CHAPTER 13

CRICK: THE GENETIC CODE IS READ THREE BASES AT A TIME

In 1961, Francis Crick and coworkers, in one of the best experiments anyone has ever done, demonstrated that the actual instructions for a protein exist as a series of overlapping, three-base code words, each "triplet" specifying one of the 20 amino acids.

THE GENETIC CODE HAS THREE DIGITS

It is one thing to understand that the genetic information encoded in DNA is translated via messenger RNA molecules into specific protein amino acid sequences, and quite another to understand how the trick is carried off. Is there one-to-one correspondence between a DNA base, an RNA base, and an amino acid? Clearly not, as there are 20 amino acids and only four types of nucleotide bases. A code of some sort has to exist to get 20 amino acids—some *sequence* of nucleotide bases must encode the information for an amino acid. Groups of two-base sequences would not do, as there are too few possible combinations (4^2=16), so attention immediately focused on the possibility that the DNA code was a three-digit code: that DNA code words specifying specific amino acids are made up of three nucleotide base groups.

DO THE CODES OVERLAP?

Within a few years of the Watson-Crick model, a logical hypothesis of DNA coding had been advanced by the physicist George Gamow, who suggested that the RNA polymerase read three-base increments of DNA while moving along the DNA one base at a time. The polymerase would therefore "read" the DNA in overlapping units. Such an *overlapping code* hypothesis was attractive because it could be tested. It predicted that certain bases should not occur side-by-side in nature (or else one triplet base sequence could code for more than one amino acid), and a study of protein amino acid sequences to see which combinations do not occur should eventually lead to a deciphering of the code and an understanding of which triplets code for which amino acids.

When amino acid sequences were examined, however, there was little evidence of forbidden two-base combinations. Also, analysis of the amino acid sequence of "mutant" proteins produced a result even more damaging to Gamow's hypothesis: a single mutation typically produced a protein with only a single amino acid different from normal, while an overlapping code would predict that three adjacent amino acids should be altered by single base change.

It seemed, then, that the DNA code was read in nonoverlapping segments, presumably of three digits (4^3=64, which is more than ample to code for 20 amino acids). There were in principle two ways in which such readings could be carried out: 1. Punctuation could be used between each three-base code word—a "*comma* code," or 2. Reading could be initiated from a fixed point in units of three bases—a "*reading frame* code." Each hypothesis was (and is) reasonable, and it was seven years before a clear experiment was devised to choose between them.

CRICK'S EXPERIMENT

The key experiment was carried out by Francis Crick and coworkers in 1961 (figure 13.1), and hinged upon the hypothetical continuous nature of a reading frame code. If a base was deleted (or added) to a nucleotide

sequence, then the reading frame code hypothesis would predict a disruption of proper read-out downstream. In other words, the reading frame would be shifted by one base, resulting in all subsequent triplet combinations being erroneous, while a comma code hypothesis would predict only a single amino acid change.

Mutagens that appeared to delete (and/or add) bases are known. Proflavin and other acridine dyes bind DNA in such a way as to "interlocate" the dye between adjacent bases of a DNA strand. This interrupts proper base-paring between strands and results in the "kink" being removed by deleting a nearby base on that strand, or by adding a base to a nearby region of the opposite strand to compensate.

