GENETIC RECOMBINATION IN THE EVOLUTION OF PROTEIN MOLECULES

H. K. JAIN

Botany Division, Indian Agricultural Research Institute, New Delhi

ON the classical theory of gene, genetic recombination is expected, at best, to have a limited role in the evolution of protein molecules. This theory postulates that recombination is restricted to between genes; on such a basis, crossing over would merely alter the sequence of coding units in the recombinant chromosomes. It would not result in the formation of new coding units. This, in turn, means that the protein molecules, information for whose synthesis is provided by individual genes, would not be affected at all in their structure. However, if there are exceptions to the one gene one polypeptide synthesis, some of the protein molecules may differ, as a result of recombination, from the parental molecules, in the sequence but not in the type and frequency of their different amino-acids.

The classical concept of the gene has been revised in recent years following the discovery of intragenic recombination, both in the eukaryotes showing chromosomal type organisation of their genetic material (*e.g., Drosophila* and *Aspergillus*) and other organisms in which a DNA molecule corresponds to the chromosome. The attempts to interpret complex loci in *Drosophila* within the framework of the classical concept of the gene continue to present difficulties (*see* Lewis, 1951). The main difficulty arises from the failure of the so-called pseudoalleles to show complementation. For this reason, such alleles are best considered to involve mutation at different sites of the same gene, rather than in closely linked but independent units, arising through duplication. Even with recombination of the intragenic type, however, only changes of sequence are expected to be produced, unless the point of cross-over is located within a coding unit (codon) which is now widely accepted to be made up of 3 nucleotides (Crick, 1963; Lanni, 1964). If it is supposed that in a short segment of the *DNA* molecule (say 18 nucleolides), a potential point of cross-over cannot discriminate between different positions, it is just as likely to give intracodon recombination as intercodon. In fact, on a purely random basis, the chance for the former type is greater by a factor of more than 2. While there is considerable evidence in the case of higher organisms to show that the points of cross-over are not randomly distributed (Darlington, 1935, 1937), the localisation appears to determine the range rather than the exact point of crossing over (Mather, 1938). There is no evidence at all to suggest that it operates at a molecular level. In micro-organisms also, non-random distribution of points of recombination is suggested in some cases by the phenomenon of negative interference (Pritchard, 1960), but here again no evidence is available for localisation at a molecular level.

The expectation on the occurrence of intracodon recombination finds striking support from the work of Henning and Yanofsky (1962). These authors have described experimental evidence for recombination of this type in *Escherichia coli*. Their elegant studies have shown that the recombinant triplets code for an amino-acid is different from that coded by the parental codon. It is obvious that genetic recombination at this level becomes indistinguishable from the process of gene mutation: and may have evolutionary implications. which the classical theory seemed to rule out.

These implications deserve attention, particularly in view of the evidence that the mechanism of recombination in higher and lower organisms may be basically alike. The partial chiasmatype hypothesis of Darlington (1937) postulates that an important step in the process of recombination is the breakage of the parental chromatids. This theory has found striking support from a number of studies (*e.g.*, Creighton and McClintock, 1931). It has also been shown by Meselson and Weigle (1962) and others that recombination in viruses involves breakage of the DNA molecules.

In view of these facts and the existence of interallelic recombination in both groups, it may be supposed that intracodon crossing over can occur in all sexually reproducing organisms. An important point for the present consideration is whether its magnitude can be such as to ensure for it a significant role in producing gene mutations.

Intracodon recombination, for a particular triplet, must necessarily be extremely rare, for nothing can be more closely linked than adjacent nucleotides. However, if recombination at the nuclear level is considered, a number of coding units can be expected to be affected in each of

the cells showing meiotic cell division, or a process corresponding to it. It has been seen that on a random basis, the chance for a cross-over point to be within a condon is twice as great as for it to lie between such units. The total number of such points in higher organisms can be readily determined by counting the number of chiasmata, which in favourable materials have been shown to have a one-to-one relationship with exchange of chromatids (see Brown and Zohary, 1955 ; Jain and Basak, 1963). In general, a median chromosome pair of average size, forms a minimum of 2 chiasmata, one in each arm. An organism having 12 such pairs, which is not a very large number, would in this way have 24 cross-over points in its chromosomes. In the absence of localisation at the molecular level, 16 of these can be expected to give intracodon recombination.

