

THE GENETIC CODE

What triplet of ribonucleotides directs the addition of, say, the amino acid alanine to a protein that is being synthesized? Of lysine? Of any one of the twenty amino acids found in proteins? That was the problem to be faced after advancement of the ideas that a gene is a string of deoxyribonucleotide triplets, that the string of deoxyribonucleotide triplets is transcribed into a string of ribonucleotide triplets, and that the string of ribonucleotide triplets is translated into a string of amino acids—a protein. The results of research on the problem is condensed in the genetic code, a listing of the sixty-four possible ribonucleotide triplets and the amino acid (or translation command) corresponding to each. Fortunately for those who worked on the problem, the genetic code is organism-independent. That is, the same genetic code is used by virtually all organisms.

Researchers began to crack the genetic code in the early 1960s. Marshall Nirenberg and his collaborators added a synthetic RNA, consisting entirely of repetitions of a single ribonucleotide, say U, to a bacterial extract that contained everything necessary for protein synthesis except RNA. The result was a string of the amino acid phenylalanine. They concluded that the ribonucleotide triplet UUU codes for phenylalanine. Other ribonucleotide triplets were decoded by performing similar experiments with synthetic RNAs containing only A's, C's, or G's or various combinations of ribonucleotides. By 1966 research teams led by Har Gobind Khorana and Marshall Nirenberg had cracked the entire genetic code.

(a) RNA Codons for the Twenty Amino Acids

		Second base						
		U	C	A	G		Amino-acid abbreviations	
U	Phe	Ser	Tyr	Cys	U	Ala = Alanine		
	Phe	Ser	Tyr	Cys	C	Arg = Arginine		
	Leu	Ser	STOP	STOP	A	Asp = Aspartic acid		
	Leu	Ser	STOP	Trp	G	Asn = Asparagine		
C	Leu	Pro	His	Arg	U	Cys = Cysteine		
	Leu	Pro	His	Arg	C	Glu = Glutamic acid		
	Leu	Pro	Gln	Arg	A	Gln = Glutamine		
	Leu	Pro	Gln	Arg	G	Gly = Glycine		
A	Ile	Thr	Asn	Ser	U	His = Histidine		
	Ile	Thr	Asn	Ser	C	Ile = Isoleucine		
	Ile	Thr	Lys	Arg	A	Leu = Leucine		
	Met (start)	Thr	Lys	Arg	G	Lys = Lysine		
G	Val	Ala	Asp	Gly	U	Met = Methionine		
	Val	Ala	Asp	Gly	C	Phe = Phenylalanine		
	Val	Ala	Glu	Gly	A	Pro = Proline		
	Val	Ala	Glu	Gly	G	Ser = Serine		

Shown in (a) is the usual representation of the genetic code. The letters U, C, A, and G are symbols for the ribonucleotides containing the bases uracil, cytosine, adenine, and guanine, respectively. The symbols in the body of the table are three-letter abbreviations for the amino acids. To find the amino acid specified by a particular codon (say the codon CAG), locate the first nucleotide (C) along the left side of the table and the second nucleotide (A) along the top of the table. Their intersection pinpoints one of four amino acids. Of those four the one aligned with the third nucleotide (G) is the amino acid in question. Thus the amino acid glutamine (Gln) is specified by the three-nucleotide sequence CAG.

Shown in (b) is another version of the genetic code, one expressed in terms of DNA

codons instead of RNA codons. Each single-stranded deoxyribonucleotide triplet listed in (b) is the sequence of the so-called sense strand of a DNA codon—the strand that does not serve as a template for synthesis of RNA. Note that most of the amino acids are specified by at least two codons. For example, phenylalanine is specified by two codons: TTT and TTC. Arginine is specified by a total of six codons: CGT, CGC, CGA, CGG, AGA, and AGG. In general, the more an amino acid is used in protein synthesis the likelier it is to be specified by more than one codon. Note also the start codon (ATG) and the three stop codons (TAA, TGA, and TAG) that are used to signal the beginning and end of protein synthesis. The substantive difference between the two versions of the genetic code is that in (b) the deoxyribonucleotide T replaces the ribonucleotide U.

(b) DNA Codons for the Twenty Amino Acids

Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ileu	Leu	Lys	Met (START)	Phe	Pro	Ser	Thr	Trp	Tyr	Val	STOP
GCA	AGA	GAT	AAT	TGT	GAA	CAA	GGA	CAT	ATA	TTA	AAA	ATG	TTT	CCA	AGT	ACA	TGG	TAT	GTA	TAA
GCG	AGG	GAC	AAC	TGC	GAG	CAG	GGG	CAC	ATT	TTG	AAG		TTC	CCG	AGC	ACG		TAC	GTG	TAG
GCT	CGA						GGT	ATC	CTA				CCT	TCA	ACT				GTT	TGA
GCC	CGG						GCC		CTG				CCC	TCG	ACC				GTC	
	CGT								CTT						TCT					
	CGC								CTC						TCC					