PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Gene Conversions and Their Relation to Homologous Chromosome Pairing

O. Smithies and Patricia A. Powers

Phil. Trans. R. Soc. Lond. B 1986 **312**, 291-302
doi: 10.1098/rstb.1986.0008

Phil. Trans. R. Soc. Lond. B **312**, 291–302 (1986)
Printed in Great Britain

Gene conversions and their relation to homologous chromosome pairing

By O. Smithies and Patricia A. Powers
Laboratory of Genetics, University of Wisconsin Madison, Wisconsin 53706, *U.S.A.*

Gene conversion is the non-reciprocal transfer of DNA sequences from one gene to a related gene elsewhere in the genome. Molecular evidence for its occurrence in higher eukaryotes was first described by our laboratory in 1980 in the two linked human foetal γ-globin genes. Over a kilobase of DNA was converted in this initial example. Other investigators have since described more examples of gene conversion including some in which the sequence that was transferred is much shorter. We have now accumulated evidence for a series of such small gene conversions in the human foetal globin gene pair. The number of small gene conversions that we have been able to detect leads us to suggest that gene conversions are the consequence of a general mechanism whereby DNA strand invasions enable chromosomes to find their homologues during meiosis.

Introduction

Gene conversion is the name that geneticists working with fungi gave to a phenomenon they observed when studying recombination between closely linked genetic markers during meiosis. Molecular biologists working with higher eukaryotes have also come to use this same term, and it is first necessary to define gene conversion in this context.

Consider a DNA segment from equivalent regions of two homologous chromosomes. The two DNA segments will, of course, be very much alike but let us assume that they differ in their nucleotide sequences by just enough for them to be distinguished (for example, by the solid and wavy lines in figure 1). The left hand side of figure 1 shows the products at the chromosomal level of a *single reciprocal crossover* within these homologous chromosomes, and of a *double reciprocal crossover*. Notice that the relative number of copies of each chromosomal region remains unchanged after the crossovers, although their linkage relationships are altered.

If we look at recombination on the scale usually considered in DNA sequences (that is, of the order of a few kilobase pairs of DNA), products looking like *double* crossovers can be formed from a *single* recombinational event. The right-hand side of figure 1 illustrates recombination at the level of DNA molecules, and is based on a model for recombination proposed by Meselson & Radding (1975) largely as the result of studies of meiotic recombination in fungi. Specific details of this model can be questioned, but its ability to illustrate the various outcomes of recombination when viewed at the DNA level is not in much doubt. Notice that two of the products (K6 and K7) of the illustrated recombinational event appear to be double crossovers. The clue that led to excluding double crossing-over as the explanation for these two products was the same as that that led to the discovery of gene conversion, and depends on observing all the products of meiosis. When this is done, as is possible in many fungi, events of this general type are often seen to be accompanied by a change in the numbers of copies of parts of the participating chromosomes. Part of one of the participating chromosomes frequently appears to have been changed to look like the equivalent part of the other chromosome involved in

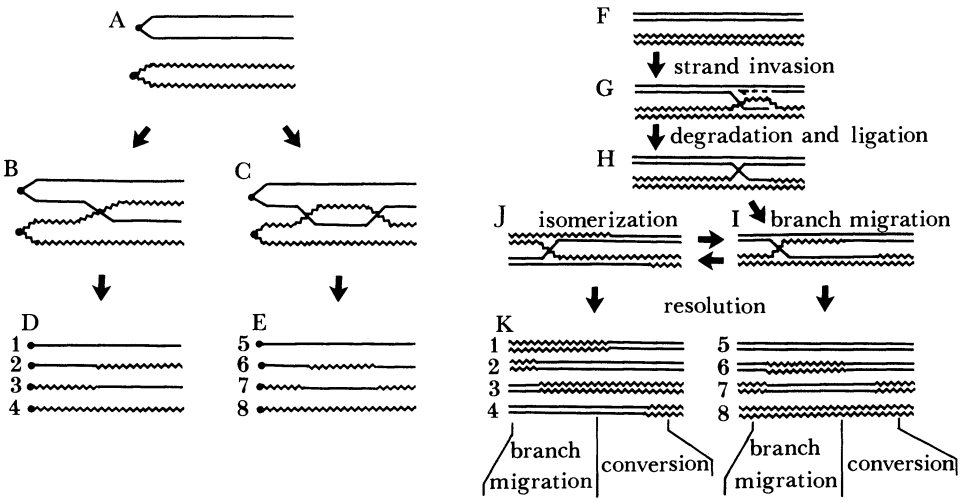

FIGURE 1. A comparison of crossovers at the level of chromosomes (A–E) and DNA molecules (F–K). (A–E) an illustration of single (B and D) and double (C and E) crossovers between homologous chromosomes. Chromatids are represented by straight and wavy lines, centromeres by solid circles. (F–K) an illustration at the level of DNA molecules of events presumed to occur during recombination between homologous regions of DNA (after Meselson & Radding 1975). Two homologous but non-identical DNA molecules are represented by the paired straight and wavy lines. Note that, at the DNA level, a region looking like a double crossover (as in K7) can result from a region of gene conversion plus a region of branch migration, or (as in K6) from a region of branch migration only.

