https://doi.org/10.1038/d41586-022-04585-5

News & views

Molecular biology

No stopping with a short-stem transfer RNA

Pavel V. Baranov & John F. Atkins

Messenger RNA has 64 possible triplet sequences, or codons, three of which usually terminate protein synthesis. But some organisms can use all codons to specify amino acids, thanks in part to a surprising feature of a transfer RNA.

When a protein is made, a chain of amino-acid residues is built on the basis of messenger RNA instructions, and this construction must be brought to a stop appropriately. Writing in *Nature*, Kachale *et al.*[1] report the discovery of an unusual mechanism that breaks with the conventional way in which the mRNA decoding process occurs.

The decoding of mRNA molecules to form proteins with specific amino-acid sequences occurs through the addition of amino acids, as specified by triplets of nucleotide sequences called codons, to the growing protein chain. There are 64 possible codons, and in nearly all organisms, the termination of protein synthesis is specified by any of three codons termed stop codons — these are the triplets UAA, UAG and UGA (U, A and G stand for the nucleotide bases uracil, adenine and guanine, respectively). The 'meaning' of stop codons as signals for termination is generally independent of where they occur in the mRNA.

In 2016, it was discovered that, for a few organisms, the meaning of these codons depends on where they occur in the mRNA molecule[2–5]. In those organisms, they are recognized as termination signals only if they occur at the end of mRNA, whereas in internal positions all 64 codons encode amino acids. However, it remains unclear how such position dependence in codon recognition is accomplished. Kachale and colleagues report their discovery of a surprising mechanism involved in the decoding of the UGA codon for an organism in which all 64 codons can encode amino acids.

The authors examined *Blastocrithidia nonstop*, which belongs to the family of unicellular parasitic organisms, known as the Trypanosomatidae, that includes the causative agent of the fatal human disease sleeping sickness. Although much remains unknown about how the decoding mechanism works for organisms in which all 64 codons can specify amino acids, it is probable that the termination of protein synthesis occurs only in proximity to the protein(s) that bind the end regions of mRNAs. In other organisms, interactions with such proteins increase termination efficiency[6,7]. In *B. nonstop*, in which only UAA located near the mRNA end specifies termination, some of the changes identified by the authors in the stop-codon recognition protein eRF1 are probably partially responsible for the change in stop-codon recognition. But how are UAA, UAG and UGA recognized as coding for amino acids at internal positions in the mRNA?

On the basis of the mRNA sequence, transfer RNA (tRNA) brings amino acids to the site of the newly forming protein, and thereby acts as an adapter molecule between mRNA and proteins. On one end of its structure the tRNA carries an amino acid, and on the other end it has a specific triplet of nucleotides called the anticodon. The anticodon matches the codons in mRNA through complementary base pairing, thus enabling the recognized codon to specify a particular amino acid.

Unsurprisingly, in *B. nonstop*, UAA and UAG specification of an amino acid, rather than a stop, uses tRNAs that have matched anticodons that pair with UAA or UAG (Fig. 1). These tRNAs are not found in standard decoding

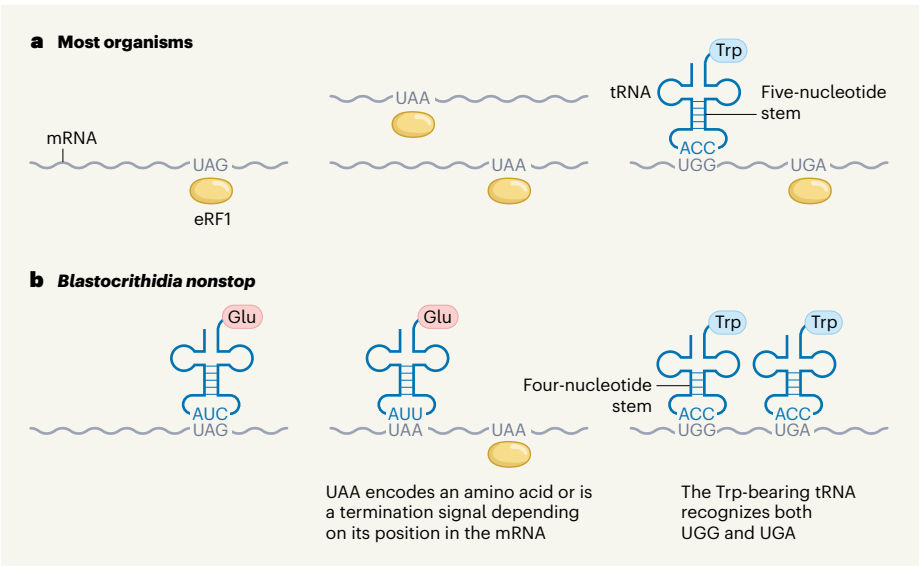

Figure 1 | An unusual system for decoding messenger RNA. a, In most organisms, out of the triplet nucleotide mRNA sequences called codons, three terminate protein synthesis. These stop codons are UAG, UAA and UGA (U, A and G represent three nucleotide bases), and, during mRNA decoding, the protein eRF1 recognizes these stop codons wherever they are positioned in the mRNA. The other 61 codons are recognized by transfer RNA (tRNA) molecules, which results in the insertion of an amino acid to a growing protein chain. For example, a tRNA that bears the amino acid tryptophan (Trp) has an 'anticodon' sequence that pairs with the codon UGG (C in the anticodon represents another nucleotide base). **b**, Kachale *et al.*[1] report that the unicellular parasite *Blastocrithidia nonstop* has adaptations that enable protein synthesis using all 64 codons in internal regions of mRNA, although the termination of synthesis can still be specified. These adaptations include tRNAs with anticodons that match UAG and UAA and which carry the amino acid glutamic acid (Glu), a termination machinery that stops protein synthesis only when UAA is near the end of the mRNA; and the stem of the Trp-bearing tRNA is a single nucleotide shorter than usual, resulting in this tRNA being able to decode both UGG and UGA.

systems, such as occur in other organisms in which UAA and UAG specify termination of protein synthesis. However, although UGA in *B. nonstop* is decoded to specify the amino acid tryptophan, no tRNA-bearing tryptophan that had a matching anticodon for UGA was found.

