SHORT GENE CONVERSIONS IN THE HUMAN FETAL GLOBIN GENE REGION: A BY-PRODUCT OF CHROMOSOME PAIRING DURING MEIOSIS?

PATRICIA A. POWERS AND OLIVER SMITHIES

Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706

Manuscript received July 12, 1985
Accepted October 3, 1985

ABSTRACT

DNA sequence comparisons of a 1200-base pair (bp) region in 14 human fetal globin genes in seven linked pairs reveal 31 nucleotide substitutions at positions where the fetal globin genes, G_γ and A_γ, usually differ. In each case, the newly substituted nucleotide is identical to the one found at the same position in the linked nonallelic gene. Most of these nucleotide substitutions are clearly the result of gene conversions, but 11 could be the result of either very short gene conversions or of point mutations. The unexpectedly frequent occurrence of these short gene conversions suggests that they may be the relics of some normal interaction between homologous but nonallelic DNA sequences, and we discuss the possibility that they result from interactions occurring between homologous sequences during the process of meiotic chromosome pairing.

GENE conversion, the nonreciprocal transfer of DNA sequence information from one DNA duplex to another, was first observed as the non-Mendelian segregation of the products of a single meiosis in fungi. Since the initial observations, genetic and molecular analyses in fungi have demonstrated conversions of the whole or portions of genes by allelic genes and by related but nonallelic genes (JACKSON and FINK 1981; KLEIN and PETES 1981; PETES and FINK 1982; KLAR and STRATHERN 1984; KLEIN 1984).

The first molecular evidence for gene conversions in higher eukaryotes was provided in 1980 by SLIGHTOM, BLECHL and SMITHIES (1980), who analyzed the DNA sequences of three human fetal globin genes from one individual and showed that part of the DNA sequence in one of the fetal globin genes had probably been replaced by the corresponding region of the linked but nonallelic fetal globin gene. All of the features of this unusual DNA sequence arrangement were consistent with the molecular models of gene conversion that had already been proposed to account for the genetic observations of gene conversion in fungi (HOLLIDAY 1964; MESELSON and RADDING 1975).

Additional molecular evidence for gene conversion in higher eukaryotes has since been reported. These additional examples of gene conversion occur between the duplicated DNA sequences of small gene families, including members of the α- and β-type globin gene clusters (LIEBHABER, GOOSSENS and KAN

1981; MICHELSON and ORKIN 1983; Hardison and MARGOT 1984), the immunoglobulin gene family (BENTLEY and RABBITTS 1983), the human haptoglobin gene family (N. MAEDA, personal communication) and the gene family encoding the mouse histocompatibility genes (SCHULZE *et al.* 1983; WEISS *et al.* 1983; MELLOR *et al.* 1983). Several other gene conversions between fetal globin genes in human (STOECKERT, COLLINS and WEISSMAN 1984), gorilla (SCOTT *et al.* 1984) and chimpanzee (J. SLIGHTOM, personal communication) have also been observed.

The human fetal globin genes are an excellent system in which to look for additional examples of gene conversion because of the nature of these two genes. The two human fetal globin genes, G_γ and A_γ, are the result of a 5-kilobase pair (kbp) tandem duplication estimated to have occurred about 35 million years ago (SHEN, SLIGHTOM and SMITHIES 1981). The genes themselves consist of three exons interrupted by a small 5′ intervening sequence (IVS1) and a larger 3′ intervening sequence (IVS2). The coding regions of the two genes differ by only a single nucleotide located in the third exon within the triplet coding for a glycine residue in the 5′ gene (G_γ) or an alanine residue in the 3′ gene (A_γ). The nucleotide sequences of the two genes differ by about 3% within their large intervening sequences, whereas over most of the nontranscribed regions, the present-day copies of the duplicated region differ by an average of 14% (SHEN, SLIGHTOM and SMITHIES 1981). The structures of the two genes are so similar that it appears unlikely that any strong evolutionary selection would exist either for or against gene conversion between them, except, perhaps, in the 5′ flanking regions. In addition, because of the high degree of homology between these two genes, gene conversion will be unhindered by regions containing extensive sequence differences. Consequently, gene conversions may be detected in this particular gene pair that might be lost by unfavorable selection or be prevented by extensive sequence differences in other genes.

