Related genes can have unrelated introns

from Athel Cornish-Bowden

IN eukaryotic gene families, such as those that code for the globins, the number and position of the introns has been found to be conserved across each family and, presumably, during evolution. A very different picture emerges from new work reported in this issue of *Nature* (p.655) by Woo and his collaborators. They have studied the introns in the genes for two distantly related proteins, human α_1-antitrypsin and chicken ovalbumin, and have found that there is no correspondence between the three introns of the ovalbumin gene and the seven of the α_1-antitrypsin gene. This result calls into question the idea that the arrangement of introns provides a record of the original structure of a eukaryotic gene; instead it suggests that introns may have been introduced into genes since the divergence of the vertebrates.

Ovalbumin is the principal protein in chicken egg white, whereas α_1-antitrypsin is a human plasma protein involved in the control of elastase; individuals deficient in it have a high risk of lung disease. There was no reason to expect any sequence similarity between the two proteins and the observation that they are in fact 24 per cent identical was thus a surprise. The degree of similarity is far too high to be ascribed to chance, as Hunt and Dayhoff (*Biochem. biophys. Res. Commun.* **95**, 864; 1980) estimated that such a possibility would occur with a probability around 10^{-41}.

If chance is excluded, there are two ways in which the similarity between ovalbumin and α_1-antitrypsin can be explained; either the two proteins have diverged from a common ancestral protein, estimated by Hunt and Dayhoff to have existed about 500 million years ago; or they have converged to similar sequences as a consequence of selection. In the former case, the lack of correspondence between the arrangements of introns in the two genes needs to be explained; in the latter it does not because no correspondence would be expected. Woo and collaborators say that it is impossible to determine conclusively whether the two proteins arose by convergent or divergent evoluion. I believe them to be unduly cautious in this regard and that the possibiiity of convergent evolution can be discounted.

Although examples of convergent evolution of protein structures do exist, they are observed at the level of catalytic mechanism and not in amino acid sequence. In the best established case, that of the bacterial proteinase subtilisin and the mammalian serine proteinases such as chymotrypsin, the catalytic mechanisms are essentially the same but there is no sequence similarity beyond a use of the same kinds of residue in the catalytic action. By contrast, if the similarity between ovalbumin and α_1-antitrypsin arose by convergent evolution it would imply convergence to a high degree of similarity over the whole lengths of two proteins without even, apparently, a similar physiological function. If this were substantiated its shattering effect on current ideas about protein structure and evolution could hardly be exaggerated; it would be comparable with the effect of the first reports of introns in eukaryotic genes and overlapping genes in viruses on our understanding of gene structure.

The chicken ovalbumin gene contains seven introns, all of them located in the 5′ half of the mRNA, whereas the three introns in the human α_1-antitrypsin gene are all in the 3′ half of the mRNA. Only one intron in the ovalbumin gene occurs in even approximately the same position as any intron in the α_1-antitrypsin gene: the introns in question do not show significant sequence similarity either with one another or with the corresponding coding regions of the other genes. It seems inescapable that a substantial reorganization of the intron structure has occurred since the time of the common ancestral gene shared by the ovalbumin and α_1-antitrypsin genes.

Could the differences in intron structure be a consequence of random losses of introns during evolution from an ancestral gene containing at least ten introns? If there were exactly ten, of which a random three were deleted on the lineage leading to the ovalbumin gene, and a random seven were deleted on the lineage leading to the α_1-antitrypsin, the probability that the two descendent genes would have no introns in common would be 3!7!/10!, or 1/120. This probability is hardly small enough to rule out random deletion of introns as an explanation, but it is small enough to suggest that alternative explanations should be considered, not only in this case but also in that of the actin gene family; for which a much smaller degree of intron variability has been observed. Woo and collaborators suggest that some at least of the introns may have been introduced into the ovalbumin and α_1-antitrypsin genes after they diverged in evolution. If so, the arrangement of introns must be a more

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

0028-0836/82/250625-02$01.00

recent aspect of gene structure than has been supposed. Light will perhaps be shed on this question by investigation of the intron structure of the gene for human antithrombin III, another plasma proteinase inhibitor in the super-family that contains ovalbumin and α_1-antitrypsin. □

stored in cytoplasmic vesicles are inserted