the recombinational event. This process is called *gene conversion*, and is most clearly seen in part K7 of figure 1. Notice that in the region denoted 'conversion' there are three copies of paired straight lines, but only one copy of paired wavy lines. The straight line part of K7 is the converted region. The formal distinction between gene conversion and various reciprocal events, including double crossovers, requires us to score all the products of meiosis. Gene conversion will have been demonstrated when it can be shown that a region of DNA has *non-reciprocally* become like the same region on the homologous chromosome: a process that usually leads to an aberrant 3:1 segregation ratio of the four meiotic products. This contrasts with double crossovers, which are reciprocal and lead only to a redistribution of the products without changing the normal 2:2 segregation ratio.

Another important phenomenon thought to take place in DNA molecules involved in recombination is *branch migration*, which is illustrated in the step between H and I in figure 1. Branch migration is the linear migration of the 'half-crossover' along two DNA molecules that have a single cross link between their double helices (as in H and I). This migration requires the formation of Watson–Crick hydrogen bonds between DNA strands that were originally in different molecules. Consequently it can only occur when the two partners in the recombination have very similar nucleotide sequences. The importance of branch migration in the present context is that it will increase the extent of the region mimicking a double crossover (see the region of paired straight lines in K7 denoted 'branch migration'). Branch migration as such does not change the segregation ratio. Indeed it can produce a region of apparent double crossover which is *not* associated with gene conversion on the *same* molecule (see K6). So we can see that, at the scale of DNA sequences, virtually indistinguishable products may be the result of either gene conversion, branch migration or double crossover.

Since, in the case of higher eukaryotes, one cannot recover all the products of any given

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

meiosis, in a formal sense one cannot prove whether a region of straight lines flanked by wavy lines represents a gene conversion, branch migration or a double crossover. However, by analogy to the fungal work and to avoid inventing a new term for two functionally equivalent processes, most molecular biologists working with higher eukaryotes will call the event a gene conversion when the region of the straight lines is in the range of a few tens to a few thousands of nucleotides. Some investigators have also used the term *gene correction* for the same phenomenon, but our preference is to avoid using the word correction, for it implies that the direction of the change is always to make things better, which is not necessarily the case.

Gene conversion between the human foetal globin genes

The first molecular evidence for gene conversion in higher eukaryotes turned up fortuitously in the course of a study in our laboratory of the genes coding for the human foetal globin chains, the β-type polypeptide chains of foetal haemoglobin (Slightom *et al.* 1980). Earlier protein studies had shown that at birth nearly all humans have two types of foetal globin chains which differ only by the presence of either a glycine residue or an alanine residue at amino acid position 136.

We were looking at the nucleotide sequences of the corresponding G_γ and A_γ globin genes which occur on chromosome 11 in humans, and cloned and sequenced the DNA from the foetal globin gene region in both copies of this chromosome from a single individual. We found, as had long been suspected from the protein sequences, that the two genes had arisen by a tandem duplication of part of the DNA within the β-globin gene cluster. The duplication turned out to be about 5 kilobase pairs in length (Shen *et al.* 1981). One of the chromosomes (we called it B) was straightforward: the non-coding regions of its G_γ gene differed in a more or less random fashion from those of its A_γ gene. But the other chromosome (A) was quite different: part of its A_γ gene was much more like a G_γ gene than it was like an A_γ gene.

At the 3′ boundary between the G_γ-like region of the A_γ gene from chromosome A there is a stretch of simple sequence DNA, and we noticed that the simple sequence of the A_γ gene of chromosome A could be derived from a similar simple sequence in the G_γ gene of chromosome A by unequal crossing over. This suggested to us that some type of recombinational event had occurred within this region, and that this event had caused the transfer of DNA from a G_γ gene to the A_γ gene to give the arrangement of DNA sequences found in chromosome A. We called the event a gene conversion because it was so like the gene conversions of fungi: a use of the term that is now accepted.

We found that the Meselson–Radding model for gene conversion could account nicely for the data on the assumption that a recombinational event had been initiated between a G_γ gene mispaired with an A_γ gene. We suggested that DNA strand transfer had occurred between the mispaired genes so that a rather extensive stretch in the A_γ gene was converted to become G_γ-like. The observed length of the conversion was about 1.5 kilobase pairs.

