Double-Faced Code

By Alexia Hwang

Students in any life science course have undoubtedly learned about the genetic code that is shared by all life on Earth. The course professor probably presented three key rules for the translation of messenger RNA to proteins: mRNA uses codons, groupings of three nucleotides, to code for amino acids; codons are read sequentially and do not overlap; and one codon encodes for only one amino acid. However, recent research has presented an exception to our understanding of the genetic code and Central Dogma – showing that even Harvard professors can need to learn new lessons.

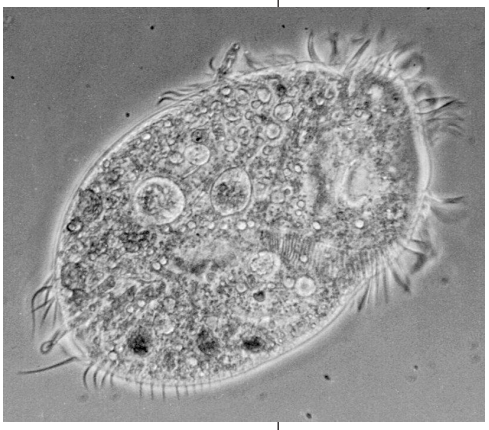


Figure 1. *Euplotes crassus*. This protozoan has been found to have one codon, UGA, encode for two different amino acids: cysteine and selenocysteine.

credit: Science 323, 259-261 (2009)

A research team led by Vadim Gladyshev of the University of Nebraska, Lincoln, has found that a protozoan from the *Euplotes* genus defies the long-held rule of "one codon, one amino acid." Instead, this protozoan has been shown to encode for two different amino acids, cysteine and selenocysteine, using just one codon: UGA.

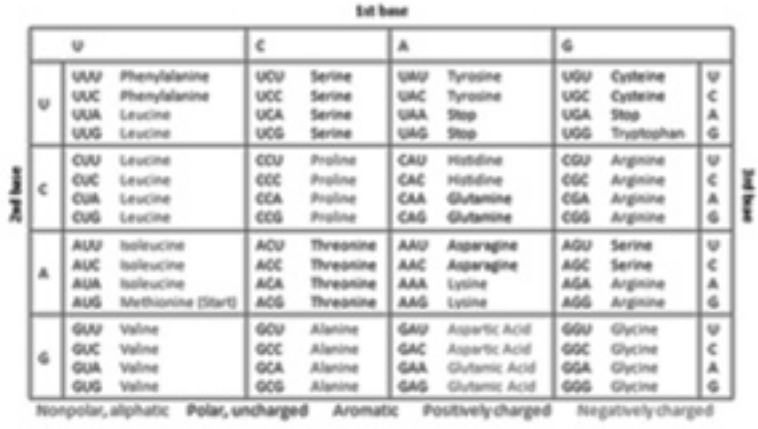
Translation, the first stage in protein biosynthesis, involves the decoding of messenger RNA (mRNA) to form a polypeptide that will ultimately become a functional protein in the cell. The mRNA presents genetic information in three-letter nucleotide combinations called codons, and the standard genetic code connects the 64 possible codon sequences to particular amino acids or stop signals. As there are only 20 standard amino acids (21 including selenocysteine), several codons must represent same amino acid, so the code is considered degenerate.

Gladyshev's work, however, has revealed a new degree of ambiguity in the genetic code. In the ciliate *Euplotes crassus*, the codon UGA dually encodes cysteine and selenocysteine (2). Which amino acid is picked to join the growing polypeptide depends on two factors: the location of the codon and the structure of the mRNA's 3' untranslated region (UTR).

The selenocysteine insertion sequence (SECIS) is a genetic element located in the 3'UTR, upstream of the protein-coding regions of an mRNA (2). The SECIS causes a physical loop in the mRNA to interact with the ribosome, which modifies the UGA signal so that selenocysteine, rather than the normally coded cysteine, becomes incorporated (1). The presence of the SECIS element at the 3' end of the mRNA may explain why selenocysteine is limited to the last 20 codons of the gene while all preceding UGA codons select for cysteine (2). The interaction of the SECIS with UGA codons may have a limited reach, affecting only the codons nearer to the element.

