

Genetic code

Contributed by: P. Schimmel, K. Ewalt

Publication year: 2014

The rules by which the base sequences of deoxyribonucleic acid (DNA) are translated into the amino acid sequences of proteins. Each sequence of DNA that codes for a protein is transcribed or copied (**Fig. 1**) into messenger ribonucleic acid (mRNA). Following the rules of the code, discrete elements in the mRNA, known as codons, specify each of the 20 different amino acids that are the constituents of proteins. In a process called translation, the cell decodes the message in mRNA. During translation (**Fig. 2**), another class of RNAs, called transfer RNAs (tRNAs), are coupled to amino acids, bind to the mRNA, and in a step-by-step fashion provide the amino acids that are linked together in the order called for by the mRNA sequence. The specific attachment of each amino acid to the appropriate tRNA, and the precise pairing of tRNAs via their anticodons to the correct codons in the mRNA, form the basis of the genetic code. *See also:* DEOXYRIBONUCLEIC ACID (DNA); PROTEIN; RIBONUCLEIC ACID (RNA).

Universal genetic code

The genetic information in DNA is found in the sequence or order of four bases that are linked together to form each strand of the two-stranded DNA molecule. The bases of DNA are adenine, guanine, thymine, and cytosine, which are abbreviated A, G, T, and C. Chemically, A and G are purines, and C and T are pyrimidines. The two strands of DNA are wound about each other in a double helix that looks like a twisted ladder (**Fig. 1**). The two posts of the ladder consist of sugar (deoxyribose) and phosphate chains that link the bases together in a directional manner. In DNA, the two strands are said to be antiparallel, because they run in opposite directions. Each rung of the ladder is formed by two bases, one from each strand, that pair with each other by means of hydrogen bonds. For a good fit, a pyrimidine must pair with a purine; in DNA, A bonds with T, and G bonds with C. *See also:* PURINE; PYRIMIDINE.

Ribonucleic acids such as mRNA or tRNA also comprise four bases, except that in RNA the pyrimidine uracil (U) replaces thymine. During transcription, a single-stranded mRNA copy of one strand of the DNA is made. That mRNA copy of the DNA is bound to a large particle known as the ribosome and decoded according to the rules of the genetic code (**Fig's. 1 and 2**). *See also:* RIBOSOMES.

If two bases at a time are grouped together, then only 4×4 or 16 different combinations are possible, a number that is insufficient to code for all 20 amino acids that are found in proteins. However, if the four bases are grouped together in threes, then there are $4 \times 4 \times 4$ or 64 different combinations. Read sequentially without overlapping, those groups of three bases constitute a codon, the unit that codes for a single amino acid.

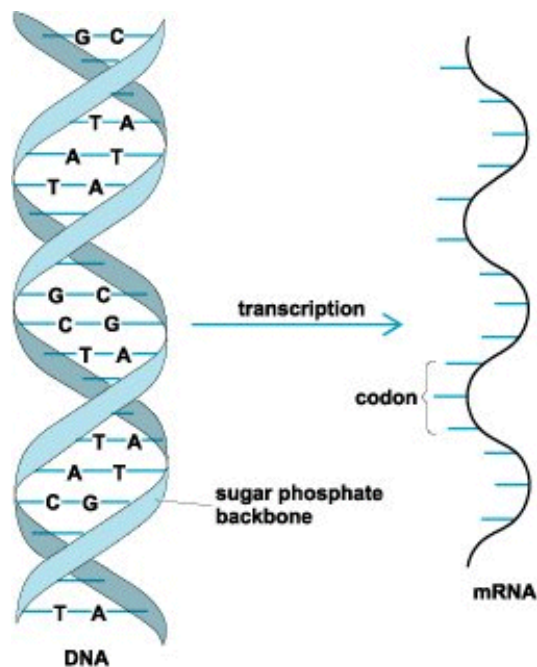


Fig. 1 DNA double helix and mRNA copy made from it by transcription.

