news and views

Why genes in pieces?

from Walter Gilbert

OUR picture of the organisation of genes in higher organisms has recently undergone a revolution. Analyses of eukaryotic genes in many laboratories [1-10], studies of globin, ovalbumin, immunoglobulin, SV40 and polyoma, suggest that in general the coding sequences on DNA, the regions that will ultimately be translated into amino acid sequence, are not continuous but are interrupted by 'silent' DNA. Even for genes with no protein product such as the tRNA genes of yeast and the rRNA genes in *Drosophila*, and also for viral messages from adenovirus, Rous sarcoma virus and murine leukaemia virus, the primary RNA transcript contains internal regions that are excised during maturation, the final tRNA or messenger being a spliced product.

The notion of the cistron, the genetic unit of function that one thought corresponded to a polypeptide chain, now must be replaced by that of a transcription unit containing regions which will be lost from the mature messenger—which I suggest we call introns (for intragenic regions)—alternating with regions which will be expressed—exons. The gene is a mosaic: expressed sequences held in a matrix of silent DNA, an intronic matrix. The introns seen so far range from 10 to 10,000 bases in length; I expect the amount of DNA in introns will turn out to be five to ten times the amount in exons.

This model immediately accommodates two aspects of the genetic structure of higher cells. Heterogeneous nuclear RNA clearly is the long transcription products out of which the much smaller ultimate messengers for expressed polypeptide sequences are spliced. The unexpected extra DNA in higher cells, the excess of DNA over that needed to code for the number of products defined genetically, now is ascribed to the introns.

What are the benefits of this intronic/exonic structure for genes? For the sake of argument let us assume that the splicing mechanism is general and independent of the specific gene or the state of the cell, reflecting simply some secondary structure in the RNA. For example, base-pairing in the messenger could generate sites which would serve as signals for enzymes, such as those that excise tRNAs from their precursors, to cut out a section. The cut would be resealed by an RNA ligase. Even if RNA processing is general, the presence of infilling sequences can speed evolutiony

Single base changes, the elementar. mutational events, not only can change protein sequences by the alteration of single amino acids but now, if they occur at the boundaries of the regions to be spliced out, can change the splicing pattern, resulting in the deletion or addition of whole sequences of amino acids. During the course of evolution relatively rare single mutations can generate novel proteins much more rapidly than would be possible if no splicing occurred.

Furthermore, the splicing need not be a hundred per cent efficient; changes in sequence can alter the process so that base pairing and splicing occurs only some of the time. Even mutations in silent third base positions, could modify the joining so that the products of a single transcription unit can be both the original gene product and a new product, also synthesised at a high rate. Evolution can seek new solutions without destroying the old. A classic problem is resolved: the genetic material does not have to duplicate to provide a second copy of an essential gene in order to mutate to a new function. Rather than a special duplication, the extra material is scattered in the genome, to be called into action at any time. After a new gene function appears, if a higher level of product is needed, there will be selective pressure for gene duplication (as well as pressure for the loss of the introns in highly repeated genes). One consequence of the intronic model is that the dogma of one gene, one polypeptide chain disappears. A gene, a contiguous region of DNA, now corresponds to one transcription unit, but that transcription unit can correspond to many polypeptide chains, of related or differing functions.

Recombination now becomes more rapid. Since the gene is spread out over a larger region of DNA, recombination, which should be hampered in higher cells by the inability of DNA molecules to get together, will be enhanced. Furthermore, if exonic regions correspond to functions put together by splicing to form special combinations in the finished protein, then recombination within introns will assort these functions independently. Middle repetitious sequences within introns may create hot spots for recombination to rearrange the exonic sequences.

Recombination within introns will generate curious genetic structures for eukaryotic genes. Structural mutations should be clustered, separated by long distances from mutations in other exons. Mutations in different functions may be interspersed, when one product's intron becomes another's exon.

According to this view, introns are both frozen remnants of history and as the sites of future evolution. Nevertheless, they could also have other roles. Specific recombinations between introns can bring together exons into a transcription unit to make special differentiation products. Specific new splicing patterns could be turned on by special gene products. A differentiation pathway may be determined by the appearance of a new splicing enzyme, calling forth new proteins out of the heterogeneous nuclear RNA.

On this can be based a striking hypothesis to explain the behaviour of immunoglobulin heavy chains. At an early stage of the immune response a single lymphocyte can synthesise two different immunoglobulins, IgM and IgD, with the same idiotype; two different constant portions attached to the same V_H region. This may be the result of a V_H region translocating by recombination within an intron near the constant genes so that a trranscription unit is formed for a V_H-C_μ-C_δ message. Splicing can then create contiguous messenger sequences for V_HC_μ and V_HC_δ chains. The switch from IgM to IgG might be a new translocation of the V_H gene, but, alternatively, it may be a new enzyme that changes the processing of a V_H-C_μ-C_δ-C_γ message to produce a V_HC-product. □

Walter Gilbert is American Cancer Society Professor of Molecular Biology at Harvard University.

1. *News and Views, Nature* **268**, 101 (1977).
2. Berget *et al. Proc. natn. Acad. Sci. U.S.A.* 74, 3171 (1977).
3. Papers in *Cell* 12 (1) (1977).
4. Aloni *et al. Proc. natn. Acad. Sci. U.S.A.* 74, 3686 (1977).
5. Leder *et al. Cold Spring Harbor Symp. quant. Biol.* (in the press).
6. *News and Views, Nature* **270**, 295 (1977).
7. Breathmark *et al. Nature* **270**, 314 (1977).
8. Deol *et al. Nucleic Acids Res.* 4, 3701 (1977).
9. Jeffreys & Flavell *Cell* 12, 1097 (1977).
10. Brack & Tonegawa *Proc. natn. Acad. Sci. U.S.A.* 74, 5652 (1977).

0028-0836/78/0271—0501$01.00