Exons encode protein functional units

from C. C. F. Blake

AN important contribution to relationship between protein structure and the newly discovered mosaic structure of eukaryotic genes, in which coding segments of DNA (exons) alternate with non-coding segments (intervening sequences or introns), is reported by Tonegawa and his colleagues in this week's *Nature* (Sakano *et al.* page 627). The protein chosen was the constant region of the immunoglobulin γ heavy chain. Sequence studies and crystallography have already shown that the constant region is composed of three homologous structural domains: the CH_1, CH_2 and CH_3 domains. Each domain is composed of about 110 amino acids, folded into almost identical tertiary structures made up from seven β-strands. The CH_1 domain pairs with the C_L domain of light chains to form, together with paired V_H and V_L domains, the Fab parts of the γ-globulin molecule. The CH_2 and CH_3 domains pair with equivalent domains of a second heavy chain to form the F_c fragment, which is separated from the Fab fragment by a hinge region. Each of these structures has a well defined function: the paired CH_3 domains act in cell surface interactions, the paired CH_2 domains in complement fixation; the hinge probably transfers information from the Fab to the F_c fragment; the CH_1 pair is the attachment point for the light chain; the V_H V_L pair forms the antigen binding site; and there is also an N-terminal signal peptide. Thus the immunoglobulin molecule is composed of six discrete structural/functional units whose relationship to the exon/intron mosaic in the coding region could be most illuminating.

The DNA was obtained from mouse MOPC 21 myeloma by complete digestion with *EcoRI* nuclease and the identification of a 6.8 kilobase segment that contained the required sequence by hybridisation with cDNA prepared from MOPC $21_{γ1}$ mRNA. The segment was recombined with the EK2 vector, phage λgt WES, and cloned. After digestion of the clone the position of the $γ_1$ coding region in the 6.8 kb fragment was determined by electron microscopy of the R-loops formed between it and the $γ_1$-chain mRNA. This revealed three small R-loops separated by two intervening double-stranded DNA segments, suggesting the presence of three exons interrupted by two introns. A restriction enzyme cleavage map showed that this interpretation was correct, but that there was an additional very short exon located between exons 1 and 2, giving a final pattern of 4 exons separated by 3 introns. Sequence analysis of the intron–exon boundaries and comparison with the known amino acid sequence of the MOPC $21_{γ1}$ chain, revealed that the exons represent, in sequence, the CH_1 domain, the 14-residue hinge region, the CH_2 domain, and the CH_3 domain. There seem to be no introns within the individual sequences coding for these units.

The conclusion from this study, that each of the six known functional units in the γ-globulin heavy chain is

coded by a separate exon is most exciting. So clear cut is the correspondence that one is tempted to conclude that the problem of exons has been solved along the lines proposed by Gilbert (*News and Views* **271**, 501; 1978) in which the presence of exons was seen as allowing new proteins to be constructed from parts of old ones. The results reported by Tonegawa are exactly in line with this hypothesis, and show that individual exons can define even very small functional units, like the 14-residue hinge region, and not just domains or folding units as I previously suggested (*News and Views* **273**, 267; 1978). In view of the precise correlation between exons and the functional units in the immunoglobulins, the presence of three exons in the β-globin gene, when the protein cannot readily be divided into individual units of either structural or functional significance seems at first very puzzling. However, closer examination shows that the central exon corresponds very closely to the protein fragment that defines the haem crevice, and contains nearly all the amino acid residues that interact with the haem. Tonegawa's work implies that exons may reveal unsuspected functional protein units—perhaps like the 'minimal globin' outlined above—even when the joins are not visible in the protein structure. The thought that the exon/intron mosaic may carry the frozen evolutionary history of proteins makes one look forward to see whether further determinations of exon–intron patterns of other proteins of known structure support this view, or point in a different direction.

C. C. F. Blake is in the Laboratory of Molecular Biophysics, Department of Zoology, University of Oxford.

RNA splicing and polynucleotide evolution

from D. Reanney

INFORMATION flow in cells, is defined by the 'central dogma' as
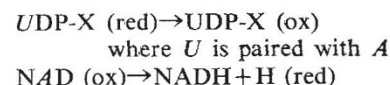
$$DNA \rightarrow RNA \rightarrow Protein$$

But although the central dogma accurately describes the flow of information in modern cells, it probably fails to reflect evolutionary perspective. Considered from the chronological viewpoint a more accurate representation of the central dogma might be

$$RNA \nearrow^{DNA}_{\searrow Protein}$$
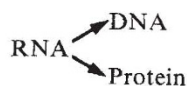
This rewriting of a basic tenet of

biology has come from several lines of evidence which strongly reinforce the view of Crick (*J. molec. Biol.* **38**, 367; 1968) and others that RNA is the primary genetic material. It is worth re-emphasising that most key features of the translation of genetic information depend on Watson–Crick pairing between (partly) complementary RNAs. Translation is mediated by interactions between different RNAs—mRNA: tRNA; mRNA:rRNA; rRNA:tRNA and rRNA:rRNA. The same was probably true of primitive 'metabolism'. With respect to a metabolic fossil such as the ribodinucleotide NAD, for example, one may ask why is a nucleotide 'handle' a necessary part of a

molecule whose active group is nicotinamide. I suggest that the first steps on the road to precellular metabolism were taken when electron-donating and electron-accepting reactions were coupled by means of molecules whose 'handles' were complementary in the Watson–Crick sense. For example

$$UDP\text{-}X \text{ (red)} \rightarrow UDP\text{-}X \text{ (ox)}$$
where U is paired with A
$$NAD \text{ (ox)} \rightarrow NADH + H \text{ (red)}$$

Here the 'genetic' monomers A and U are used to bring the two elements of the redox reaction into appropriate juxtaposition; this is effectively what an enzyme does. The addition of

D. Reanney is Senior Lecturer in the Department of Microbiology, La Trobe University, Australia.