Expanding codon size

Engineering transfer RNAs to read codons consisting of four bases requires changes in tRNA that go beyond the anticodon sequence.

TARANA SIDDIKA, ILKA U HEINEMANN AND PATRICK O'DONOGHUE

Cells use a genetic code to translate the information contained in DNA and RNA sequences into the amino-acid building blocks that make up a protein (*Nirenberg et al., 1965*; *Söll et al., 1965*). DNA molecules are chain-like structures consisting of two entwined strands that encode information using an 'alphabet' of four nucleotide building blocks (made up of a sugar and a phosphate group, and one of four nitrogenous bases, A, C, G and T).

Some segments of the DNA genome are instructions for making proteins, and a messenger RNA molecule (mRNA) is produced from the DNA template for each protein-coding gene. On the ribosome, molecules known as transfer RNAs (tRNAs) decode the information in mRNAs by reading it in three-letter groups, also known as codons, allowing for 64 unique triplet combinations (three of which are used as stop signals). Despite this wealth of codons, most organisms usually use just 20 amino acids, making the code redundant as several codons can code for the same amino acid.

Expanding or modifying the genetic code may enable scientists to use cells as factories for making an array of molecules, which could further therapies based on proteins, and in even grander schemes, to enable the creation of artificial life forms. Recent engineering efforts have successfully produced proteins using 22 or even 23 different amino acids, rather than the usual 20 (*Wright et al., 2018*; *Tharp et al., 2021*). However, the standard triplet codons cannot be easily reassigned into new amino acids, because even though decoding the 64 codons allows for redundancy, all codons are assigned to a specific amino acid in an organism. Reassigning a codon to a new amino acid would drastically change the organisms' protein composition.

A quadruplet system based on four-letter codons rather than three has 256 total codons that could encode many more amino acids, independent of the natural triplet codon system. Expanding the number of protein building blocks would help to produce highly specialized proteins containing unnatural amino acids, potentially opening the door to advances in both basic biology and therapeutic applications (*Hohsaka et al., 1996*).

One approach to expanding the genetic code is based on a natural process called +1 frame-shifting, where four rather than three nucleotide bases are effectively decoded as a single amino acid (*Riyasaty and Atkins, 1968*). The insertion of a single base generates a frameshift in an otherwise triplet codon gene, shifting the frame by one letter. So-called frameshift suppressor tRNAs allow protein synthesis to continue past this insertion to produce a normal full-length protein. For example, some tRNAs can read frameshift mutations, including insertions or deletions of one or two nucleotides in the mRNA. A tRNA that suppresses a +1 frameshift effectively reads or decodes a quadruplet codon (*Figure 1*).