At the time we described these findings we believed that unequal crossing over leading to a chromosome with three foetal globin genes and another with only a single foetal globin gene might be an equally likely outcome of the type of event that led to the gene conversion. Our belief stemmed from the observation in fungi that gene conversion is accompanied by crossing over in some cases, and is unaccompanied by crossing over in other cases. As we shall show below, we now think that gene conversion can be dissociated from crossing over. That is not

to say that unequal crossing over leading to single and triple foetal globin genes does not occur: it does. Instead, we shall consider the possibility that crossovers may require additional quite complex steps.

Gene conversions in other gene systems

Following our original description of this gene conversion event in the human foetal globin genes, other investigators also began to observe instances of sequences in different genes of other higher eukaryotes where the simplest explanation for the particular arrangement of DNA sequences was that one of the genes had been converted by a related gene.

In some of these cases the conversion appears to be short. A particularly convincing example of a short gene conversion was described by two groups studying the major histocompatibility gene complex of the mouse. The H2 genes, which encode type I histocompatibility antigens, are highly polymorphic in mice and have a relatively high rate of spontaneous mutation. Schulze *et al.* (1983) and Weiss *et al.* (1983) independently studied a mutant gene, $H2K^{bm-1}$, which had arisen spontaneously from the class I gene $H2K^{b}$. Cloning and DNA sequencing of the mutant allele showed that it was the same as the parent allele except for a stretch of between 13 and 50 nucleotides that appeared to be derived from another class I gene. Subsequent work showed that the donor gene was present in the genome of the original mouse strain but was at a previously unrecognized locus (Mellor *et al.* 1983). This work established that gene conversions can be quite short, of the order of tens of nucleotides, as well as of the order of several kilobases. It also illustrates the 'unrighteousness' of gene conversions, for this particular gene conversion produced a mutant allele (if any functional histocompatibility alleles can ever be regarded as mutants).

More gene conversions between the human foetal globin genes

Over the last few years we have been looking for other examples of gene conversion between the two human foetal globin genes. This pair of genes is well suited for such a search because: (i) an example within this gene pair is already known; (ii) the 5 kilobase pair duplicate regions of DNA encompassing the G_γ and A_γ globin genes do not differ greatly so that the formation of hybrid double-stranded DNA between homologous regions of the two genes should be possible and branch migration should not be inhibited; and (iii) the two genes code for almost identical products so that there is not likely to be any strong evolutionary selection for or against gene conversions between them. We hoped, therefore, to find conversions in the foetal globin genes which, in other systems, might be prevented by extensive sequence mismatching or be lost by selection.

Some chromosomes were available to us, or could be identified, with features suggesting that a gene conversion might have occurred in the foetal globin gene region at some time in their past histories. Dr Titus Huisman, in Augusta, Georgia, with whom we have collaborated in the past, has kept records of the levels of foetal haemoglobin in thousands of newborns and has in many cases measured the ratio of the G_γ and A_γ types of foetal globin chains. From these records he was able to identify several individuals who, together with other members of their families, had unusual G_γ to A_γ globin chain ratios: either abnormally high or abnormally low (Huisman & Altay 1981). Two different chromosomes were identified in which both foetal genes

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

were of the G_γ variety, in the arrangement G_γ-G_γ, and one chromosome was found with both genes of the A_γ variety, in the arrangement A_γ-A_γ (Powers *et al.* 1984). These three chromosomes were excellent candidates for having had a gene conversion that transferred coding sequences between the linked foetal globin genes.

Another chromosome was identified with restriction sites indicative of gene conversion. Jeffreys (1979) described a polymorphic *Hind*III restriction enzyme site that was present in some foetal globin genes but not in others. He found individuals who had the *Hind*III site in both their foetal genes (that is, were $++$), or had the *Hind*III site in the G_γ gene only (that is, were $+-$), or lacked the *Hind*III site in both genes (that is, were $--$). Clearly, if the single ancestral globin gene lacked the site before the gene duplication, then the duplicate genes would also lack the site after the duplication. Subsequent gain of one site could be a point mutation, but gain of the other would require either the same point mutation or a gene conversion to transfer the site from one gene to the other. (The same type of argument would apply if the single ancestral foetal gene already had the *Hind*III site.) Inspection of the DNA sequences of foetal globin genes from previously sequenced chromosomes having the $(--)$ and $(+-)$ haplotypes showed no evidence of conversion involving the polymorphic *Hind*III site. We therefore decided to study a chromosome of haplotype $(++)$.

