

Cracking the Genetic Code

You learned about the genetic code in Biology. It's the mapping from nucleotide triplets in DNA sequences (via messenger RNA) to individual amino acids in the protein encoded by a given gene. You may recall that there are 64 "codons" (distinct triplets of G, A, C, and T) but only 20 amino acids, and that the code is redundant or "degenerate" in the sense that several different codons typically specify a given amino acid. There are also a few codons that specify *no* amino acid and thereby cause the growing polypeptide chain to be terminated.

In the 1950s biologists learned how DNA is structured (the famous "double helix") and how it is replicated, and they knew that the amino-acid sequences of proteins must be specified by the sequence of nucleotide bases in the DNA sequence of a gene, as transcribed (copied) into messenger RNA. But they had no idea what the mapping was, which is to say, how the DNA/RNA sequence was "read" to "mean" a particular protein.

Marshall Nirenberg then realized that he could "ask" the translational machinery of *E. coli* how it "understood" a given RNA sequence by synthesizing a simple known sequence and adding it to a cell-free extract of *E. coli* that was capable of synthesizing proteins. Nirenberg gave polyuridine (UUUUU...) to such an extract, and got polyphenylalanine (Phe-Phe-Phe ... or FFF ...). This clearly meant that "words" consisting purely of U somehow caused the incorporation of phenylalanine.

However, this finding did *not* reveal how the "words" in RNA are *structured*. For example, how *long* are the words, and are they all the *same* length? Do they consist of *consecutive* bases, or ones that are somehow *interdigitated*? Do words *overlap* with each other, or are they *non-overlapping*? Is there some kind of "punctuation" between words, to designate their beginnings and ends? These and other basic questions remained completely open, even after Nirenberg's great experiment.

H. Gobind Khorana then conceived and carried out the experiment that decisively revealed the nature of the genetic code. He synthesized artificial messages more complex than Nirenberg's and analyzed the resulting polypeptides. His data are shown below. "(XY)_n" means "XYXYXY ...", and the resulting amino-acid couplet also repeats indefinitely (e.g., Ser-Leu-Ser-Leu-Ser-Leu ...). Please use these data (plus Nirenberg's result) to figure out as much of the code as you can. In particular, show that the code must be structured as consecutive, non-overlapping triplets without punctuation, that the third position is often degenerate, and that some codons mean "stop!"

mRNA	Polypeptide(s) synthesized
(UC) _n	(Ser-Leu)
(UG) _n	(Val-Cys)
(AC) _n	(Thr-His)
(AG) _n	(Arg-Glu)
(UUC) _n	(Ser-Ser) and (Leu-Leu) and (Phe-Phe)
(UUG) _n	(Leu-Leu) and (Val-Val) and (Cys-Cys)
(AAG) _n	(Arg-Arg) and (Lys-Lys) and (Glu-Glu)
(CAA) _n	(Thr-Thr) and (Asn-Asn) and (Gln-Gln)
(UAC) _n	(Thr-Thr) and (Leu-Leu) and (Tyr-Tyr)
(AUC) _n	(Ile-Ile) and (Ser-Ser) and (His-His)
(GUA) _n	(Ser-Ser) and (Val-Val)
(GAU) _n	(Asp-Asp) and (Met-Met)
(UAUC) _n	(Tyr-Leu-Ser-Ile)
(UUAC) _n	(Leu-Leu-Thr-Tyr)
(GAUA) _n	None
(GUAA) _n	None

The standard amino acids		
alanine	ala	A
cysteine	cys	C
aspartic acid	asp	D
glutamic acid	glu	E
phenylalanine	phe	F
glycine	gly	G
histidine	his	H
isoleucine	ile	I
lysine	lys	K
leucine	leu	L
methionine	met	M
asparagine	asn	N
proline	pro	P
glutamine	gln	Q
arginine	arg	R
serine	ser	S
threonine	thr	T
valine	val	V
tryptophan	trp	W
tyrosine	tyr	Y



Khorana, Har Gobind



Har Gobind Khorana

Copyright Hulton Getty/Tony Stone Images

(b. Jan. 9, 1922, Raipur, India), Indian-born American biochemist who shared the [Nobel Prize in Physiology or Medicine](#) in 1968 with [Marshall W. Nirenberg](#) and [Robert W. Holley](#) for research that helped to show how the nucleotides in nucleic acids, which carry the [genetic code](#) of the cell, control the cell's synthesis of proteins.

Khorana was born into a poor family and attended Punjab University at Lahore and the University of Liverpool, England, on government scholarships. He obtained his Ph.D. at Liverpool in 1948. He began research on nucleic acids during a fellowship at the University of Cambridge (1951) under Sir Alexander Todd. He held fellowships and professorships in Switzerland at the Swiss Federal Institute of Technology and the universities of British Columbia, Can. (1952-59), and Wisconsin, U.S. In 1971 he joined the faculty of the Massachusetts Institute of Technology.

In the 1960s Khorana confirmed Nirenberg's findings that the way the four different types of nucleotides are arranged on the spiral "staircase" of the DNA molecule determines the chemical composition and function of a new cell. The 64 possible combinations of the nucleotides are read off along a strand of DNA as required to produce the desired amino acids, which are the building blocks of proteins. Khorana added details about which serial combinations of nucleotides form which specific amino acids. He also proved that the nucleotide code is always transmitted to the cell in groups of three, called codons. Khorana also determined that some of the codons prompt the cell to start or stop the manufacture of proteins. Khorana made another contribution to genetics in 1970, when he and his research team were able to synthesize the first artificial copy of a yeast gene.



Copyright © 1997 Encyclopædia Britannica, Inc. All Rights Reserved.