Translation using quadruplet codons exists in both natural and synthetic systems, so why did nature favor a triplet system overall? Now, in eLife, Erika DeBenedictis, Dieter Söll and Kevin Esvelt at the Massachusetts Institute of Technology and
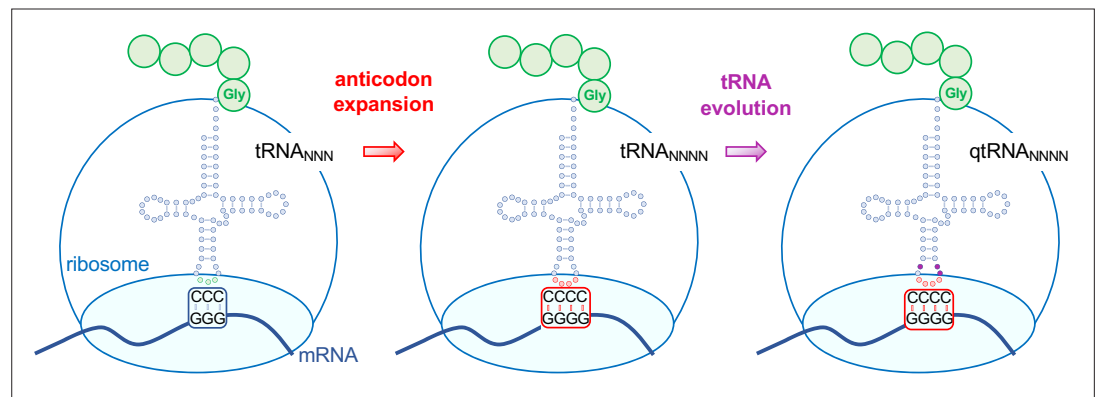
Figure 1. Evolving tRNAs to efficiently read quadruplet codons. The schematic illustrates the approach used by DeBenedictis et al. to evolve triplet-decoding tRNAs (left) into tRNAs with expanded anticodon loops (red dots) to decode quadruplet codons consisting of four nucleotide bases (N; middle), and ultimately into efficient quadruplet codon decoders (right). To test the efficiency of four-base translation, DeBenedictis et al. created reporter genes (such as luciferase) with a single base insertion (also referred to as +1 frameshift). The ability of a particular tRNA variant to read that +1-frameshift mutation as a four-base codon can be measured as a function of how much full length and active reporter protein (e.g., luciferase) is made in cells. DeBenedictis et al. then recorded the translation efficiency of tRNAs with simple mutations to expand the anticodon loop (middle, red dots). Next, tRNAs were evolved through various mutations into more effective quadruplet-decoding tRNAs (qtRNAs, purple dots, right). The work represents an important step towards engineering a quadruplet genetic code with 256 codons.

Yale University, report how tRNAs can be evolved to read quadruplet codons more efficiently (*DeBenedictis et al., 2022*). DeBenedictis et al. made simple mutations to several natural and synthetic tRNAs to generate various quadruplet-decoding RNAs (qtRNAs) based on the natural frameshifting variants and tested their efficiency in the bacterium *Escherichia coli*. This revealed that quadruplet codons only had a translation efficiency of 1%–3% compared to triplet codons. Some qtRNAs also showed significantly slowed growth rates in *E. coli*, while others were better tolerated despite their ability to translate quadruplet codons.

DeBenedictis et al. then tested whether tRNAs could evolve into qtRNAs through replacement of the entire anticodon (the complementary sequence to the codon located on the tRNA) with quadruplet codons using a phage-based library selection approach. This involved encoding modified qtRNAs into the genome of bacteriophages, which in turn infected the *E. coli* bacteria. Only functional qtRNAs enabled phages to successfully reproduce, thus highlighting effective quadruplet codon/anticodon pairs and comparing the efficiency of quadruplet decoding. Their data revealed that functional qtRNAS that can successfully decode quadruplet codons can arise through just a few mutations.

Next, the positions surrounding the anticodon were mutated across a large library of different qtRNAs (*Figure 1*). These experiments identified qtRNA variants that displayed up to 40-fold improvement in quadruplet decoding efficiency compared to unmutated quadruplet decoding tRNAs. Finally, mass spectrometry was used to show how some qtRNAs provided 'high-fidelity' four-base decoding and inserted just a single type of amino acid in response to a quadruplet codon. Other qtRNAs were less selective and the same four-base codon was 'read' with multiple different amino acids, revealing ambiguity as a potential limitation in quadruplet decoding.

Overall, the findings highlight the potential and limitations of a genetic code based on quadruplet codons. Quadruplet codons can be used to produce highly specialized proteins with new functionalities, but low translation fidelity and limited efficiency remain challenges in the field. DeBenedictis et al. use multiple approaches to mutagenize tRNAs, and they generated many improved qtRNA variants that represent a full palate of starting points to reassign quadruplet codons to new amino acids. The researchers demonstrated the power of tRNA variations on their own, and significant improvements in quadruplet decoding by engineering tRNAs alone. Combining qtRNAs with engineered

tRNA synthetases to improve incorporation of unnatural amino acids (reviewed in *Kim et al., 2022*) or with genetically modified ribosomes to enhance 4-base translation (*Dunkelmann et al., 2021*) could further improve efficiency and decrease ambiguity in translating proteins with an increased codon size, thus vastly expanding the genetic code.