If all of these 16 cross-over points resulted in the formation of new coding units, recombination should constitute a significant source of new triplets. It is obvious, however, that not all of them would be effective. For them to be effective, the two parental codons involved must differ from each other in respect of at least two of their three nucleotides. The relative frequency of recombinant and parental triplets which would result from different pairs of codons of this type is shown in Fig. 1. The cross-over point has been shown to be randomly distributed over the two locations, and no double cross-overs have been considered in view of their improbability. The parental and recombinant types of nucleotide sequences which correspond to some of the hypothetical triplets shown here are presented in Fig. 2. In those cases, where the two parental triplets differ in all of their three nucleotides, the codons formed as a result of crossing over would be wholly of the recombinant type.

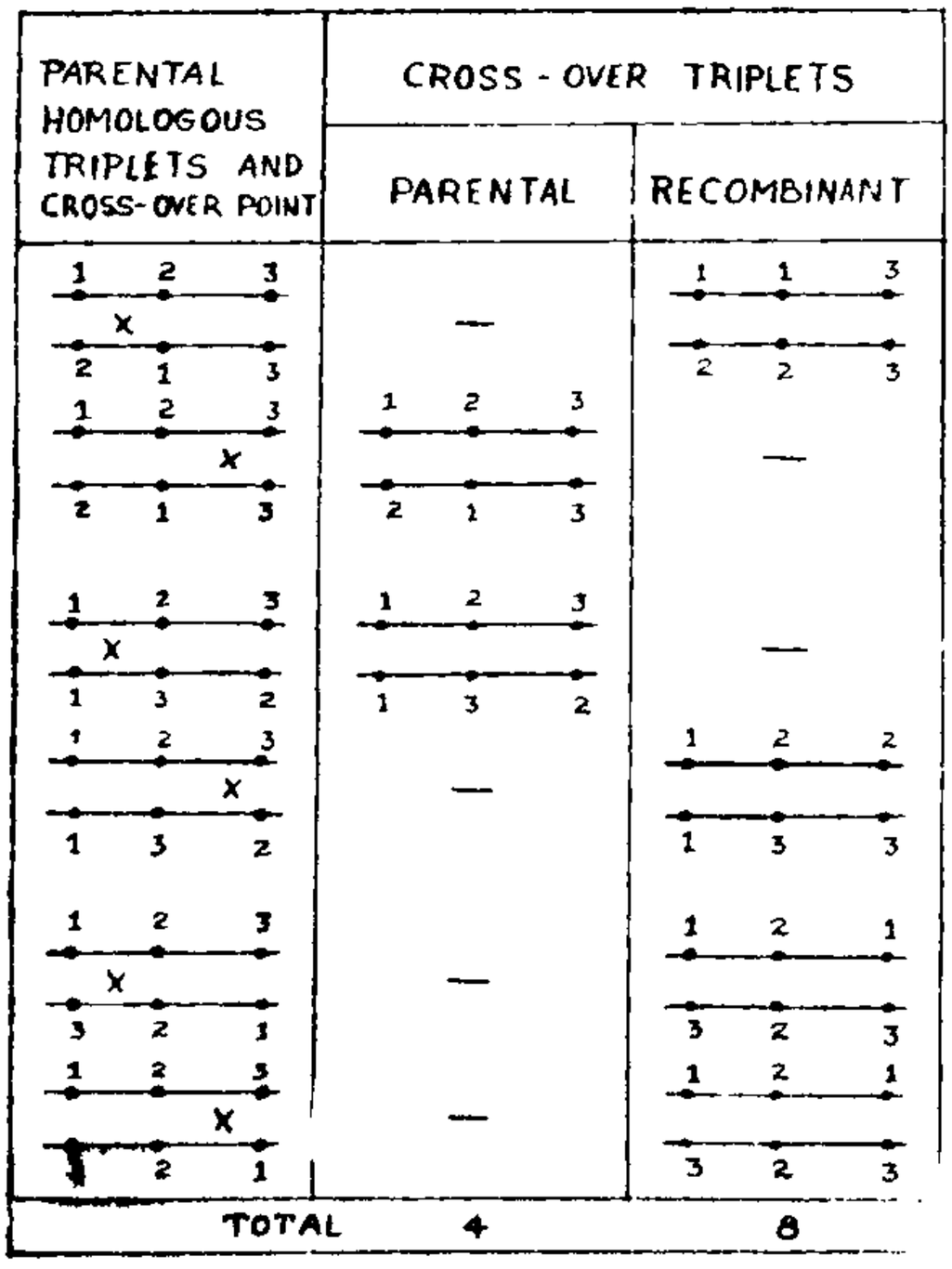

FIG. 1. Shows the expected formation of new types of triplets, when intracodon recombination involves coding units, which differ from each other in two of their three bases at various positions. The numbers correspond to different hypothetical nucleotides.