In order to explore further the incidence of gene conversions between the human fetal globin genes, we chose to compare regions including the large intervening sequence, the third exon and some 3′ flanking DNA in the 14 paired genes of seven chromosomes. This particular region of the fetal globin genes was chosen for detailed study because previous work has shown that the large intervening sequences and the 3′ flanking regions of globin genes are poorly conserved during evolution (EFSTRATIADIS *et al.* 1980), so that selective forces are not likely to play a great role in determining acceptable mutations within much of the region we have studied. Some of the chromosome pairs were selected for study because we already had evidence indicating that they might have been involved in gene conversions. Other pairs were from randomly selected chromosomes. Our comparison has revealed several new examples of gene conversions between the human fetal globin genes. The relative ease of detecting these gene conversions, and their lengths (most are short), suggests to us that they may be the result of some general cellular process; and we discuss the possibility that these short gene conversions may be by-

products of the normal interactions occurring between homologous sequences during the process of meiotic chromosome pairing.

MATERIALS AND METHODS

DNA preparation: Large molecular weight DNA was prepared from fibroblasts by the method of BLATTNER *et al.* (1978) or from peripheral blood leukocytes by the procedure of PONCZ *et al.* (1982).

Cloning the linked fetal globin genes: Cloning into bacteriophage vectors, and subcloning into plasmid vectors, was performed essentially as described by POWERS *et al.* (1984).

DNA sequencing and analysis: DNA sequence analysis was performed as described by MAXAM and GILBERT (1977), using the modifications of SLIGHTOM, BLECHL and SMITHIES (1980). The nucleotide sequences of all of the regions that differ between the genes were determined at least twice. Nucleotide sequences were analyzed using software provided by the University of Wisconsin Genetics Computer Group (DEVEREUX, HABERLI and SMITHIES 1984).

RESULTS

DNA sequence analysis: The nucleotide sequences of 1176 bp of DNA, which includes the approximately 900-bp large intervening sequence (IVS2), the third exon and some 3′ flanking DNA, were determined for both γ genes from four different chromosomes. These four chromosomes were chosen because each contained some feature already suggesting gene conversion events. Two of the chromosomes (*CG* and *MP*) had both fetal globin genes of the G_γ variety, in the arrangement G_γ-G_γ, and one (*AR*) had both fetal globin genes of the A_γ variety, in the arrangement A_γ-A_γ. These chromosomes had been characterized in an earlier study (POWERS *et al.* 1984), in which we had also shown by DNA sequencing that the coding sequences of two of the unusual genes, the 5′ A_γ gene of *AR* and the 3′ G_γ gene of *MP*, differed from their normal counterparts by a nucleotide substitution that could be the result of either a gene conversion or of a single nucleotide replacement. The fourth chromosome (*HG3*) in the present study was obtained as a cosmid clone provided by FRANK GROSVELD (GROSVELD *et al.* 1981). The fetal globin gene region in this cosmid has the usual G_γ-A_γ arrangement of fetal globin genes, but we selected it for study after we had determined that the polymorphic *Hin*dIII restriction site within IVS2 was present in both of its γ genes. JEFFREYS (1979) had originally discussed the possibility that this arrangement of polymorphic *Hin*dIII restriction sites in the large intervening sequence of the fetal globin genes might be the result of a gene conversion.

Previously determined sequences of the linked γ genes from three additional human chromosomes were combined with the four chromosomes described above to make our analysis more comprehensive. Two of these previously described chromosomes, *563A* and *563B*, were the chromosomes *A* and *B* of a single individual, *563*, already studied extensively in our laboratory (SLIGHTOM, BLECHL and SMITHIES 1980; SHEN, SLIGHTOM and SMITHIES 1981; J. L. SLIGHTOM, S. SHEN and O. SMITHIES, unpublished data). The third previously described chromosome, which we refer to as *HbS*, was originally studied by

