

mRNA readout at 40

John F. Atkins and
Raymond F. Gesteland

On 30 December 1961, 40 years ago, Francis Crick and colleagues published the results of a remarkable experiment that established without doubt the general features of the genetic code for protein synthesis. It was already known that the amino-acid sequence along the polypeptide chain of a protein is determined by the bases along the nucleic acid (DNA or RNA) of the genetic material, and that the code for this correspondence is non-overlapping along the base chain. Crick *et al.* established that a group of three bases encodes one amino acid, and that the chain of bases is 'read' in a sequential manner without punctuation from a fixed starting point.

These findings, in conjunction with the non-overlap, mean that a 'reading frame', set by the starting point, determines which particular sets of three bases specify the different amino acids (see figure) — a brilliant culmination of many years' work originating different coding schemes and trying to distinguish between them. In today's era of cloning, rapid sequencing and atomic-resolution structures, it is salutary that this cornerstone was put in place not by ordering a kit or sending samples to a sequencing centre, but rather through simple genetic crosses of a bacteriophage, combined with great insight in interpreting

the properties of the resulting mutants.

The extent to which certain phage mutants could be reverted to wild type by different mutagens revealed a class of 'deletion' (minus) and 'insertion' (plus) mutants. Crosses between representatives of the two classes (plus/minus or minus/plus) would give wild-type progeny because the reading frame must have been restored, whereas crosses within a class (plus/plus or minus/minus) gave no wild-type progeny — the reading frame remained disrupted (showing, incidentally, that the code was not based on two or a multiple of two bases). The crucial insight came from the discovery that a triple mutant within either class (minus/minus/minus or plus/plus/plus) could give wild type: the frame had been restored. Hence the reading unit, or codon, must be three (or a multiple of three) bases.

In the 40 years since this dramatic discovery, the detailed mechanism of the decoding of the non-overlapping sequential triplets is still not understood, despite recent success in obtaining atomic-level structural information about the subunits of the ribosome. The ribosome is a gigantic complex of more than 50 proteins and RNA, the latter having the primal role in the decoding process. Understanding the dynamics between the ribosomal RNA, transfer RNA and messenger RNA (mRNA) that establish and maintain reading frames in the ribosome remains a challenge today.

Evolutionary variants of triplet readout that exploit alterations of reading frames, fascinating in their own right, can also provide clues to how the standard process works so reliably. In the decoding of a minority of genes, signals in the mRNA dictate that a proportion of translating ribosomes shift, at a specific site, into one or other of the alternative two reading frames (see figure). The resulting trans-frame protein product is often produced in a set ratio to that synthesized by standard decoding. One example of such a product is the GagPol precursor of HIV and many other retroviruses.

In other cases, the programmed frameshifting occurs early in the readout process, providing a regulatory mechanism in which the efficiency of the shift responds to cell physiology; that is, it acts as a 'sensor'. Antizymes, negative regulators of polyamine levels, are an example of this process.

Nearly all the frameshifting used for gene expression involves moving the reading frame one base either 'backwards' or 'forwards' on mRNA. In rare cases, a constellation of signals causes decoding to resume after a block of nucleotides are traversed without specifying any amino acid. Signals in mRNA specify the resumption site. The current record distance

Genetic code

The importance of the concept of non-overlapping triplet reading of the genetic code, proved 40 years ago, remains huge.

efficiently 'bypassed' this way is a 50-nucleotide non-coding sequence. There is even a case in all bacteria where the resumption of decoding after interruption is on a different, specialized RNA.

These gymnastic feats are not the only way in which standard genetic readout is locally reprogrammed. Signals in mRNA can redefine 'stop' codons to specify an amino acid. Sometimes, this is to permit ribosomes to access coding sequence from further on, in which case the identity of the amino acid specified by the redefinition is unimportant. In other cases, redefinition is used to specify selenocysteine, the twenty-first directly encoded amino acid.

Apart from these cases of code extension where standard readout is dynamically and temporarily overridden (recoding), the standard reading frame is maintained with high accuracy. Errors in frame maintenance have a high cost, as they result in truncated proteins. Truncated proteins also derive from substitution mutants that create a premature stop codon: this type of mutant is responsible for about a quarter of human genetic diseases. Nevertheless, errors that give read-through of a premature stop codon, and certain framing errors near a frameshift mutation can lead to a small amount of full-length protein. The low efficiency of read-through of a stop codon can be increased by aminoglycoside antibiotics that decrease the fidelity of protein synthesis. The potential of this process for ameliorating genetic disease is now being assessed in clinical trials.

The importance of the concept of the principles of genetic decoding, proved 40 years ago, remains huge. Even when current advances in molecular biological and structural tools reveal the dynamics of its mechanism in ribosomes, the permutations will continue to challenge and surprise us. ■

John F. Atkins and Raymond F. Gesteland are in the Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112-5330, USA.

FURTHER READING

- Crick, F. H. C. *et al.* General nature of the genetic code for proteins. *Nature* **192**, 1227–1232 (1961).
- Crick, F. *What Mad Pursuit. A Personal View of Scientific Discovery* Ch. 12, Triplets (Basic Books, New York, 1988).
- Judson, H. F. *The Eighth Day of Creation* (Simon & Schuster, New York, 1979).
- Atkins, J. F. *et al.* Dynamics of the genetic code. In *The RNA World* 2nd edn (eds Gesteland, R. F. *et al.*) 637–673 (Cold Spring Harbor Laboratory Press, New York, 1999).

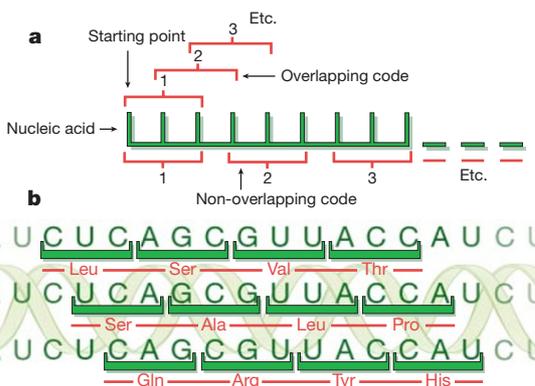


Figure 1 Reading frames. a, The difference between an overlapping code and a non-overlapping code. The short vertical lines represent the bases of the nucleic acid. The case illustrated is for a triplet code. b, Three possible reading frames in protein synthesis. A sequence of nucleotides in RNA is read in sequential sets of three nucleotides and is thereby translated into amino acids. The same RNA sequence can therefore specify three completely different amino-acid sequences, depending on the 'reading frame'.