In deciding where in the sequences of the foetal globin genes in the four selected chromosomes we should look for possible gene conversions between the foetal genes, we were guided by the need to look at codon 136 which is responsible for the glycine or alanine that distinguishes the G_γ and A_γ globin chains and at the region of the polymorphic *Hind*III site. We also decided to look at other parts of the genes where selection would be minimal. Globin genes are coded in three parts, with a small and a large intervening sequence dividing the coding sequences. Previous work (Efstratiadis *et al.* 1980) had shown that the large intervening sequence and the 3′ flanking sequences show a great deal of variation between different globin genes. We therefore chose to concentrate our DNA sequencing efforts on the approximately 1200 base pair length of DNA illustrated in figure 2 that includes 19 nucleotides from the second exon, about 900

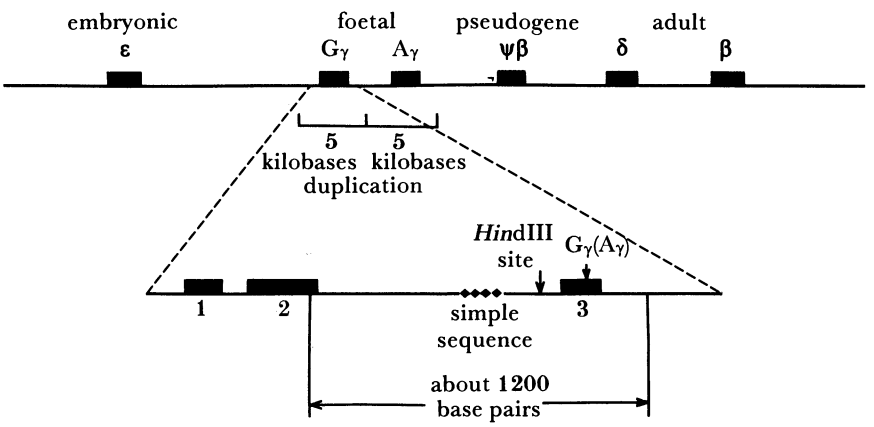

FIGURE 2. A diagram showing the human β-globin gene cluster (upper line) and the locations of the sequenced regions within the foetal globin genes (lower line). The five productive genes, ε, G_γ, A_γ, δ and β, and the non-productive pseudogene, ψβ, are shown as solid bars in the upper line which also illustrates the duplicated 5 kilobase pairs of DNA that led to the formation of G_γ and A_γ globin genes. In the lower line the solid bars represent the three coding regions of the foetal globin genes. The sequenced region of about 1200 base pairs is shown by a line. The region of simple sequence (diamonds) within the large intervening sequence, the location of the polymorphic *Hind*III site, and the position where the coding regions of the G_γ and A_γ genes differ are indicated.

nucleotides making up the large intervening sequences, 129 nucleotides making up the third exon (coding for amino acids 105 through 146) and 121 nucleotides from the 3′ flanking region. We determined this DNA sequence for both foetal globin genes from each of the four chromosomes described above and combined our data with data available from the literature to give a total of seven human chromosomes carrying 14 foetal globin genes (Powers & Smithies 1985).

To sort out what mutational or recombinational events could account for the evolution of these 14 foetal globin genes we first had to establish their inter-relatedness. This is most simply done by arranging them into an evolutionary tree on the branches of which we can then insert the mutational and recombinational events. Four categories of mutational event (N, C, Q and S) were used in describing the data. Examples of each of these types of mutational events are illustrated in figure 3. N is used to depict the occurrence of a *single nucleotide substitution* at some

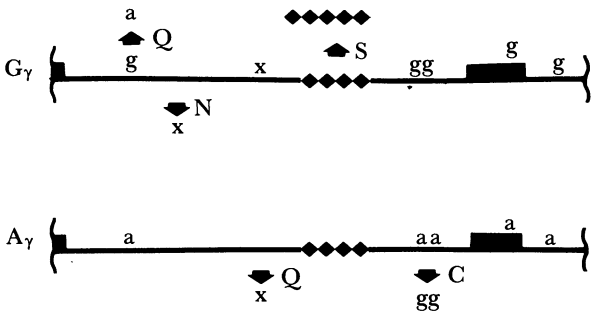

FIGURE 3. A diagram to illustrate the definitions of the four categories (N, C, Q and S) of mutational events used in the evolutionary tree. The lower case letters, g and a, represent nucleotides at positions where the two foetal globin genes differ characteristically: g represents a nucleotide characteristic of G_γ genes, a represents a nucleotide characteristic of A_γ genes. The letter x represents a nucleotide difference at a position where the G_γ and A_γ genes are usually the same. A change in the length of the simple sequence is shown by a change in the number of diamonds. The letters N, C, Q and S alongside the arrows indicate how the respective change is classified. More detailed definitions of the four categories of mutation are given in the text.

position in the sequence that is usually the same in all the human foetal globin genes studied. C is used to depict an event which is a *gene conversion* as judged by the *replacement of more than one nucleotide* (or a length difference) in a given location in either a G_γ or an A_γ gene by the equivalent nucleotides (or length difference) from the other foetal globin gene. Q represents mutational events that are questionably either *gene conversions* or single base pair substitutions that *mimic gene conversions*, for example where a single nucleotide position is changed from being G_γ-like to being A_γ-like (or vice versa). S denotes a change in the *length of the stretch of simple sequence* (all events of this category involve a change in an integral number of the dinucleotides TG or GG). The tree was made so as to minimize the number of these events. Some ambiguities in trees of this type are not uncommon, but we found that two trees were better than all the others. Since the two best trees do not differ by any feature that affects our conclusions, it does not matter which we use. One of these trees is shown in figure 4.

