CHAPTER 14

NIRENBERG/KHORANA: BREAKING THE GENETIC CODE

When it became known that each amino acid was coded for by a sequence of three nucleotide bases, scientists eagerly sought to determine which *triplets went with* which *amino acids. In 1964, Marshall Nirenberg and Har Gobind Khorana worked out the puzzle of the genetic code. By using radioactively-labeled synthetic mRNA molecules, they were able to assign specific triplets to each of the 20 amino acids.*

BREAKING THE CODE REQUIRED ORGANIC CHEMISTRY

The key breakthrough in deciphering the genetic code came from an unexpected direction. In 1960, Marshall Nirenberg and J. H. Matthaei developed a system for synthesizing proteins *in vitro*. They had learned that preparation of disrupted cells soon ceased to make protein, and, in an attempt to prolong the short period during which *in vitro* synthesis continued, they added RNA to the preparations (rRNA, as it happens). rRNA indeed prolonged the period of *in vitro* protein synthesis and all 20 amino acids were actively incorporated into newly-made protein. As a control, they used an artificial RNA, reasoning that only RNA sequences with physiological significance should be active in *in vitro* protein synthesis. Artificial RNA, because it was not naturally occurring, should not prolong *in vitro* protein synthesis. Well, an experiment is only as good as its controls, and in this case the control proved far more important than the experiment itself (the effect of rRNA on *in vitro* protein synthesis was later shown to be indirect). Nirenberg and Matthaei used the enzyme polynucleotide phosphorylase, which synthesizes RNA chains randomly from available precursors without a template, to make the artificial RNA polyuridylic acid (poly-U) from UDP. They added the poly-U to a fresh, disrupted cell suspension (*cell-free extract*), expecting the rapid decay of *in vitro* protein synthesis (they monitored the ^{14}C amino acid into acid-precipitable protein to detect protein synthesis). Instead, protein synthesis was stimulated! Activity was so great as to make the rRNA activity levels seem miniscule by comparison. *Only 10 micrograms* of poly-U yielded approximately 13,000 ^{14}C amino acid counts per minute (CPM is a measurement of radioactivity; higher levels of radioactivity are indicated by higher counts per minute), while 2,400 micrograms of rRNA yielded only about *200* CPM!) Most importantly, *only ^{14}C phenylalanine was incorporated into protein*. The acid-precipitable 14C label was in polyphenylalanine (PHE-PHE-PHE--). This immediately provided additional confirmation of Brenner, Jacob, and Meselson's mRNA hypothesis, and suggested an additional hypothesis of first importance: that the ribosomes could not distinguish an artificial mRNA from a naturally-derived one. When an artificial mRNA was presented carrying the code word for phenylalanine (evidently UUU), the ribosomes proceeded to read it with high efficiency. In a similar manner, AAA = LYS, and CCC = PRO. It is this approach, the synthesis of synthetic mRNA molecules, which led directly and quickly to the full deciphering of the genetic code.

INFORMATION FROM RANDOM SEQUENCES

At first, attempts were made to deduce the code from more complex artificial mRNA molecules. By presenting polynucleotide phosphorylase with two nucleotides present in varying proportions, RNA chains could be obtained with the two nucleotides present in *random sequence*. This mRNA could then be employed in *in vitro* protein synthesis and protein isolated with several amino acids present. Their composition provided direct code information. Imagine an initial mix of 3:1 U to G. The possibility of UUU is (3/4)(3/4)(3/4), or 27/64; the probability of two U's and one C is (3/4)(1/4)(1/4) or 3/64. Thus, the ratio of PHE to the three codons with two U's and one C should be 3:1, and the ratio of PHE to the codons carrying one C should be 9:1. When one tries poly-UG, 3:1 in *in vitro* protein synthesis, one obtains valine, leucine, and cysteine incorporated about 1/3 as often as phenylalanine, suggesting that the codons for VAL,

LEU, and CYS each obtain two U's and one C. But which is which? This approach cannot tell you that. Artificial mRNA of random sequence can provide information only about codon composition, not codon sequence. What was required then was a sequence-specific probe.

NIRENBERG'S EXPERIMENT

The first such probes were indirect, but powerful. Marshall Nirenberg and Philip Leder showed in 1964 that the simple trinucleotide UUU, while it was incapable of acting as mRNA, would bind with ^{14}C PHE-tRNA (the phenylalanine-specific transfer RNA, charged with ^{14}C labeled phenylalanine) to ribosomes (figure 14.1). The binding required the presence of several additional binding factor proteins and GTP, and was specific: only ^{14}C PHE-tRNA was bound to ribosomes when the UUU trinucleotide triplet was employed. It was thus possible to carry out a simple *triplet binding assay*. A specific triplet (say UGU) was added to a mix containing ribosomes, binding factors, GTP, and a variety of ^{14}C amino acid-charged tRNAs. This mixture was then passed through a filter. While most radioactivity passed through the filter, a small amount remained trapped on the filter surface because the ribosomes adhered to the filter, and the ribosomes had bound to *them* the ^{14}C amino acid-tRNA that recognized UGU. When the filter was analyzed, it contained ^{14}C-cysteine, so UGU = CYS. Because all possible trinucleotides could be readily synthesized, it was possible to decode most three-base codons, despite the indirect nature of the assay. Some 47 of the 64 possible combinations gave unambiguous results.