The work of DeBenedictis et al. points towards several areas worth further investigation to improve translation efficiency and fidelity with quadruplet codons and they show that in-roads can be made in both areas though tRNA engineering. Together with recent efforts to use quadruplet codons to encode synthetic amino acids (*DeBenedictis et al., 2021*; *Dunkelmann et al., 2021*), the current study suggests that the expansion of the genetic code with the help of quadruplet codons may be within reach.

Tarana Siddika is in the Department of Biochemistry, The University of Western Ontario, London, Canada

Ilka U Heinemann is in the Department of Biochemistry, The University of Western Ontario, London, Canada

http://orcid.org/0000-0003-1721-1667

Patrick O'Donoghue is in the Department of Biochemistry, Department of Chemistry, The University of Western Ontario, London, Canada

patrick.odonoghue@uwo.ca

http://orcid.org/0000-0002-9500-8640

Competing interests: The authors declare that no competing interests exist.

Published 11 May 2022

References

DeBenedictis EA, Carver GD, Chung CZ, Söll D, Badran AH. 2021. Multiplex suppression of four quadruplet codons via tRNA directed evolution. *Nature Communications* **12**:5706. DOI: https://doi.org/10.1038/s41467-021-25948-y, PMID: 34588441

DeBenedictis EA, Söll D, Esvelt KM. 2022. Measuring the tolerance of the genetic code to altered codon size. *eLife* **11**:e76941. DOI: https://doi.org/10.7554/eLife.76941, PMID: 35293861

Dunkelmann DL, Oehm SB, Beattie AT, Chin JW. 2021. A 68-codon genetic code to incorporate four distinct non-canonical amino acids enabled by automated orthogonal mRNA design. *Nature Chemistry* **13**:1110–1117. DOI: https://doi.org/10.1038/s41557-021-00764-5, PMID: 34426682

Hohsaka T, Ashizuka Y, Murakami H, Sisido M. 1996. Incorporation of nonnatural amino acids into streptavidin through in vitro frame-shift suppression. *Journal of the American Chemical Society* **118**:9778–9779. DOI: https://doi.org/10.1021/ja9614225

Kim S, Yi H, Kim YT, Lee HS. 2022. Engineering translation components for genetic code expansion. *Journal of Molecular Biology* **434**:167302. DOI: https://doi.org/10.1016/j.jmb.2021.167302, PMID: 34673113

Nirenberg M, Leder P, Bernfield M, Brimacombe R, Trupin J, Rottman F, O'Neal C. 1965. RNA codewords and protein synthesis, VII on the general nature of the RNA code. *PNAS* **53**:1161–1168. DOI: https://doi.org/10.1073/pnas.53.5.1161, PMID: 5330357

Riyasaty S, Atkins JF. 1968. External suppression of a frameshift mutant in salmonella. *Journal of Molecular Biology* **34**:541–557. DOI: https://doi.org/10.1016/0022-2836(68)90179-4, PMID: 4938557

Söll D, Ohtsuka E, Jones DS, Lohrmann R, Hayatsu H, Nishimura S, Khorana HG. 1965. Studies on polynucleotides, XLIX stimulation of the binding of aminoacyl-sRNA's to ribosomes by ribotrinucleotides and a survey of codon assignments for 20 amino acids. *PNAS* **54**:1378–1385. DOI: https://doi.org/10.1073/pnas.54.5.1378, PMID: 5325653

Tharp JM, Vargas-Rodriguez O, Schepartz A, Söll D. 2021. Genetic encoding of three distinct noncanonical amino acids using reprogrammed initiator and nonsense codons. *ACS Chemical Biology* **16**:766–774. DOI: https://doi.org/10.1021/acschembio.1c00120, PMID: 33723984

Wright DE, Altaany Z, Bi Y, Alperstein Z, O'Donoghue P. 2018. Acetylation regulates thioredoxin reductase oligomerization and activity. *Antioxidants & Redox Signaling* **29**:377–388. DOI: https://doi.org/10.1089/ars.2017.7082, PMID: 29117711