In the tree shown in figure 4 there are four examples of foetal globin genes with category C mutations, as we defined them above. One is the long 1.5 kilobase pair gene conversion that started the whole story. (We have now identified two chromosomes (563A and AR) which are descendants of the chromosome in which this event initially occurred.)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

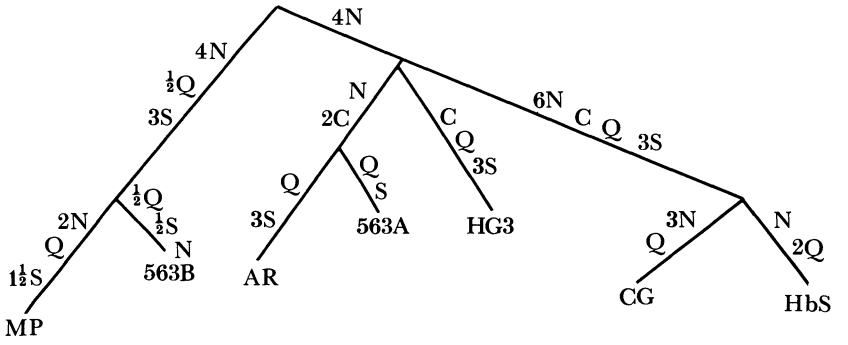

FIGURE 4. An evolutionary tree for the seven human chromosomes considered (after Powers & Smithies 1985). The seven chromosomes are: AR, with the A_γ-A_γ gene arrangement; CG and MP, each with G_γ-G_γ gene arrangements; HG3, with the (++) arrangement of *Hind*III sites; 563A and 563B, originally described by Slightom *et al.* (1980); and HbS, originally described by Stoeckert *et al.* (1984). Mutational events of the four categories defined in figure 3 and in the text are shown, with the numbers signifying how many events (other than one) of each category are required in a given part of the tree. A total of 22N, 4C, 9Q and 15S category events are required to account for all the data.

A new short gene conversion event was identified that transferred the nucleotides TCAC normally seen in the sequence of G_γ genes to the equivalent positions in the A_γ gene of chromosome HG3. The data require that this be a short gene conversion of from 4 to 282 nucleotides. Incidentally both γ genes on chromosome HG3 have the polymorphic *Hind*III restriction site and so this chromosome already had signs elsewhere in its structure of having had a gene conversion.

A different short gene conversion of from 7 to 344 nucleotides was identified in an ancestor common to chromosomes CG and HbS: it transfers the nucleotides GC...AA from a G_γ gene to an A_γ gene.

The fourth category C event is a length change in an ancestor common to chromosomes AR and 563A: an insertion of the single nucleotide A in a G_γ gene which makes it like an A_γ gene at the corresponding position. We could find no repetitive or other features in the local nucleotide sequence that might have predisposed an exact recurrence of the insertion of this single nucleotide in this particular position, hence we classify it as a gene conversion.

We found nine examples of Q category events that could either be the result of gene conversions or single nucleotide changes that mimic gene conversions. Four were preselected by us: they were the mutational events resulting in the two selected G_γ-G_γ chromosomes, the selected A_γ-A_γ chromosome and the selected *Hind*III (++) chromosome, but five were unselected. Statistical arguments, to which I shall return, make it very unlikely that most of the Q category events are due to random changes that mimic gene conversion.

The most common type of mutation is of the category N, that is, a nucleotide substitution at a position where normally the two foetal genes are alike. Since the two genes are alike at about 98% of the positions we have sequenced it is not surprising that this category is the most frequent (it was seen 22 times). However, in fact it is less common than the Q category *when one considers the number of positions at risk*.

There are about 800 nucleotides in the region we sequenced that can reasonably be considered as neutral, in the sense that comparing different globin genes shows no evidence for it mattering which nucleotides occupy these positions. Among these 800 positions, about 20 are usually different in the G_γ and A_γ genes. Thus the probability that a *random* nucleotide

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

substitution will occur at a position where the G_γ and A_γ genes do not differ is roughly 40 times more than the probability that it will occur at a position where they do differ (a necessary prerequisite if a nucleotide substitution is to mimic a gene conversion). Yet the unselected N category mutations are only about five times more frequent than the unselected Q category mutations. Consequently it is very likely that several or all of the Q category mutations are really gene conversions.