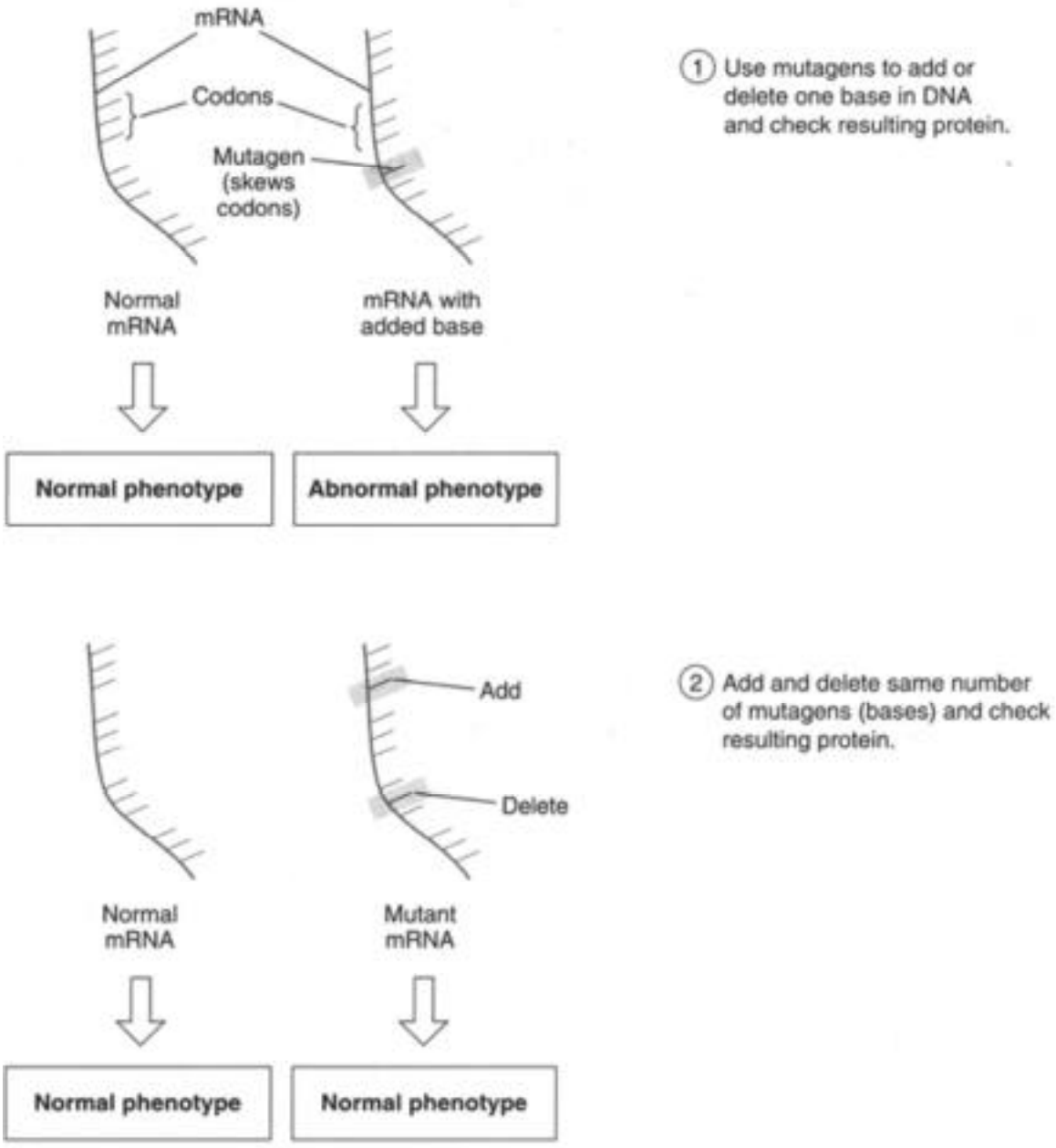

Figure 13.1
Crick's 1961 experiment.

Crick and coworkers examined such mutants in T4 viruses, and showed that while base addition or base deletion gave a mutant phenotype, a combination of a single base addition and single base deletion near to one another on the DNA always produced a normal phenotype! This result, on the face of it, disproved the comma code hypothesis and established that the genetic code is indeed a reading frame code, with code reading starting from some fixed point.

They went on to show that the code words had three digits. Combinations of two base deletions or two base additions were still mutant, but combinations of three different single base deletions or three different single base additions gave a wild-type phenotype. This could only mean that the third deleted (or added) base restored the original reading frame! This proved beyond question that the code words occurred as multiples of three nucleotide bases.