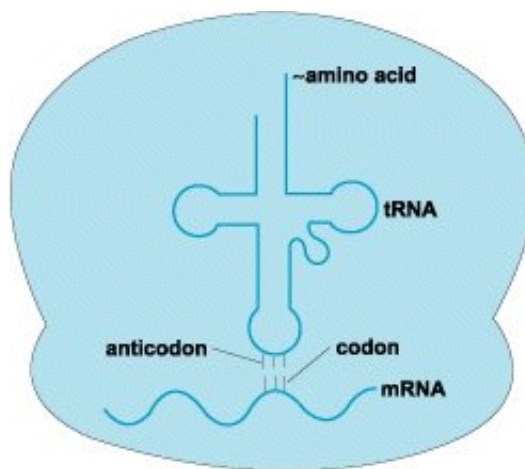


Fig. 2 Translation process in the ribosome, showing the tRNA with attached amino acid interacting with the mRNA via the codon-anticodon interaction.

The 64 codons can be divided into 16 families of four (**Fig. 3**), in which each codon begins with the same two bases. With the number of codons exceeding the number of amino acids, several codons can code for the same amino acid. Thus, the code is degenerate. In eight instances, all four codons in a family specify the same amino acid. In the remaining families, the two codons that end with the pyrimidines U and C often specify one amino

	U	C	A	G
U	UUU } Phe UUC } UUA } Leu UUG }	UCU } Ser UCC } UCA } UCG }	UAU } Tyr UAC } UAA } Stop UAG }	UGU } Cys UGC } UGA } Stop UGG } Trp
C	CUU } Leu CUC } CUA } CUG }	CCU } Pro CCC } CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } Arg CGC } CGA } CGG }
A	AUU } Ile AUC } AUA } AUG } Met	ACU } Thr ACC } ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }
G	GUU } Val GUC } GUA } GUG }	GCU } Ala GCC } GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } Gly GGC } GGA } GGG }

Fig. 3 Universal (standard) genetic code. Each of the 64 codons found in mRNA specifies an amino acid (indicated by the three-letter abbreviation) or the end of the protein chain (*stop*). The amino acids are phenylalanine (Phe), leucine (Leu), isoleucine (Ile), methionine (Met), valine (Val), serine (Ser), proline (Pro), threonine (Thr), alanine (Ala), tyrosine (Tyr), histidine (His), glutamine (Gln), asparagine (Asn), lysine (Lys), aspartic acid (Asp), glutamic acid (Glu), cysteine (Cys), tryptophan (Trp), arginine (Arg), and glycine (Gly).

acid, whereas the two codons that end with the purines A and G specify another. Furthermore, three of the codons, UAA, UAG, and UGA, do not code for any amino acid but instead signal the end of the protein chain.

Decoding mRNAs with tRNAs

The nucleic acid code of an mRNA is converted into an amino acid sequence with the aid of tRNAs. These RNAs are relatively small nucleic acids that vary from 75 to 93 bases in length. Some of the bases in a tRNA pair with each other, so that a tRNA is folded into a pattern that can be represented in two dimensions as a cloverleaf (Fig. 2). This pattern is further folded in three dimensions to form an L-shaped molecule (Fig. 4), to which an amino acid can then be attached. At the other end, three bases in a single-stranded loop are free to pair with a codon in the mRNA. These three bases of a tRNA constitute the anticodon.

Each amino acid has one or more tRNAs, and because of the degeneracy of the code, many of the tRNAs for a specific amino acid have different anticodon sequences. However, the tRNAs for one amino acid are capable of pairing their anticodons only with the codon or codons in the mRNA that specify that amino acid. The tRNAs act as interpreters of the code, providing the correct amino acid in response to each codon by virtue of precise codon-anticodon pairing. The tRNAs pair with the codons and sequentially insert their amino acids in the exact order specified by the sequence of codons in the mRNA. That takes place in the ribosome.

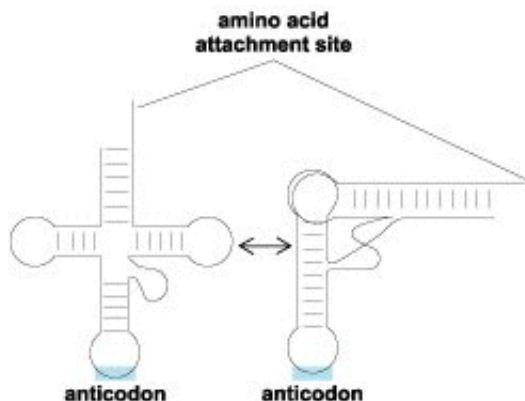


Fig. 4 A transfer RNA in the cloverleaf representation is shown with a two-dimensional projection of how the structure is folded into an L-shaped molecule in three dimensions. The short horizontal and vertical lines represent base pairs that form within the structure. (Courtesy of Dr. Douglas Buechter)

