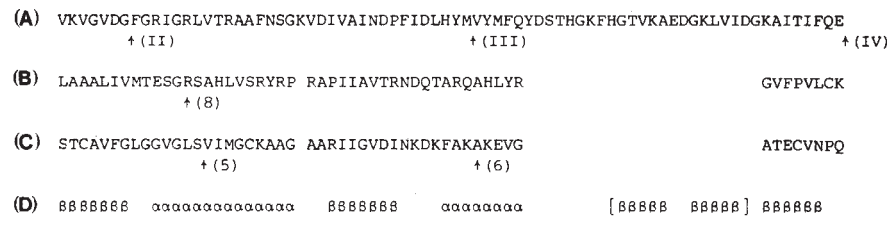
Eukaryotic genes

Are introns structural elements or evolutionary debris?

from Athel Cornish-Bowden

THE discovery of introns — sequences of non-coding DNA that interrupt the coding sequences of genes but are excised from gene transcripts — burst upon an unsuspecting world in 1977. Since then, much information has accumulated about the ubiquity of introns in the genes of higher eukaryotes and their absence from those of prokaryotes and lower eukaryotes, but it is still not really clear why they are there or what they do. There is still argument over whether they were present in the primordial genes and have been eliminated in the interests of efficiency by prokaryotes, or whether, instead, they were introduced after the eukaryotes separated from the prokaryotes. Papers in this issue of *Nature* and the January issue of *Cell* contribute to the debate.

It is scarcely tenable to propose that all introns are functionless, opportunistic pieces of selfish DNA which have invaded eukaryotic genes by taking advantage of the biochemical machinery that normally excises them. If so, why then did the excision machinery evolve in the first place? At least some introns, therefore, must have had a function, at some stage in the evolution of the gene. The most plausible suggestion is that the existence of introns makes it much easier for large and complex proteins to evolve, as they can be assembled from the small functional units or domains into which introns split genes. Moreover, new functions could in principle be produced by rearranging such domains. Once introns evolved, however, it became possible for parasitic DNA to take advantage of the excision machinery. Thus although some, perhaps most, introns must have a functional origin, it is not necessary to assume this for every intron.



Intron arrangement in the mononucleotide binding-site region of pig glyceraldehyde 3-phosphate dehydogenase (A), chicken pyruvate kinase (B) and horse alcohol dehydrogenase (C), together with the consensus secondary structure (D) in terms of α-helices β-pleated sheets. Amino acids are designated by the one letter code; the positions of introns in the corresponding genes are shown by arrows and numbered according to the authors.

On page 498 of this issue, Schwartz and co-workers suggest that both kinds of intron are recognizable in the glyceraldehyde 3-phosphate dehydrogenase gene of the chicken. They classify the introns as type A, which result from the original assembly of the gene from smaller units, and type B, which have never had any function. Moreover, they believe that type A introns can be further classified according to whether they derive from the creation of a domain from smaller units (Al), from the duplication of a complete domain (A2), or from the joining of two dissimilar domains (A3). Classification of the three introns that occur at or near the boundaries of the four recognized structural domains of the enzyme is straightforward: one is type A2, two are type A3. In addition, the catalytic domain contains a region that makes about 85 per cent of its hydrogen bonds within itself and which also corresponds to an exon — the part of the gene that is translated into protein. Thus, the two introns that flank this exon can be regarded as resulting from the joining of protein domains and so of type A3.

Apart from one intron in the non-coding region of the mRNA, all of the remaining six introns occur within domains and are more difficult to classify, because both A1 and B type introns would be expected in similar positions. Nonetheless, Schwartz and co-workers believe that sub-domains can be recognized within the domains, and that several of the remaining introns can be interpreted as separating them, so that they are of type A1. Only one intron in the coding region is assigned to type B.

This type of classification can be tested by examining the genes of comparable proteins. Types A2 and A3 introns should occur in other genes at domain boundaries; type A1 introns should occur at homologous positions within homologous domains; but type B introns should occur at arbitrary and unrelated positions. It is of interest, therefore, to examine the hypothesis of Schwartz and co-workers in relation to new work by Lonberg & Gilbert (*Cell* **40**, 81; 1985). These workers have determined the structure of the gene for chicken pyruvate kinase and compared it with that of the gene for maize alcohol dehydrogenase (Dennis E.S.*et al.*, *Nucleic Acids Res.* **12**, 3983; 1984). All three proteins contain a mononucleotide binding domain with alternating regions of β-pleated sheet and α-helix, though this domain in glyceraldehyde 3-phosphate dehydrogenase is considerably longer because it contains an extra piece of β-sheet (which does not interfere with the structure of the rest of the domain). The comparison shown in the figure, with arrows indicating the location of introns, is disappointing: intron II of the glyceraldehyde 3-phosphate dehydrogenase gene, classified as type A1 by Schwartz and co-workers, occurs at the boundary between β and α regions, but there is no intron at this boundary in either of the other two genes; intron III, on the other hand, suggested to be of type B, corresponds almost exactly to an intron of the alcohol dehydrogenase gene.

Examining the pyruvate kinase gene as a whole, it is clear that introns are not placed at random in the sequence, and at least two systematic characteristics are evident. The ten exons are of a much more uniform length than one would expect if the coding region were interrupted at random, and the introns tend to fall between discrete regions of secondary structure in the protein. Surprisingly, there is little tendency for introns to occur at domain boundaries: most noticeably, the boundary between domains B and A_2 occurs near the middle of exon 5, the longest exon in the gene. It would be appealing, though entirely speculative, to suggest that an intron at this boundary has been lost, because if it were present it would not only separate the two domains but would also make the exon size even more uniform.

Although there is now much to suggest that introns are an ancient relic of primordial genes, convincing proof must await the discovery of clearly corresponding intron arrangements in genes for that arose by duplication before the separation of prokaryotes and eukaryotes. □

Athel Cornish-Bowden is in the Department of Biochemistry, University of Birmingham, PO Box 363, Birmingham B15 2TT, UK.