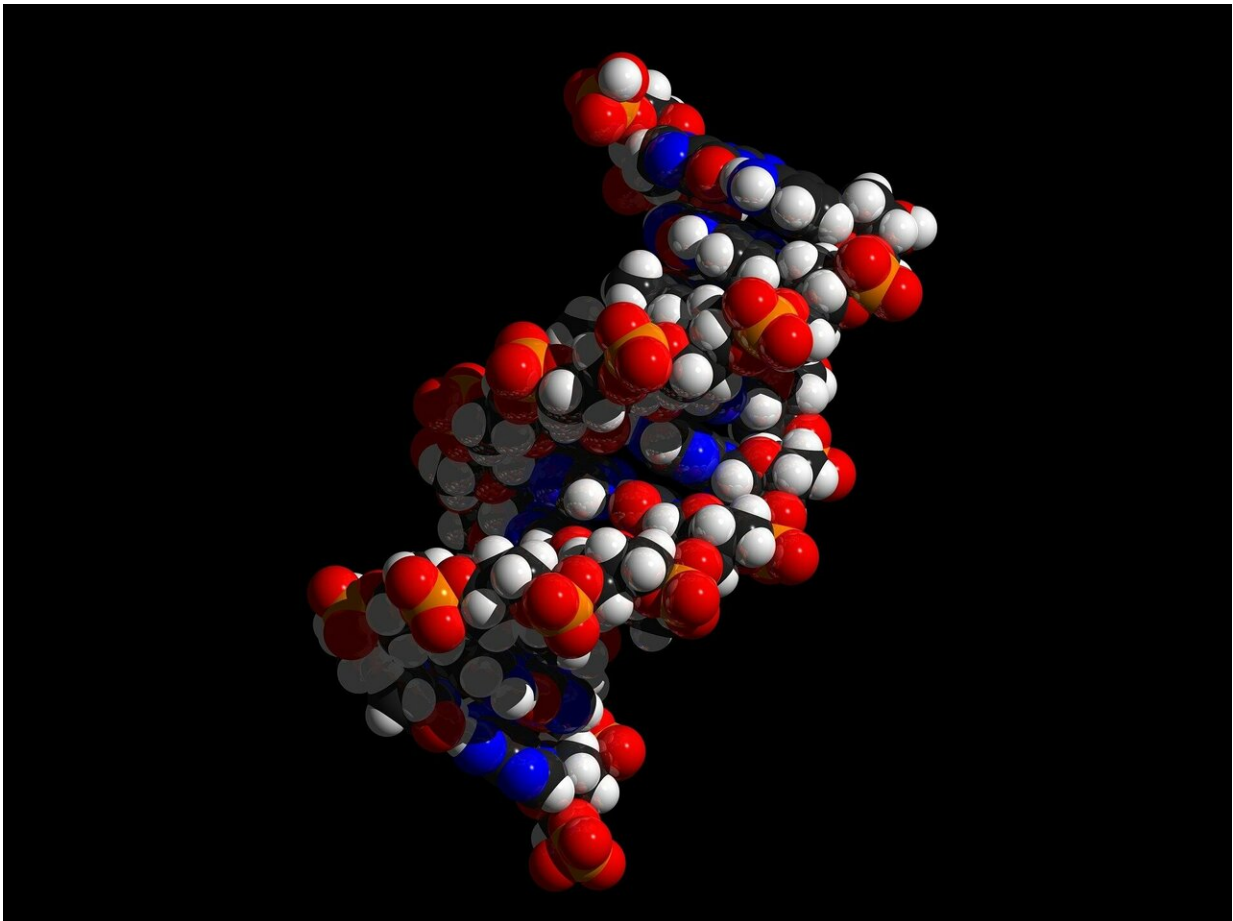# Linguistics may help us to understand some 'strangeness' of the genetic code

September 5 2023

Credit: CC0 Public Domain

Linguists have developed the comparison of the genetic code with

language where nucleotides act as letters, and introduced the concept of "a semiotic nucleotide"—the minimal element that makes it possible to distinguish between codons—coding units of DNA. According to this approach, the biochemical characteristics of DNA operate as informational ones.