In addition, the physical structure of the 3'UTR seems to convey information about the amino acid to be selected. To test this hypothesis, the research team selected an *E. crassus* gene that incorporated both selenocysteine and cysteine. They replaced the gene's 3'UTR with one from another species whose SECIS element results in a different RNA structure. This caused selenocysteine be incorporated earlier in the sequence, suggesting that the amino acid incorporation depends on the RNA structure (2).

Forms of codon duality are not a completely novel finding, however. In the standard universal genetic code, UGA codes for a stop signal, although in several species, such as *E. crassus*, the triplet codes for cysteine (2). Stop codons in some organisms have a "dual function," in which the codon competes either with termination of translation or with a single amino acid. Additionally, in *Candida* yeast species, the CUG codon signals for both leucine and serine in the same gene (4). Unlike Gladyshev's find-



1st base	2nd base U	2nd base C	2nd base A	2nd base G	3rd base
U	UUU Phenylalanine	UCU Serine	UAU Tyrosine	UGU Cysteine	U
	UUC Phenylalanine	UCC Serine	UAC Tyrosine	UGC Cysteine	C
	UUA Leucine	UCA Serine	UAA Stop	UGA Stop	A
	UUG Leucine	UCG Serine	UAG Stop	UGG Tryptophan	G
C	CUU Leucine	CCU Proline	CAU Histidine	CGU Arginine	U
	CUC Leucine	CCC Proline	CAC Histidine	CGC Arginine	C
	CUA Leucine	CCA Proline	CAA Glutamine	CGA Arginine	A
	CUG Leucine	CCG Proline	CAG Glutamine	CGG Arginine	G
A	AUU Isoleucine	ACU Threonine	AAU Asparagine	AGU Serine	U
	AUC Isoleucine	ACC Threonine	AAC Asparagine	AGC Serine	C
	AUA Isoleucine	ACA Threonine	AAA Lysine	AGA Arginine	A
	AUG Methionine (Start)	ACG Threonine	AAG Lysine	AGG Arginine	G
G	GUU Valine	GCU Alanine	GAU Aspartic Acid	GGU Glycine	U
	GUC Valine	GCC Alanine	GAC Aspartic Acid	GGC Glycine	C
	GUA Valine	GCA Alanine	GAA Glutamic Acid	GGA Glycine	A
	GUG Valine	GCG Alanine	GAG Glutamic Acid	GGG Glycine	G

Nonpolar, aliphatic | Polar, uncharged | Aromatic | Positively charged | Negatively charged

Figure 2. The Universal Genetic Code. Historically, it was thought that one codon signals for one amino acid. Gladyshev and his lab found a protozoan that encodes to two amino acids using one codon. The lab speculate that the finding indicates the possibility of an evolutionarily expanded genetic code (2).

ings, though, the duality in these examples comes from an ambiguity in transfer RNAs (the molecules that shuttle amino acids to the mRNA). With *E. crassus*, two separate tRNAs are charged with their specific amino acids, and the ambiguity arises from the mRNA sequence.

Though this dual signaling codon may seem like a rare phenomenon, Gladyshev and his lab speculate that the new function is an indication of an evolutionarily expanded genetic code (2). The duality shown by *E. crassus* may be the beginning of a more flexible, variable, and informative genetic code—and an impetus for change in Harvard professors' life science syllabi. H

—Alexia Hwang '11 is a Molecular and Cellular concentrator in Currier House.

References:

1. Stephenson, H. "Genetic Code Sees Double." *ScienceNOW Daily News* 08 Jan 2009. *Science* 11 Apr 2009 <http://sciencenow.sciencemag.org/cgi/content/full/2009/108/3>.
2. Turanov, A. A. *et al.* (2009). "Genetic Code Supports Tar geted Insertion of Two Amino Acids by One Codon." *Science* 323(5911): 259 - 261.
3. Copeland, P. R. (2005). "Making sense of nonsense: the evolution of selenocysteine usage in proteins." *Genome Biology* 6(6): 221.
4. Santos, M. A. S. *et al.* (1999). "Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in Candida spp." *Molecular Microbiology* 31(3): 937-947.

credit: Dave15o1, WikiMedia Commons