

Genetic code

Genetic code, the sequence of nucleotides in deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) that determines the amino acid sequence of proteins. Though the linear sequence of nucleotides in DNA contains the information for protein sequences, proteins are not made directly from DNA. Instead, a messenger RNA (mRNA) molecule is synthesized from the DNA and directs the formation of the protein. RNA is composed of four nucleotides: adenine (A), guanine (G), cytosine (C), and uracil (U). Three adjacent nucleotides constitute a unit known as the codon, which codes for an amino acid. For example, the sequence AUG is a codon that specifies the amino acid methionine. There are 64 possible codons, three of which do not code for amino acids but indicate the end of a protein. The remaining 61 codons specify the 20 amino acids that make up proteins. The AUG codon, in addition to coding for methionine, is found at the beginning of every mRNA and indicates the start of a protein. Methionine and tryptophan are the only two amino acids that are coded for by just a single codon (AUG and UGG, respectively). The other 18 amino acids are coded for by two to six codons. Because most of the 20 amino acids are coded for by more than one codon, the code is called degenerate.

The genetic code, once thought to be identical in all forms of life, has been found to diverge slightly in certain organisms and in the mitochondria of some eukaryotes. Nevertheless, these differences are rare, and the genetic code is identical in almost all species, with the same codons specifying the same amino acids. The deciphering of the genetic code was accomplished by American biochemists Marshall W. Nirenberg, Robert W. Holley, and Har Gobind Khorana in the early 1960s.

Nucleotide triplets (codons) specifying different amino acids are shown in the table.

The genetic code: Nucleotide triplets (codons) specifying different amino acids in protein chains*

DNA triplet	RNA triplet	amino acid
AAA	UUU	phenylalanine
AAG	UUC	
AAT	UUA	
AAC	UUG	
GAA	CUU	

GAG	CUC	leucine
GAT	CUA	
GAC	CUG	
AGA	UCU	
AGG	UCC	
AGT	UCA	
AGC	UCG	serine
TCA	AGU	
TCG	AGC	
GGA	CCU	
GGG	CCC	
GGT	CCA	proline
GGC	CCG	
TAA	AUU	
TAG	AUC	isoleucine (Ileu)
TAT	AUA	
TAC	AUG	methionine
TGA	ACU	
TGG	ACC	
TGT	ACA	threonine
TGC	ACG	
CAA	GUU	
CAG	GUC	
CAT	GUA	valine
CAC	GUG	
CGA	GCU	
CGG	GCC	
CGT	GCA	alanine
CGC	GCG	
ACA	UGU	cysteine
ACG	UGC	

ACC	UGG	tryptophan
ATA	UAU	tyrosine
ATG	UAC	
ATT	UAA	
ATC	UAG	(termination: end of specification)
ACT	UGA	arginine
GCA	CGU	
GCG	CGC	
GCT	CGA	
GCC	CCG	
TCT	AGA	
TCC	AGG	histidine
GTA	CAU	
GTG	CAC	
GTT	CAA	glutamine (GluN)
GTC	CAG	
TTA	AAU	asparagine (AspN)
TTG	AAC	
TTT	AAA	lysine
TTC	AAG	
CCA	GGU	glycine
CCG	GGC	
CCT	GGA	
CCC	GGG	aspartic acid
CTA	GAU	
CTG	GAC	
CTT	GAA	glutamic acid
CTC	GAG	

*The columns may be read thus: The DNA triplet is transcribed into an RNA triplet, which then directs the production of an amino acid.

The Editors of Encyclopaedia Britannica This article was most recently revised and updated by

Kara Rogers, Senior Editor.

CITATION INFORMATION

ARTICLE TITLE: Genetic code

WEBSITE NAME: Encyclopaedia Britannica

PUBLISHER: Encyclopaedia Britannica, Inc.

DATE PUBLISHED: 22 November 2019

URL: <https://www.britannica.com/science/genetic-code>

ACCESS DATE: April 25, 2020