One complication that faces this type of statistical argument is the possibility that some of the mutations that led to the G_γ versus A_γ differences might occur more than once because they are predisposed by local features of the DNA sequence. If this were true, any inferences drawn from assuming that the mutations are random would be in error. We therefore examined the nucleotide sequences associated with the Q category of mutational events for any features that might make them particularly mutable. We could find no direct repeats or palindromic sequences that might facilitate the observed mutations. Nor could any of the Q category events be explained as the consequence of deamination of the 5-methylcytosine in CpG dinucleotides. In addition we examined the nucleotides at the corresponding positions in the DNA sequences of the foetal globin genes of gorilla (Scott *et al.* 1984) and chimpanzee (J. Slightom, personal communication) and could find no evidence that these positions were significantly predisposed to the occurrence of the observed mutations. Thus our conclusion that many of the Q type events are the result of gene conversions appears to be justified.

Gene conversions are frequent

The total number of gene conversions in our tree is between 4 (the sum of all the C events) and 13 (the sum of C and Q events) with a low likelihood that the smaller number is correct. When one considers the fact that we can only detect gene conversions in regions where the two foetal globin genes already differ, then the molecular events that lead to these gene conversions must be frequent: perhaps several times more frequent for a given length of sequence than are conventional base pair substitutions (otherwise the N type events would outnumber the Q type events by a factor approaching 40).

Our evidence thus leads us to conclude that gene conversions are quite frequent events when one looks for them in an appropriate system. The question arises why they would be so frequent. It is hard to convince oneself that gene conversions have a selective advantage of sufficient immediacy and magnitude to lead to the direct evolution of a mechanism specifically to increase their frequency. That is not to say that gene conversions are without interest in an evolutionary sense (see Dover & Tautz this symposium) but it does imply that we should look elsewhere for an explanation of their frequency. We thought that they might instead be the products of some other general biological process having an immediate benefit on which selection can act. The best general process we have been able to imagine as being a suitable candidate is embodied in the requirement in *multichromosomal* organisms for a mechanism to facilitate recombination between related DNA sequences in homologous chromosomes and to avoid recombination between related sequences on non-homologous chromosomes. Homologous recombination has the selectable advantage of producing new combinations of genes, while recombination between non-homologous chromosomes produces chromosomal translocations with greatly reduced fertility, an immediate selective disadvantage. Our suggestion is that small gene conversions may be by-products of the process used to obtain pairing between homologous chromosomes.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Homologous chromosome pairing

The means whereby homologous chromosomes are paired during meiosis has long been the subject of interest and debate. A substantial part of the discussion meeting on the meiotic process held by the Royal Society in 1975 (Riley *et al.* 1977) was devoted to this question. It is not possible to discuss all the subtleties of the topic, which can be found in the cited discussion and in more recent works (Maguire 1984; Von Wettstein *et al.* 1984), but we do wish to emphasize that virtually all who have considered the subject are puzzled by what appear to be long range interactions between homologous chromosomes. Special processes have been invoked to account for this feature, such as the involvement of filamentous proteins attached at specific sites determined by local nucleotide sequences. The avoidance of problems caused by chromosomal entanglements during meiosis has also been a source of puzzlement, although possibly compaction of chromosomes before synapsis may reduce the incidence of entanglements (discussed in detail by Stern 1977). Whether formation of the synaptonemal complex precedes, follows, or is the cause of synapsis is also subject to debate.

We suggest that the required long range interactions are possible if chromosomal synapsis is mediated at least in part through interactions at the DNA level. If we assume that the mutual recognition of homologous chromosomes occurs through DNA–DNA interactions at a time when some of the chromatin is still in an uncompacted state, then subsequent condensation of the chromatin during chromosomal compaction could draw the homologues together. The total physical length of a typical mammalian haploid genome if it is in the form of essentially uncompacted 30 nm chromatin is approximately 26 mm ((number of base pairs × axial length per base pair in millimetres)/packing ratio of 30 nm chromatin = $[(3 \times 10^9) \times (3.4 \times 10^{-7})]/40 = 26$). The smallest chromosomes each contain about 1% of the genome. Even if we assume that only 10% of their chromatin is in the uncompacted form at the time of synapsis, the range available (26 μm) is still likely to be sufficient for these chromosomes to come into contact with each other from any position they might occupy in the nucleus. These figures suggest that compaction of chromosomes, if it is a prerequisite to avoid chromosomal entanglements during meiosis, could be extensive before distance problems are likely to be encountered. Similarly the existence of enzymes such as the topoisomerases, which effectively permit one DNA molecule to pass through another without losing its integrity, suggests that any residual interlocking of synapsed chromosomes can be resolved at the DNA level (see Von Wettstein *et al.* (1984) for a detailed discussion of this matter). With these comments in mind we would like to present a model of the molecular basis for the pairing of homologous chromosomes. We claim no originality for our ideas except in their relation to the prevalence of gene conversions.