Codon-anticodon interaction

The anticodon of the tRNA pairs with the codon in the mRNA in an antiparallel fashion so that the first base of the codon pairs with the third base of the anticodon, the second base of the codon pairs with the second base of the anticodon, and the third base of the codon pairs with the first base of the anticodon. For the first two bases of the codon, pairing follows the usual rule: A pairs with U, and G with C. However, at the third base of codon, pairing occurs according to a set of rules first suggested by F. Crick and known as the wobble hypothesis:

<i>First base of anticodon</i>	<i>Third base of codon</i>
C	G
U	A or G
G	C or U
I	U, C, or A

The purine base inosine (I) is often found instead of A in the first position of the anticodon. There, inosine can pair with any of three bases in the third position of the codon, thus allowing a single tRNA to read three out of the four codons for a single amino acid. However, inosine is not found in tRNAs in which it would allow pairing with a codon that specifies another amino acid. For example, the histidine tRNA does not contain inosine in the first position of the anticodon because that would allow pairing with the CAA codon which codes for glutamine, not histidine.

Variations

The rules of the genetic code are virtually the same for all organisms, but there are some interesting exceptions. In the microorganism *Mycoplasma capricolum*, UGA is not a stop codon; instead it codes for tryptophan. This alteration in the code is also found in the mitochondria of some organisms. In contrast, UGA is the sole termination codon in some ciliated protozoa, where UAA and UAG are translated as glutamine. A special case involves the unusual amino acid selenocysteine. This metal (selenium)-containing amino acid has a unique tRNA that inserts selenocysteine in response to a UGA codon, in a limited number of proteins in bacteria and mammals. Because these organisms also use UGA as a stop codon, it is thought that the base sequence of the mRNA around selenocysteine-encoding UGAs influences reading to favor amino acid insertion instead of termination. Other specific changes in the meaning of codons that normally encode isoleucine, leucine, and arginine have also been observed in the mitochondria of some organisms.

In addition to changes in the meanings of codons, a modified system for reading codons that requires fewer tRNAs is found in mitochondria. Following the wobble rules, a minimum of two tRNAs are normally needed to read each family of four codons. For example, two tRNAs, with I or C in the first position of the anticodon, could read all four members of a family. In contrast, each of the eight four-codon families that codes for a single amino acid in mitochondria is read by a single tRNA with U in the first position of the anticodon. An explanation for this apparent relaxation of the wobble rules is that the presence of a chemical modification of U in the universal code restricts pairing of U to A or G. In mitochondrial tRNAs, U is unmodified and is therefore able to pair with codons ending in U, C, A, or G. In those cases where two codons in a family specify one amino acid and the other two members specify another amino acid or a stop codon (Fig. 3), the U in the first position of the anticodon is modified and thus pairs only with A and G, as in the universal system. *See also:* MITOCHONDRIA.

Aminoacyl tRNA synthetases

The specific attachment of amino acids to tRNAs is catalyzed by a class of enzymes known as aminoacyl-tRNA synthetases. Each amino acid has a different aminoacyl-tRNA synthetase, which catalyzes a reaction known as aminoacylation. Each enzyme attaches its particular amino acid to all of the tRNAs that have anticodons corresponding to that amino acid. Each aminoacyl-tRNA synthetase must distinguish between the various tRNAs so that only the proper ones are aminoacylated. Thus, the accurate recognition of tRNAs by these enzymes establishes the rules of the genetic code.

Aminoacyl tRNA synthetases are divided into two classes of 10 synthetases each. The classification of a synthetase is based on which of two distinct three-dimensional structures it has for the active-site region that catalyzes attachment of an amino acid to its cognate tRNA. There is no evidence that the two distinct structures for the active-site region came from a common ancestor, and it is believed that the two classes arose independently. Also, the occurrence of two classes of synthetases appears to be universal; that is, there is no known example of an organism where all synthetases are of the same class. The historical rationale for the two classes of synthetases

is a question of great interest. An attractive hypothesis that is gaining support postulates that, in an early environment, synthetases from opposite classes paired themselves on a single tRNA. One member from each class bound to opposite sides of a given tRNA to form a two-synthetase-one-tRNA complex. From this perspective, the distinct architectures of the two classes of synthetases are seen as needed in order to fit two synthetases to distinct sides of the tRNA.