KHORANA'S EXPERIMENT

The remaining 17 triplets gave ambiguous results on triplet binding assays, and decoding them required a more direct approach. Har Gobind Khorana provided such an approach by setting out to directly construct a series of artificial mRNA molecules of *defined sequence* (figure 14.2). He first constructed short defined sequences of DNA. He knew the sequences of the DNA molecules that he synthesized because he made the DNA from special chemical groups blocked so that only certain base combinations were possible. An over-simplified example might be to imagine G bound to a column matrix, but T blocked chemically so that it could not bind to the column. The blocked T was added to the column under conditions that promoted the nucleotide condensation reaction, and GT was obtained, with unused T washed out the bottom of the column and all the initial G's then bound by T. Blocked G was then added to yield –GTG. In this way, defined DNA double-helical models of 6 to 8 base pairs were constructed. Khorana then used those DNA oligonucleotides as templates for RNA polymerase, and produced specific RNA molecules such as GUGUGUGU-----. Very long mRNA molecules of known sequence could be produced in this fashion.

From an mRNA segment such as GUGUGUGU---, there are two alternating codons, GUG and UGU. When employed in *in vitro* protein synthesis, this mRNA yielded a polypeptide of alternating CYS-VAL-CYS-VAL---. Which was which? From the triplet binding assay, Khorana knew that UGU coded for CYS. Therefore, GUG must code for valine (VAL). By constructing these and more complicated defined-sequence mRNAs, Khorana was able to verify the entire code (figure 14.3).