Instead, the tryptophan-bearing tRNA present in *B. nonstop* retains the standard anticodon for 'reading' UGG (the standard codon for tryptophan). However, Kachale and colleagues report that this tRNA also recognizes the codon UGA. This tRNA is different from other tRNAs: the 'arm' that projects the tRNA anticodon is shorter than normal by one base pair (Fig. 1). This 'short arm' feature greatly enhances UGA reading by the tRNA. Interestingly, this feature is also present in an organism called *Condylostoma*, for which it was previously unknown how UGA might specify an amino acid.

Why did such drastic changes evolve in how *B. nonstop* interprets the genetic code? Infectious agents lead to natural selection of successive host variants that are resistant to infection and can thwart an invader's countermeasures. *B. nonstop* is a host for many viruses, and Kachale *et al.* suggest that the adoption of such variant genetic decoding might have helped it in the 'arms race' against infectious invaders.

The authors also point out another possible selective force. The DNA of *B. nonstop* has a notably lower content of G- and C-containing nucleotides (C is cytosine, another nucleotide base) than do other species in the Trypanosomatidae family, suggesting that, during its evolution, *B. nonstop* gained many mutations that increased its content of A- and T-containing nucleotides (T is the base thymine). This probably resulted in amino-acid-specifying codons changing their sequences to become richer in A or T sequences. For example, G-to-A mutations in the UGG codon, which normally encodes tryptophan, would turn the codon into UGA, UAG or UAA. Making each of these three codons specify amino acids rather than prematurely signifying termination was probably advantageous.

Mutations in tRNAs at positions distant from the anticodon are known to affect various aspects of the fidelity of protein synthesis, with one in a tryptophan tRNA characterized in detail[8]. Kachale and colleagues' finding is striking, in that it provides a natural example of how a change in tRNA shape alters codon recognition without altering the anticodon itself.

Although tRNA databases contain sequences and secondary structures of many tRNA molecules, codon specificity is assigned mainly, although not entirely, on the basis of the molecules' anticodons[9]. The curious case of *B. nonstop* tryptophan tRNA reminds us of the outstanding challenges in making an accurate assignment of a tRNA's codon specificity. For example, the human genome encodes more than 400 tRNA-encoding genes and many more tRNA-like sequences[9]. For most tRNAs, their codon specificity is predicted purely computationally, and it is not known how variations in the sequences of these genes might affect their codon-decoding properties.

Part of the challenge in the correct assignment of a tRNA's codon specificity is due to the presence of numerous nucleotide modifications introduced after tRNA synthesis. Moreover, predicting modification status is not trivial, especially considering that previously unknown types of modification are continually being discovered[10]. Perhaps the eventual solution to this challenge can be facilitated by the deep-learning approaches that in the past few years have propelled an enormous leap in scientists' ability to predict protein 3D structures.

However, that deep-learning success was enabled by the availability of vast relevant experimental data. Although work over many decades has provided extensive information about the relationship between tRNA structures and codon recognition, the scope of the available 3D structures of tRNAs interacting with their codons is limited by comparison. Better availability of cryo-electron microscopy, given the increasing power of this technique to produce high-resolution atomic structural data, might eventually make a major contribution in generating the data sets required for machine learning to tackle this problem.

Pavel V. Baranov and **John F. Atkins** are in the School of Biochemistry and Cell Biology, University College Cork, Cork T12 XF62, Ireland.
e-mails: p.baranov@ucc.ie; j.atkins@ucc.ie

1. Kachale, A. *et al. Nature* https://doi.org/10.1038/s41586-022-05584-2 (2023).
2. Swart, E. C., Serra, V., Petroni, G. & Nowacki, M. *Cell* **166**, 691–702 (2016).
3. Heaphy, S. M., Mariotti, M., Gladyshev, V. N., Atkins, J. F. & Baranov, P. V. *Mol. Biol. Evol.* **33**, 2885–2889 (2016).
4. Záhonová, K., Kostygov, A. Y., Ševčíková, T., Yurchenko, V. & Eliáš, M. *Curr Biol.* **26**, 2364–2369 (2016).
5. Bachvaroff, T. R. *PLoS ONE* **14**, e0212912 (2019).
6. Wu, C., Roy, B., He, F., Yan, K. & Jacobson, A. *Cell Rep.* **33**, 108399 (2020).
7. Ivanov, A. *et al. J. Biol. Chem.* **294**, 8630–8639 (2019).
8. Schmeing, T. M., Voorhees, R. M., Kelley, A. C. & Ramakrishnan, V. *Nature Struct. Mol. Biol.* **18**, 432–436 (2011).
9. Chan, P. P. & Lowe, T. M. *Nucleic Acids Res.* **44**, D184–D189 (2016).
10. Suzuki, T. *Nature Rev. Mol. Cell Biol.* **22**, 375–392 (2021).

The authors declare no competing interests.