Because synthetases are responsible for the genetic code and are present in all living organisms, it is likely that they are among the oldest enzymes on Earth. The active-site region of a synthetase is thought to be the primordial, or earliest, form of the synthetase. In addition to the active-site-containing region that is conserved among all members of the same class, each synthetase has a region that is referred to as a nonconserved domain. This domain is more or less unique to the synthetase, even within a particular class.

Recognition of tRNAs

Aminoacyl tRNA synthetases engage in a process called recognition of tRNAs, analogous to the recognition of different buildings or trees by the human eye. Each tRNA has one or more distinguishing characteristics that are determined by the nucleotide sequence and associated three-dimensional structure. The distinguishing features are recognized by the appropriate synthetase that catalyzes the attachment of its cognate amino acid.

One feature that distinguishes a particular tRNA is the anticodon trinucleotide. Were the anticodon trinucleotide the only feature that synthetases use to distinguish among tRNAs, then the interaction of a tRNA with a synthetase would establish a direct connection between a particular amino acid and its complementary trinucleotide of the genetic code. However, there are synthetases that do not interact with the anticodon and yet still correctly identify their tRNA. In these cases, the relationship between a particular trinucleotide and a given amino acid is indirect, and suggests the existence of another code, namely, an operational RNA code for amino acids that was the precursor to the genetic code.

Second genetic code

In the case of the tRNA specific for alanine with an anticodon GGC, a single G-U base pair, at a specific position far removed from the anticodon, is needed for aminoacylation with alanine. This base pair is located in a portion of the tRNA molecule relatively close to the site of amino acid attachment (**Fig. 5**). The part of the tRNA structure that is near the amino acid attachment site is known as the acceptor stem, and consists of a short helix joined to a single-stranded piece that terminates at the site of amino acid attachment. A truncated version of alanine tRNA that contains the critical base pair together with the adjacent structure near the amino acid attachment site can be synthesized as a minihelix. This tRNA fragment is recognized by alanine tRNA synthetase.

Even smaller RNA fragments that recapitulate the acceptor ends of several tRNAs are aminoacylated with the correct amino acid (**Fig. 6**). Because aminoacylation is specific and depends on the order of bases in the short

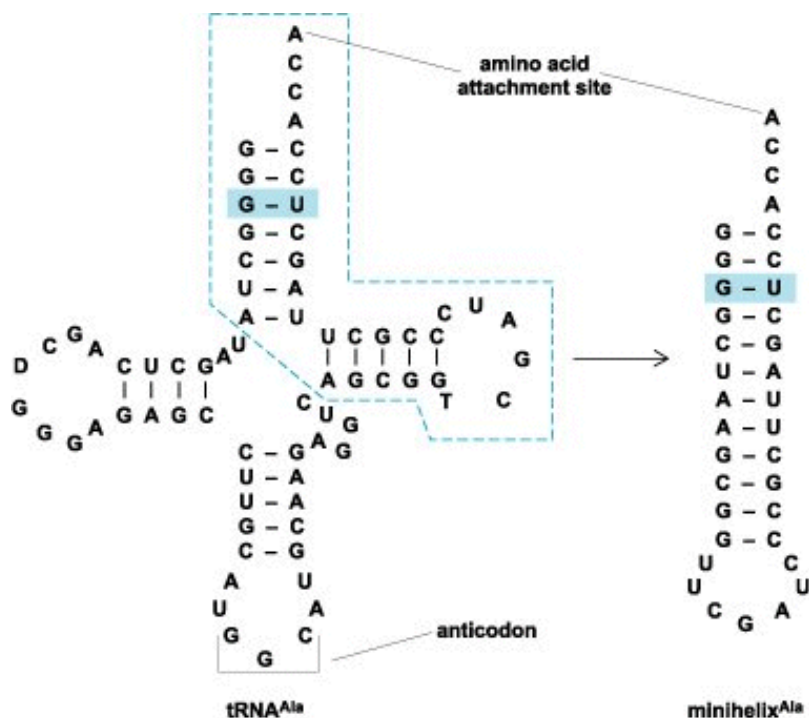


Fig. 5 Cloverleaf representation of the nucleotide sequence, comprising four monomer units (A = adenine, G = guanine, C = cytosine, and U = uracil), of an alanine tRNA with a GGC anticodon and of a synthetic minihelix based on a limited portion of the whole tRNA (enclosed by a broken line). D, T, and Ψ are modified uracil bases in the tRNA, and T and Ψ have been replaced by U in the minihelix. The critical G-U base pair recognized by the alanine tRNA synthetase is highlighted in both structures. (Courtesy of Dr. Douglas Buechter)