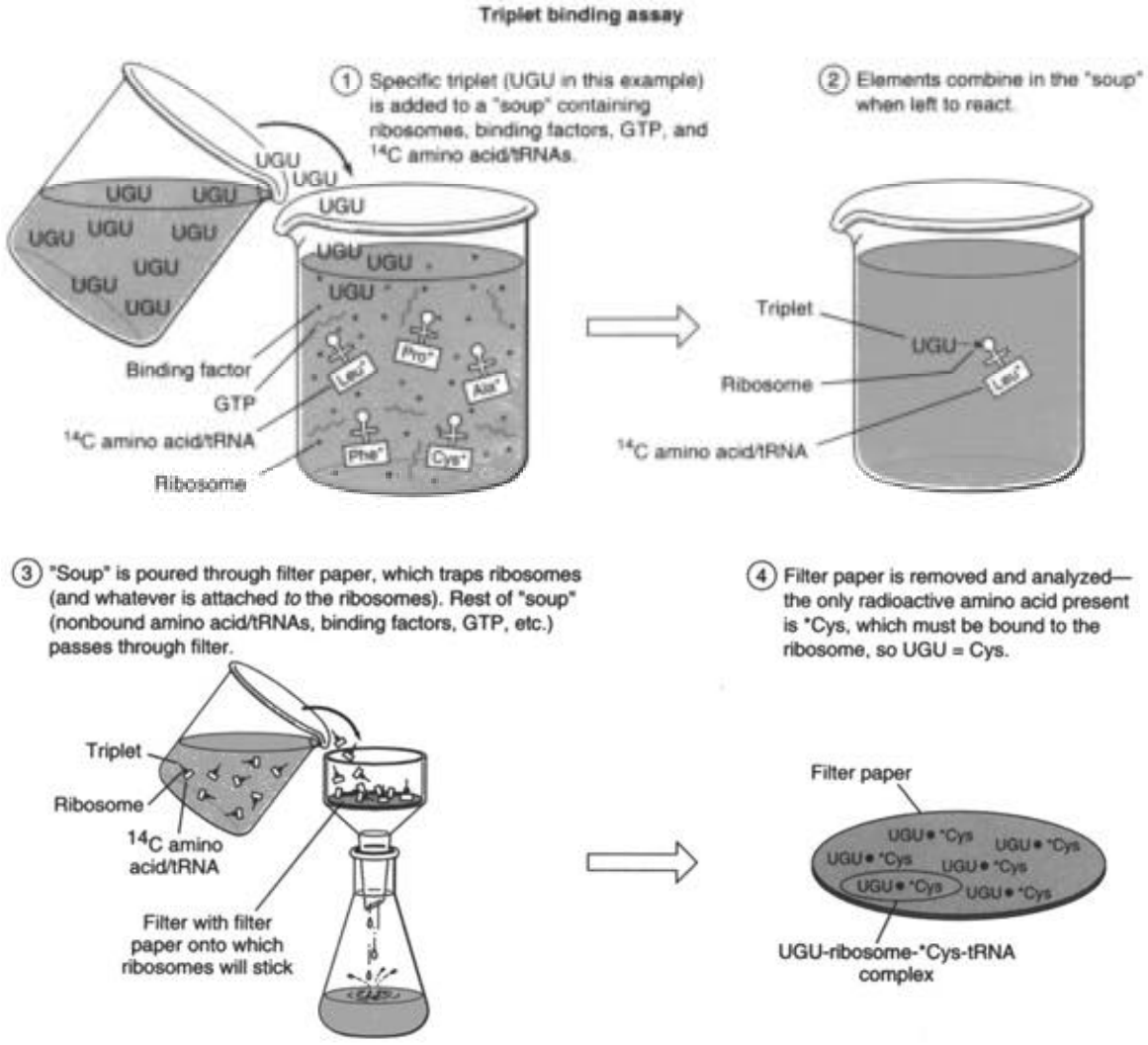

Figure 14.1
Nirenberg's experiment.

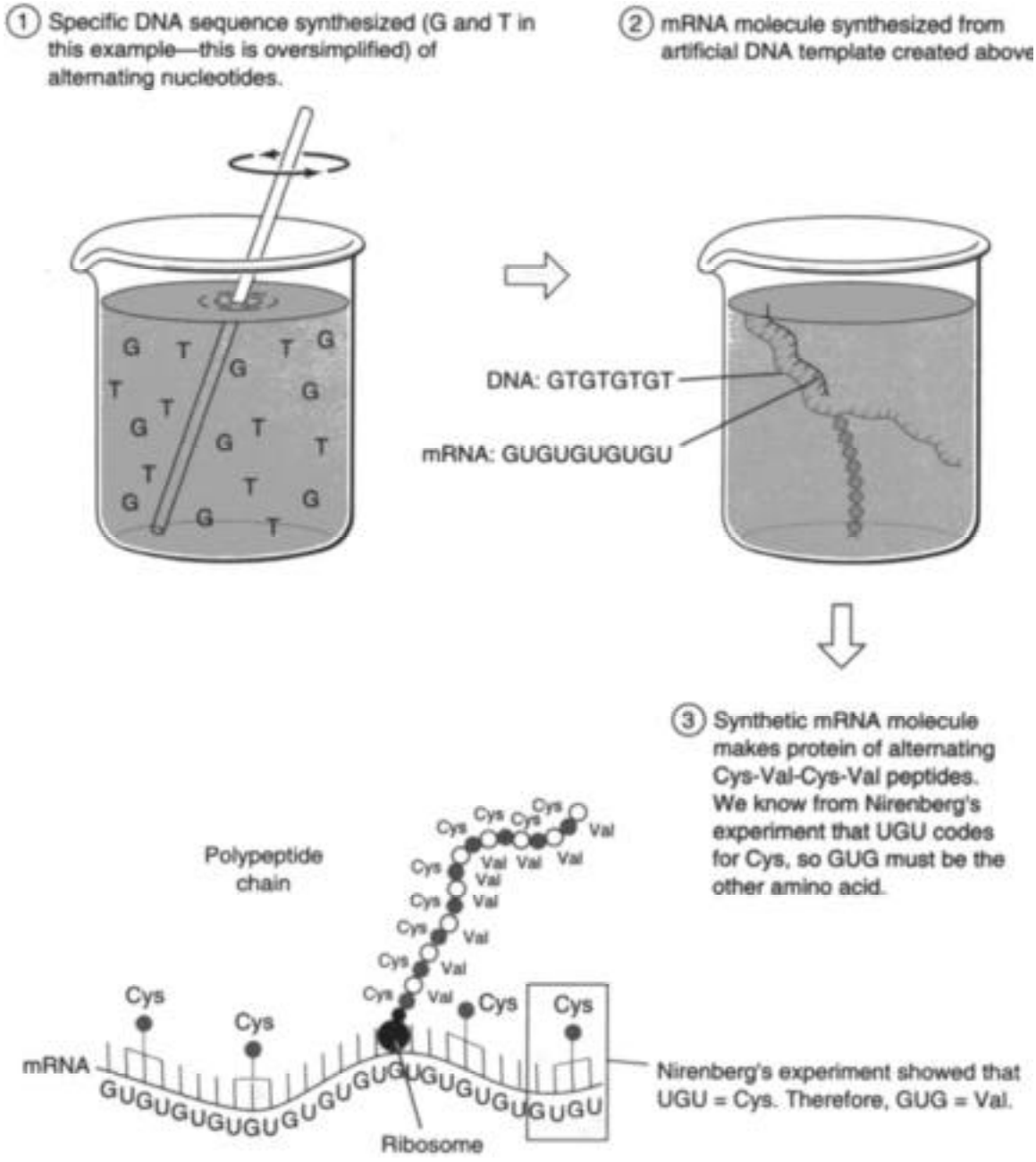

Figure 14.2 Khorana's experiment.

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

Figure 14.3
The genetic code.