A model for the pairing of homologous chromosomes

Figure 5 presents an illustration of our model for homologous pairing, which draws on previous models for gene conversion (see, for example, Holliday 1964; Meselson & Radding 1975; Szostak *et al.* 1983) and has similarities to parts of a model of meiotic chromosome behaviour discussed by Hotta *et al.* (1984). The specific details of our model are as follows. We suggest that:

(i) during meiosis single-stranded 'feelers' are extruded from many sites along DNA molecules.

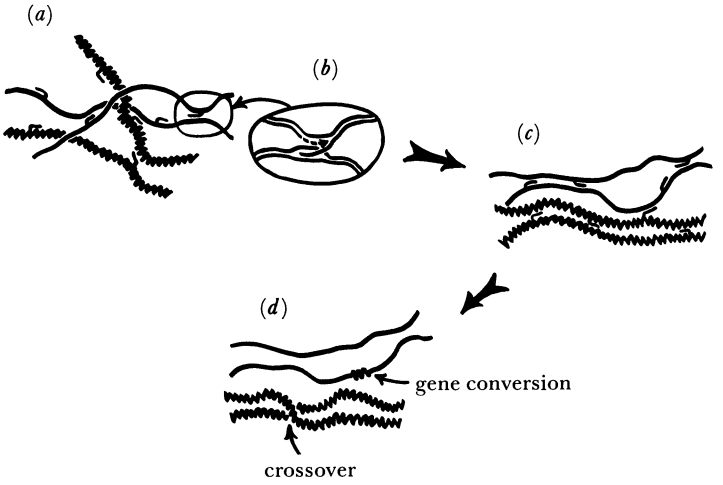

FIGURE 5. An illustration of the proposed relation between chromosomal pairing and gene conversions. Two homologous pairs of chromatids, represented by straight lines and zig-zag lines, are shown. (*a*) Single-stranded DNA 'feelers' are extruded at many positions along the unsynapsed chromatids and invade nearby DNA molecules. (*b*) An enlarged view of a typical extrusion and invasion. The single-stranded feeler, extruded from the upper chromatid by newly synthesized DNA shown as a dashed line, invades the lower chromatid. (*c*) 'Scanning' leads to some stable invasions. 'Zipping' of the homologous chromosomes, complete for the zig-zag and partial for the straight chromatids, has occurred because of cooperation between nearby invasions. (*d*) Once the homologous chromatids have paired, the machinery permitting crossing over can be set in motion. Stable invasions between non-homologues will not be sufficiently frequent to permit the zipping effect. In these cases, the 'footprints' of the invading feelers may occasionally lead to gene conversions.

(ii) These feelers can invade any DNA duplex that they encounter and (perhaps under the influence of a rec A-type protein) then can scan the invaded duplex for homologous sequences with which the feeler can form a Watson–Crick double helix.

(iii) Scanning is halted when a nucleotide sequence is encountered that can form a stable double helix.

(iv) If nearby invasions have also been successful, then a kind of zipping effect will be possible. Since multiple adjacent stable heteroduplexes can form between homologous chromosomes, but not between non-homologous chromosomes, the zipping effect will lead to the pairing of homologues. Once the homologues have been correctly paired, the other machinery necessary to achieve crossing over can be set in motion. Note that chiasma formation and reciprocal crossing over can be regulated independently of the frequency of the strand invasion leading to synapsis.

(v) If, on the other hand, a local stable heteroduplex has been formed as a consequence of a related sequence being found at a non-homologous chromosomal location, then any nearby invasions that have occurred between the same DNA molecules will not in general be able to find stable positions. Consequently full pairing and crossing over will not be able to follow. The distinction between pairing involving homologues and non-homologues will have been achieved based on the presence or absence of many adjacent invasions that find stable positions.

(vi) The relics of some of the stable heteroduplexes formed between related sequences at non-homologous chromosomal locations give rise to short gene conversions.

Proximity is likely to influence the chances of two DNA sequences interacting as described above. We expect, therefore, that gene conversions between closely linked genes will occur more frequently than conversions between genes at widely separated loci. Indeed we have evidence in another very closely linked gene pair, the human haptoglobin and haptoglobin-related genes,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

that more gene conversions and crossing over events have occurred in the history of these two loci than have single nucleotide substitutions (N. Maeda and O. Smithies, unpublished results). Our model is also consistent with gene conversions occurring between closely linked genes within a single chromatid or between sister chromatids, as well as between homologous chromosomes.

An important corollary of the model is its potential for separating the process of synapsis (and of gene conversions) from the formation of chiasmata and the occurrence of crossing over, even though synapsis is a prerequisite of crossing over. We are suggesting that DNA strand invasions initiate synapsis and can lead to gene conversions, but that more complex machinery, including the synaptonemal complex and recombinational nodules, is likely to be necessary for the occurrence of crossing over. In this regard, it is of considerable interest to note that although crossovers show classical interference (that is, the presence of one crossover interferes with the occurrence of a second crossover within its immediate neighbourhood), gene conversions do not. Nor do gene conversions interfere with crossovers or vice versa (see Holliday (1977) for a more extensive discussion of these points). The alignments of homologous chromosomes and of their multiple strands in the polytene nuclei of *Drosophila* and similar species may also be mediated by DNA strand invasions of the type we have discussed here.