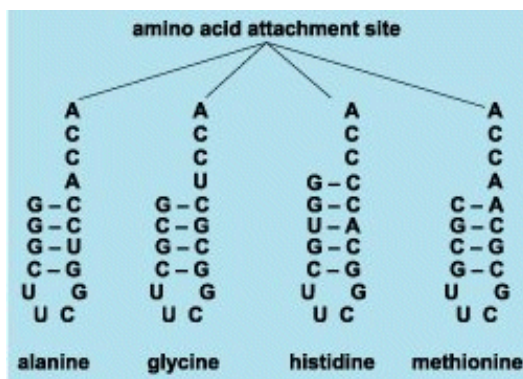


Fig. 6 The smallest RNA oligonucleotides that have been shown to be specifically aminoacylated. The amino acid which can be attached to each RNA oligonucleotide is indicated. (Courtesy of Dr. Barry Henderson)

fragments, the relationship between the structure-sequence of RNA fragments and each amino acid can be viewed as an operational RNA code, referred to also as the second genetic code.

Several synthetases recognize the anticodon as one of the elements needed for identification of the tRNA, but even in these cases nucleotides in the acceptor stem are still important for precise and efficient recognition of the whole tRNA and small RNA fragments. The class-defining and active-site-containing domain of aminoacyl tRNA synthetases is largely responsible for interactions with the acceptor end of a tRNA and, therefore, for interpretation of the second genetic code. After establishment of the second genetic code in a primitive world, RNA fragments with signals for attachment of specific amino acids may have become incorporated into RNA structures which eventually became transfer RNAs. That step, in turn, gave birth to the modern genetic code. In addition, the synthetases developed editing functions that are tRNA-dependent. The development of these editing functions was one of the events that linked the evolution of tRNAs with that of the synthetases. These editing functions enable the synthetases to achieve high accuracy of aminoacylation. During evolution, the earliest tRNA synthetases apparently acquired additional, unique domains which allowed for editing. *See also:* GENE; GENETICS.

P. Schimmel, K. Ewalt

Bibliography

J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry*, 7th ed., 2012

S. Cusack, RNA-protein complexes, *Curr. Opin. Struc. Biol.*, 9:66–73, 1999

DOI: [http://doi.org/10.1016/S0959-440X\(99\)80009-8](http://doi.org/10.1016/S0959-440X(99)80009-8)

A. J. F. Griffiths et al., *Introduction to Genetic Analysis*, 9th ed., 2008

P. Schimmel and L. Ribas de Pouplana, Formation of two classes of tRNA synthetases in relation to editing functions and genetic code, *Cold Spring Harbor Symp. Quantitative Biol.*, 66:161–166, 2001

DOI: <http://doi.org/10.1101/sqb.2001.66.161>

P. Schimmel and L. Ribas de Pouplana, Transfer RNA: From minihelix to genetic code, *Cell*, 81:983–986, 1995

DOI: [http://doi.org/10.1016/S0092-8674\(05\)80002-9](http://doi.org/10.1016/S0092-8674(05)80002-9)

Additional Readings

R. H. Garrett and C. M. Grisham, *Biochemistry*, 5th ed., Brooks/Cole, Belmont, CA, 2013

B. Lewin et al., *Lewin's Essential Genes*, 2d ed., Jones & Bartlett Learning, Burlington, MA, 2010

C. C. Liu and P. G. Schultz, Adding new chemistries to the genetic code, *Annu. Rev. Biochem.*, 79:413–444, 2010

DOI: <http://doi.org/10.1146/annurev.biochem.052308.105824>

M. Ram, *Fundamentals of Cytogenetics and Genetics*, PHI Learning, New Delhi, India, 2010