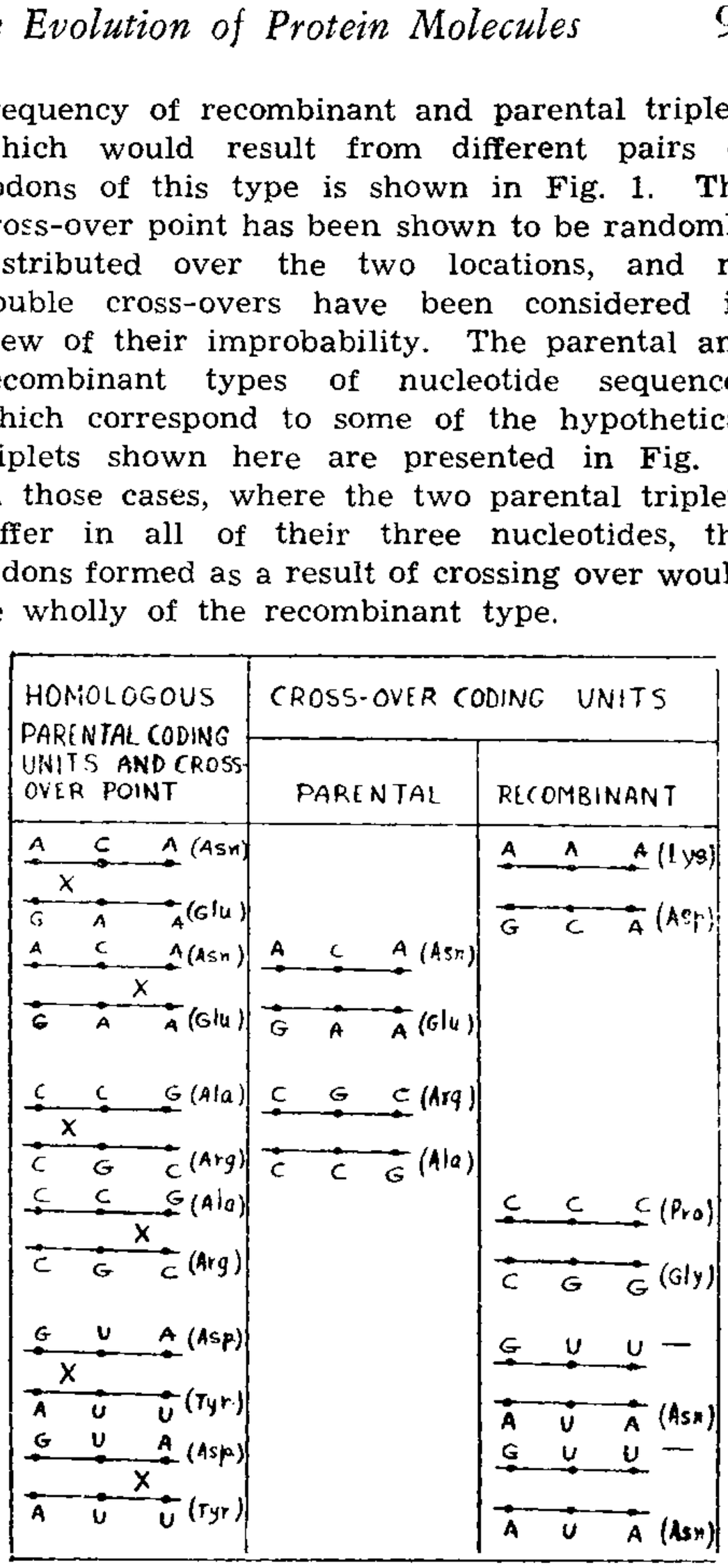

FIG. 2. The parental and recombinant coding units have been selected to correspond to the hypothetical triplets of Fig. 1. It should be made clear that the coding units shown here are of the messenger RNA; crossing over would involve them only indirectly, through the corresponding DNA segments, priming their synthesis. The amino acids for which they code have been shown in parenthesis. The nucleotide sequence in the units is arbitrary (after Nirenberg *et al.*, 1963).

Since the base composition of DNA in the genes cannot be analysed at present, it is not possible to determine the extent of such differences for the coding units in the different alleles of a gene. Theoretically, however, the difference need not be very great. It has been shown by Ingram (1963) that the hæmoglobin molecules, conditioned by the sickle cell mutant gene and its wild type allele, differ in respect of only one of their amino-acids. On the basis

of our present knowledge of the genetic code, it is possible to suggest that the two alleles may differ from each other in respect of no more than a single triplet out of the 300 or more which they may have. If this conclusion is correct, and if differences of this order are common for alleles at most loci, it would follow that the 16 intracodon points of recombination, randomly distributed along chromosome length, would have little or no chance of producing new coding units. A number of considerations, however, suggest that the inter-allelic differences in respect of the nucleotides may be of a greater magnitude.