Conclusions

To summarize our present views, we propose that local DNA strand invasions form the molecular basis for the recognition of homologous chromosomes during meiosis. If several of these strand invasions in a given region of DNA lead to stable heteroduplexes, then their co-operation can lead to extensive pairing (synapsis) followed by crossing over in some of the paired chromosomes. Local mispairing between non-allelic regions of DNA which have locally homologous nucleotide sequences are likely to occur from time to time, but co-operative pairing will not be able to follow. None the less the invading single-stranded molecules might occasionally become incorporated into the invaded DNA double helix, leading to a gene conversion of limited extent. The short gene conversions we have found so commonly between the human foetal globin genes may be the relics of such abortive pairing. We expect that similar small gene conversions will prove to be generally common within the genomes of higher eukaryotes.

This is paper number 2805 from the Laboratory of Genetics, University of Wisconsin-Madison. This work was supported by N.I.H. grants GM20069 and AM20120 to O. S. In addition P. A. P. was supported by N.I.H. training grant T326GM07133.

References

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. 1980 The structure and evolution of the human β-globin gene family. *Cell* **21**, 653–668.

Holliday, R. 1964 A mechanism for gene conversion in fungi. *Genet. Res.* **5**, 282–304.

Holliday, R. 1977 Recombination and meiosis. *Phil. Trans. R. Soc. Lond.* B **277**, 359–370.

Hotta, Y., Tabata, S. & Stern, H. 1984 Replication and nicking of zygotene DNA sequences: control by a meiosis-specific protein. *Chromosoma* **90**, 243–253.

Huisman, T. H. J. & Altay, C. 1981 The chemical heterogeneity of the fetal hemoglobin of black newborn babies and adults: a reevaluation. *Blood* **58**, 491–500.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Jeffreys, A. J. 1979 DNA sequence variants in the G_γ-, A_γ-, δ- and β-globin genes of man. *Cell* **16**, 1–10.
Maguire, M. P. 1984 The Mechanism of meiotic homologue pairing. *J. theor. Biol.* **106**, 605–615.
Mellor, A. L., Weiss, E. H., Ramachandran, K. & Flavell, R. A. 1983 A potential donor gene for the *bm1* gene conversion event in the C57BL mouse. *Nature, Lond.* **306**, 792–795.
Meselson, M. S. & Radding, C. M. 1975 A general model for genetic recombination. *Proc. natn. Acad. Sci. U.S.A.* **72**, 358–361.
Powers, P. A., Altay, C., Huisman, T. H. J. & Smithies, O. 1984 Two novel arrangements of the human fetal globin genes: G_γ-G_γ and A_γ-A_γ. *Nucl. Acids Res.* **12**, 7023–7034.
Powers, P. A. & Smithies, O. 1985 Small gene conversions between human fetal globin genes suggest a mechanism for homologous chromosome pairing. (In preparation.)
Riley, R., Bennett, M. D. & Flavell, R. B. (eds) 1977 A discussion on the meiotic process. *Phil. Trans. R. Soc. Lond.* B **277**, 183–376.
Schulze, D. H., Pease, L. R., Geier, S. S., Reyes, A. A., Sabmiento, L. A., Wallace, R. B. & Nathenson, S. G. 1983 Comparison of the cloned *H-2K*^{bm1} variant gene with the *H-2K*^{b} gene shows a cluster of seven nucleotide differences. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2007–2011.
Shen, S., Slightom, J. L. & Smithies, O. 1981 A history of the human fetal globin gene duplication. *Cell* **26**, 191–203.
Slightom, J. L., Blechl, A. E. & Smithies, O. 1980 Human fetal G_γ- and A_γ-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21**, 627–638.
Stern, H. 1977 Concluding remarks. *Phil. Trans. R. Soc. Lond.* B **277**, 371–376.
Stoeckert, C. J., Jr, Collins, F. S., & Weissman, S. M. 1984 Human fetal globin DNA sequences suggest novel conversion event. *Nucl. Acids Res.* **12**, 4469–4479.
Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. 1983 The double-strand-break repair model for recombination. *Cell* **33**, 25–35.
Von Wettstein, D., Rassmussen, S. W. & Holm, P. B. 1984 The synaptonemal complex in genetic segregation. *Ann. Rev. Genet.* **18**, 331–413.
Weiss, E., Golden, L., Zakut, R., Mellor, A., Fahrner, K., Kvist, S. & Flavell, R. A. 1983 The DNA sequences of the *H-2K*^{b} gene: evidence for gene conversion as a mechanism for the generation of polymorphism in histocompatibility antigens. *EMBO J.* **2**, 453–462.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES