

 MOLECULAR BIOLOGY

Nature of the genetic code finally revealed!

This December marks the fiftieth anniversary of the publication of the landmark article by Francis Crick *et al.*, describing the nature of the genetic code. At that time, many fundamental questions in molecular biology were yet to be answered. The central dogma, that protein sequence is determined by DNA sequence, had been articulated by Crick in 1958, but how the information was encoded was unknown. In their paper, Crick *et al.* demonstrated that the codons which encode the individual amino acids are non-overlapping and consist of three bases.

Amazingly, this fundamental finding required just two strains of *Escherichia coli* — a K strain and a B strain — and mutants of bacteriophage T4. Wild-type T4 grows on both *E. coli* strains, but a mutant lacking cistron B grows only on the B strain. To mutagenize T4, the authors used acridines, which were thought to be likely to remove or add a single base. Mutagenesis of the T4 mutant FC0,

which carries a mutation in the B cistron, produced suppressors, bacteriophages that could once again infect a K strain. Combining the suppressor mutations showed that a base was either added (+ mutation) or removed (– mutation) and that a + mutation could rescue a – mutation, and vice versa. Hence, the authors had identified frameshift mutations, in which the addition or deletion of a base in the coding DNA sequence causes a frameshift for the translation machinery, leading to a non-functional protein. Correction of the reading frame by an additional frameshift mutation allowed a functional protein to be produced. The authors noticed that this worked well with mutations that were closely linked, but not with mutations that were further apart. They speculated that frameshifting introduces ‘unacceptable’ or ‘nonsense’ codons, which we now refer to as stop codons. This confirmed the previously held notion that the reading frame was non-overlapping.

However, the size of the codons remained unknown. The authors found that a combination of three+ or three – mutations, but not two+ or two – mutations, could restore the wild-type phenotype. On the basis of these findings, the authors concluded that a codon was most likely to be three nucleotides long. With a triplex codon made up of four potential bases, 64 possible codons can be formed, which is more than enough to cover the 20 standard amino acids; thus, the authors also identified the redundancy in the genetic code.

But the nature of the code itself remained a mystery. In their conclusions, the authors stated, optimistically, that “If the coding ratio is indeed 3, as our results suggest, and if the code is conserved throughout Nature, then the genetic code may well be solved within the year”. At the time, only a single codon could be assigned (namely, the phenylalanine codon, UUU), as earlier in the year Marshall Nirenberg had shown that a continuous string of uracils was translated into a string of phenylalanines. But by 1966 a complete codon table was produced, based primarily on the continued work of Nirenberg, who worked with cell-free extracts and chemically synthesized RNA, work for which he received the Nobel Prize in 1968.

Perhaps not much thought is given to these experiments today when translating a DNA sequence to protein, but this stands as a shining example of how basic, and elegant, microbiological techniques have provided insights into fundamental biological questions.

Christiaan van Ooij

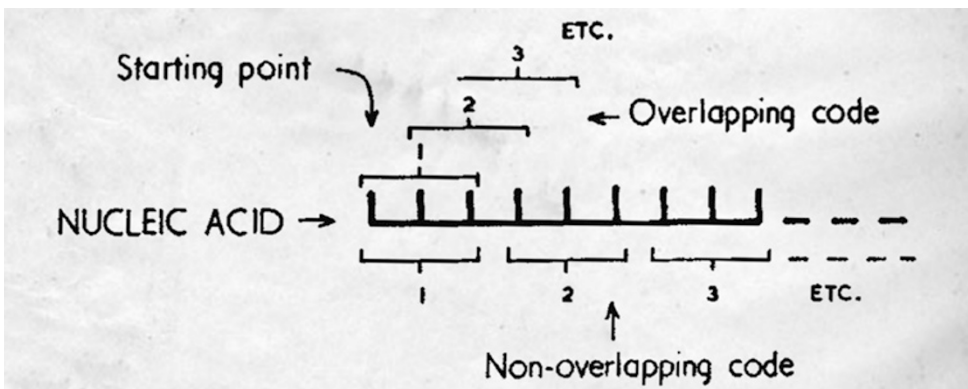


Figure 1 from Crick *et al.*, reproduced, with permission, from Crick, F. H. C. *et al.* © (1961) Macmillan Publishers Ltd. All rights reserved.

ORIGINAL RESEARCH PAPER Crick, F. H. C. *et al.*
General nature of the genetic code for proteins.
Nature **192**, 1227–1232 (1961)