The most important consideration in this respect is the degeneracy of the code. The code is recognised to be degenerate (*see* Crick, 1963 for a critical review) with many of the amino-acids having more than one triplet coding for them. This means that amino-acid differences, of the type found in the two hæmoglobin molecules, cannot be a reliable indication of the difference at the level of nucleotides in the corresponding DNA segments, which may be considerably greater. *Mutational* alterations of nucleotides in alleles of a gene can be expected to arise throughout the evolutionary history of the species, and such of these as have no immediate genetic consequences, because of degeneracy and other factors, would be preserved.

In this context, it is also relevant to consider the phenomenon of isoallelism. The isoalleles may be expected to differ in several of their nucleotides, the difference being such that it does not affect the corresponding enzymes at critical sites, which would, therefore, show functional similarity. An example of alleles showing such similarity is provided by one of the mutants of the *r*II locus in phage T_4. The mutant 1589 is known to involve a deletion which extends over parts of two adjacent cistrons A and B (Benzer, 1962). It is observed that only the A gene behaves as a mutant; the B gene retains much of its wild type activity. The occurrence of isoalleles shows that many homologous chromosome segments, which appear similar in their genetic effects, may be having nucleotide differences offering possibilities for intracodon recombination to be effective.

The above considerations suggest that the possibility of intracodon recombination giving rise to the formation of new non-parental type of coding units is not as negligible as it may appear at first sight. It may be concluded that at least a few of the 16 points of cross-over, in the example considered above, may give rise to units of this type, particularly over a number of generations of sexual reproduction. This in turn means that a few of the several thousand enzymes and other protein molecules, which an individual carries, may be different from those of its parents, both because of spontaneously arising mutations and also because of the process of intracodon recombination. It would appear that of these two sources of variability, the one based on recombination has a more definite basis. While mutations for the most part arise irregularly through factors which remain obscure, intracodon recombination can be expected to arise regularly in all the thousands of cells giving rise to gametes in the course of sexual reproduction. Another important consideration is that neither gene mutations nor recombinant codons may give rise to amino-acid changes, if the DNA segment involved has a function other than that of coding for them. The existence of regulatory genes in *E. coli* (Jacob and Monod, 1961) and other organisms (Ames and Hartman, 1963) is of obvious interest in this context. However, the nature of the regulatory materials remains obscure, and indirect evidence, based on complementation studies, indicates that a repressor may also be a protein.

1. Ames, B. N. and Hartman, P. E., *Cold Spring Harbour Symp. Quant Biol.*, 1963, **28,** 349.
2. Benzer, S. and Champe, S. P., *Proc. Natl. Acad. Sci. U. S.*, 1962, **48,** 1114.
3. Brown, S. W. and Zohary, D., *Genetics*, 1955, **40,** 850.
4. Creighton, H. B. and McClintock; B., *Proc. Natl. Acad. Sci. U. S.*, 1931, **17,** 492.
5. Crick, F. H. C., *In Progress in Nucleic Acid Research*, Ed. Devidson and Cohn, Academic Press, 1963, **1,** 163.
6. Darlington, C. D., *J. Genet.*, 1935, **31,** 185.
7. —, *Recent Advances in Cytology*, 2nd Ed., Churchill Ltd., London, 1937.
8. Henning, U. and Yanofsky, C., *Proc. Natl. Acad. Sci. U. S.*, 1962, **48,** 1497.
9. Ingram, V. M., *The Haemoglobins in Genetics and Evolution*, Columbia University Press, New York and Loudon, 1963.
10. Jacob, F. and Monod, J., *Cold Spring Harbour Symp. Quant. Biol*, 1961, **26,** 193.
11. Jain, H. K. and Basak, S. L., *Genetics*, 1963, **48,** 329.
12. Lanni, F., *Adv. Genetics*, 1964, **12,** 1.
13. Lewis, E. B., *Cold Spring Harbour Symp. Quant. Biol.*, 1951, **16,** 159.
14. Mather, K., *Biol. Rev.*, 1938, **13,** 252.
15. Meselson, M. and Weigle, J. J., *Proc. Natl. Acad. Sci. U. S.*, 1961, **47,** 857.
16. Nirenberg, M. W., Jones, P. L., Clark, B. F. C., Sly, W. S. and Pestka, S., *Cold Spring Harbour Symp. Quant. Biol.*, 1963, **28,** 495.
17. Pritchard, R. H., *Genet. Res.*, 1960, **1,** 1.