to scientists is how the codons were assigned to amino acids in the beginning? There are several hypothesis have been proposed earlier in this regard. All these hypotheses have tried to answer by connecting codons with amino acids directly. So it becomes difficult for the reader to imagine the relationship. We strongly believe that amino acyl tRNA synthetase (aaRS), the enzyme that attaches amino acids to tRNA, is the main decider for the codon assignment. The enzyme had evolved prior to the code. This is able to explain the non-random codon distribution in the genetic code and the recent evolution of the genetic code in some organisms and organelles [3, 4].

It is the anticodon of the tRNA that recognizes codon on the mRNA inside the ribosome and incorporates the amino acid in the nascent polypeptide during translation. So the amino acid attached to the tRNA is assigned to the codon to which the tRNA recognizes. E.g. amino acid aa1 is attached to a tRNA whose anticodon recognizes the codon1 and therefore codon1 is assigned to aa1. If there is change in anticodon sequence in the tRNA that no longer recognize codon1 but recognizes codon2, and if the tRNA with the changed anticodon can still be charged with aa1, then codon2 is assigned to aa1 not the codon1. So the charging of tRNA with amino acid and its anticodon is the key factor for codon assignment. The best example of change in codon assignment is observed in case of suppressor tRNA that

recognizes non-sense codon in suppressor mutants and assigned an amino acid.

The enzyme that catalyzes the charging of tRNA with amino acid is known as amino acyl tRNA synthetase (aaRS). In a cell there is usually 20 different aaRS are present for attaching 20 amino acids with specific tRNAs. While amino acid binding site is fixed in aaRS, there is some flexibility in binding to tRNA. E.g. trpRS attaches tryptophan only to the trp-tRNA in cell. But if there is mutation in the anticodon of trp-tRNA, still the enzyme attached tryptophan to the tRNA. This example we observed in case of suppressor tRNA that we have mentioned before. In other example of stating flexibility of aaRS tRNA binding cite is that leuRS attache's leucine to minimum three different leu-tRNAs present inside the cell. Similarly serRS can bind to three different ser-tRNA and catalyze attachment of ser to ser-tRNA. So aaRS can bind to tRNAs having difference in their anicodons. aaRS does not recognize tRNA from its anticodon rather it interacts at multiple points in a tRNA to recognize it. Had it recognize anticodon, and then there would have more number of aaRS in cell.

So aaRS charges tRNA with an amino acid and therefore the aaRS is most likely the deciding factor for codon assignment. aaRS has two sites; one strictly binds to a specific amino acid and the other with more flexibility that binds to tRNA. Considering Higgs' hypothesis on the four column theory of the origin of the genetic code [4], in the beginning there may be four/five aaRS for five amino acids that each could charge only one amino acid out of the five amino acids such as Gly, Ala, Asp/Glu, and Val. Initially all the sixty four codons were divided to these four/five amino acids only. So in the beginning code was much more degenerative than in the present code. At the early stage, there might be less number of tRNA genes, wobble pairing both at the 1st and at the 3rd position of the codon where the middle one was following the Watson-Crick base pairing and codons for Asp and Glu were same. Later duplication of the aaRS genes followed by mutation both at the amino acid binding site and at the tRNA binding site, resulted into 20 types of aaRS. Similarly duplication of tRNA

genes and mutation results in evolution of different types of tRNA. Latter the loss of the wobble pairing at the 1st codon position and restriction of wobble pairing at the 3rd codon position might have been evolved after the evolution of tRNA modification enzymes [5-9]. This resulted into the assignment of more number of amino acids in the genetic code table as we see today. As the code started working well in living organisms, any change could result lethality, so it maintained as almost universal.

References

1. Crick FHC, 1968, The origin of the genetic code, *J. Mol. Biol.* **38**, 367-379.
2. Hartman H, Smith TF. 2014, The evolution of the ribosome and the genetic code, *Life*, **4**
3. Caetano-Anollés G, Wang M, Caetano-Anollés D, 2013, Structural phylogenomics retrodicts the origin of the genetic code and uncovers the evolutionary impact of protein flexibility, *PLoS ONE* **8**, e72225.
4. Higgs PG, 2009, A four column theory for the origin of the genetic code: tracing the evolutionary pathways that gave rise to an optimized code, *Biology Direct*, **4**, 16.
5. Koonin EV and Novozhilov AS, 2009, Origin and evolution of the genetic code: the universal enigma, *IUBMB Life*, **61**, 91-111.
6. Osawa S, Jukes TH, Watanabe K, Muto A, 1992, Recent evidence for evolution of the genetic code, *Microbiol Rev.*, **56**, 229.
7. Rodnin AS, Szathmary E, Rodin SS, 2011, On origin of genetic code and tRNA before translation, *Biology Direct*, **6**:14.
8. Wong JT-F, 2005, Coevolution theory of genetic code at age thirty. *Bioessays*, **27**, 416-325.
9. Wolf YI, Koonin EV, 2007, On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization, *Biology direct*, **2**, 14.