BioEssays
www.bioessays-journal.com

AUG as the Translation Start Codon in Circular RNA Molecules: A Connection between Protein-Coding Genes and Transfer RNAs?

Paweł Mackiewicz

The AUG codon is a classical signal that tells the ribosome where protein synthesis (translation) should start. Despite in-depth knowledge about translation initiation, the reason for the evolutionary assignment of this role to AUG is not clear. It was proposed that high metabolic costs of methionine synthesis, the amino acid coded by AUG, allow for the precise regulation of translation initiation.[1] Alternatively, this codon, together with an adenine occurring just after it, can easily generate the stop translation codon UGA in +1 frameshift, which can immediately terminate aberrant protein synthesis or adjust the ribosome to the appropriate frame.[2] However, that proposed mechanisms could have evolved later, when the AUG codon already existed, as a result of other selection pressures.

An interesting hypothesis, which provides a valuable insight into this subject, has been recently described by Demongeot and Seligmann in this issue.[3] Using simulation methods, the authors claim that short 22-nucleotide-long circular RNAs, termed "RNA rings," can easily acquire AUG as the start codon if they adopt a stem-loop hairpin structure and the number of coded amino acid is maximized when these RNA sequences are read in three frames. The first assumption is reasonable, because such a structure could protect the molecules against features of the prebiotic environment that damage macromolecules. The second stipulation also seems sensible, because the diversity of amino acids in primordial peptides was surely positively selected for enhanced functionality. Under these assumptions, 10 out of 25 generated RNA rings started with AUG, and all rings gained a stop codon at their ends.

This concept is especially noteworthy because it does not postulate mechanisms directly related to translation initiation. Interestingly, the RNA rings that are generated share properties with protein-coding genes. The RNA sequences prefer 20 codons that are overrepresented in coding genes and enable detection of the ribosomal translation frame. The codon content is also biased toward codons encoding amino acids that are attached to appropriate tRNAs by aminoacyl-tRNA synthetases of class II—considered to be ancient molecules. In this sense, the RNA rings are reminiscent of early stages in the origin and evolution of life. It should be emphasized that they may also be good candidates for proto-tRNAs because they bear similarities to ancestral tRNA loops. Furthermore, the circular RNAs might also play a role in originating replication, because stem-loop RNA structures possessing an anticodon resemble replication origin loops. If this feature were confirmed experimentally, it would suggest that these molecules were originally able to auto-replicate and propagate, an essential property for their evolution.

The findings that the theoretical RNA rings could be both short protein-coding sequences and primordial tRNAs find strong support in the form of modern RNA molecules called transfer-messenger RNA (tmRNA). These are used as templates for continuation of translation when mRNA transcripts are broken. Additionally, the RNA ring structure is a well-known motif in some mitochondrial tRNAs, whose sequences contain conserved stop codons between two stems and codon AUG, that is, between the hypervariable loop and a sidearm. The region encompassed by these codons seems capable of cyclization, and could code for a peptide, just like the RNA rings obtained by the authors.[3]

The presented hypothesis assumes that RNA molecules encoded potential peptides in three overlapping reading frames, which maximized information capacity in one stretch of RNA, as in present-day viral genomes. This feature is well established in properties of the standard genetic code, which shows a tendency to minimize effects of frameshift mutations to a similar extent as point mutations.[4] It is conceivable that frameshift mutations were quite frequent at the early stages of life on Earth, but the mechanism of transformation from frameshifting to one-frame translation is still unclear. The hypothesis does not explain whether precise frame reading would still support AUG as a start codon. It would be worth checking this. Moreover, the model claims that all 20 amino acids could be used in protein coding. However, only a small repertoire of amino acids produced by simple biochemical reactions was likely available at the origin of life, and other amino acids were incorporated into the genetic code with the evolution of more complex metabolic networks.[5] It would, therefore, be interesting to verify whether RNA rings with an AUG start codon and the afore-mentioned properties are generated in a scenario where these rings encode smaller amino acid sets.

Prof. P. Mackiewicz
Department of Bioinformatics and Genomics
Faculty of Biotechnology
University of Wrocław
14a Fryderyka Joliot-Curie Street, Wrocław 50-383, Poland
E-mail: pamac@smorfland.uni.wroc.pl

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/bies.202000061

This article comments on the Hypothesis paper by Jacques Demongeot and Hervé Seligmann, https://doi.org/10.1002/bies.201900201.

DOI: 10.1002/bies.202000061

Conflict of Interest

The author declares no conflict of interest.

Received: March 20, 2020
Published online: April 27, 2020

[1] S. Bhattacharyya, U. Varshney, *RNA Biol.* **2016**, *13*, 810.
[2] L. Abrahams, L. D. Hurst, *Mol. Biol. Evol.* **2017**, *34*, 3064.
[3] J. Demongeot, H. Seligmann, *BioEssays* **2020**, *42*, 1900201.
[4] M. Wnetrzak, P. Blazej, P. Mackiewicz, *Biosystems* **2019**, *181*, 44.
[5] M. Di Giulio, *Biol. Direct* **2008**, *3*, 37.