Flexibility of informational approach enables researchers to highlight facts that are not explained by biochemical features, and usually they are considered as deviations from the universal regularities of the genetic code. The research is published in the journal *Biosystems*.

The genetic code has dual characteristics: It contains not only biochemical properties, but it also has a semiotic or semantic dimension. Semiotics is a science that studies general regularities of processing of information through signs. Researchers find analogies between a text and the genetic code, for example, in the fact that genes carry a program of organism development, and that program resembles texts written according to some rules.

The semiotic theory allows consideration of nucleotides not as biological molecules but as information carriers. The crucial genetic processes can be described from the viewpoint of operations with text: reading, transcription, translation, proofreading, editing.

Researchers from Immanuel Kant Baltic Federal University and Institute of Scientific Information on Social Sciences of the Russian Academy of Sciences paid attention to the fact that the same nucleotide in DNA according to its position has a different value in genetic information processing.

Thus, when proteins are synthesized in a cell according to "recipe," written in genes, special cell "machines"—ribosomes—read out nucleotides three by three, and for each such triplet, called codon, they

choose a particular amino acid. In 32 cases from 64 possible combinations of nucleotides "A," "T," "G," and "C," the third position can be occupied by any of them and this does not affect on the result—the recognized amino acid. It happens because the same amino acid can be encoded by several different triplets of nucleotides.

As a result, in order to understand which amino acid is needed, a ribosome, while reading out each letter, focuses first of all on the "meaning" of its combination within triplets. This is called wobbling, due to the "wobbling" position of the last nucleotide in codons. In order to describe it from the point of view of data transmission, the linguists introduced the term "a semiotic nucleotide"—a minimal element that enables them to distinguish one triplet of nucleotides from another one.

In this connection, instead of comparing nucleotides with letters as is usually done, the scientists suggested correlating them with other language entities—sounds, or more precisely, phonemes (a language element that includes only those features that are necessary for distinguishing of signs). A letter is not a language unit; it only serves to designate sound in writing.

The analogy with phonemes enables an explanation of how two distinctive features of a nucleotide correlate with their changeable significance depending on nucleotide's position within a codon.

This presupposes that minimal units of the genetic code are not nucleotides but their distinctive features. These features have different relevance according to their position within a triplet—maximal in the second position and minimal, down to zero, in the third. A nucleotide in the third position is present in the physical sense but it may be absent in semiotic sense (from the point of view of its distinctive value).

Each nucleotide has two distinctive features: the number of hydrogen

bonds (two or three) and carbon rings (one or two). These features are relevant for binding nucleotides with each other. Thus, nucleotides that have two rings correspond to those that have one ring (and vice versa) but with the same number of hydrogen bonds. However, this regularity can be transformed as far as the third position in a codon is concerned.

"Use of semiotic approach makes it possible to identify which role each of nucleotides play for distinguishing of codons, and considering wobbling as a special reading mode. As a product of evolution, the genetic code is semiotically heterogeneous—in the half of codons (32) the third position is irrelevant, in thirty cases it acts in its half-strength (only one feature, the number of rings is relevant); and only in the case of the tryptophan the two features participate equally."

"Informational-semiotic approach enables [us] to complement the common description of the genetic code. Early Francis Crick, speaking of deviations from regularities of the genetic code connected with the third position, called them 'out of the obvious sense.' However, from the point of view of semiotics, the special position of nucleotide may have a meaningful explanation, as its primary function is to separate one codon from another, and only the second function is to distinguish between codons," says Suren Zolyan, Doctor of Philological Sciences, Professor, chief researcher of Immanuel Kant Baltic Federal University, Institute for the Humanities.