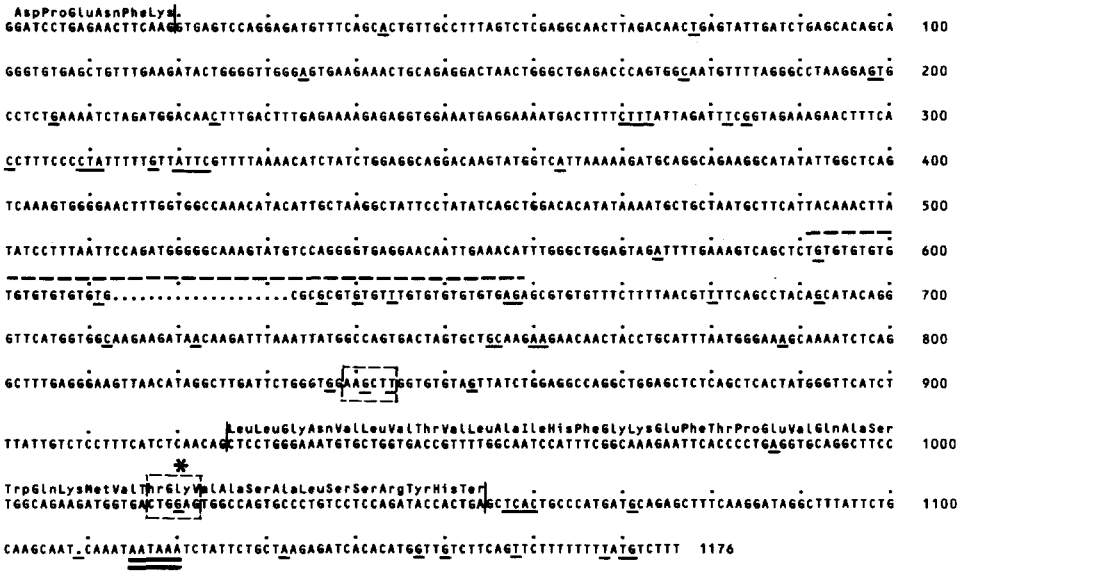

FIGURE 1.—The nucleotide sequence of a normal 5′ G_γ gene. The 1176 bp of DNA sequence representing the 3′ end of the second exon, the large intervening sequence, the third exon and the 3′ untranslated region is shown here for the G_γ gene from chromosome *HG3*. Gaps, represented by dots, were introduced to maximize the alignment of the nucleotide sequences of the 14 γ genes considered in this study. The amino acids encoded by the exons are printed above the sequence. Variable positions shown in Figure 2 and discussed in the text are underlined. Boxes mark the locations of the polymorphic *Hin*dIII recognition sequence (present in the large intervening sequence of this gene) and of the *Pst*I recognition sequence (present in the third exon of A_γ genes). The region of simple sequence DNA is overlined with a dashed line. An asterisk marks the site of the single coding region difference between G_γ and A_γ genes. Ter is the terminator codon. The poly(A) addition signal is underlined twice.

STOECKERT, COLLINS and WEISSMAN (1984), who determined the sequences of the large intervening sequence of the G_γ and A_γ genes linked to a β-globin gene with the sickle cell mutation ($\beta^{6\ \mathrm{Glu} \rightarrow \mathrm{Val}}$).

Figure 1 presents the nucleotide sequence of the relevant region of one of the genes that we sequenced (the G_γ gene from chromosome *HG3*) for reference and as a key to our numbering system. Figure 2 presents a compilation of all of the nucleotide differences in the sequences that we have determined, together with those in the other available human fetal globin gene sequences. Nucleotides at the tabulated positions in the sequences of the gorilla G_γ and A_γ genes (SCOTT *et al.* 1984) also are included.

Inspection of the data shows, as expected, that at many of the variable positions (*e.g.*, positions 657, 658, 680 and 692) all of the 5′ γ genes differ from all of the 3′ γ genes. Such differences are simply interpreted as nucleotide substitutions that have accumulated since the two linked fetal globin genes last shared a common ancestor. At a number of positions (*e.g.*, positions 363, 788 and 1057–1060) this pattern is disturbed by the single occurrence in either the 5′ or 3′ γ gene of a nucleotide or group of nucleotides characteristic of the other nonallelic gene. Such variations could either represent gene conver-

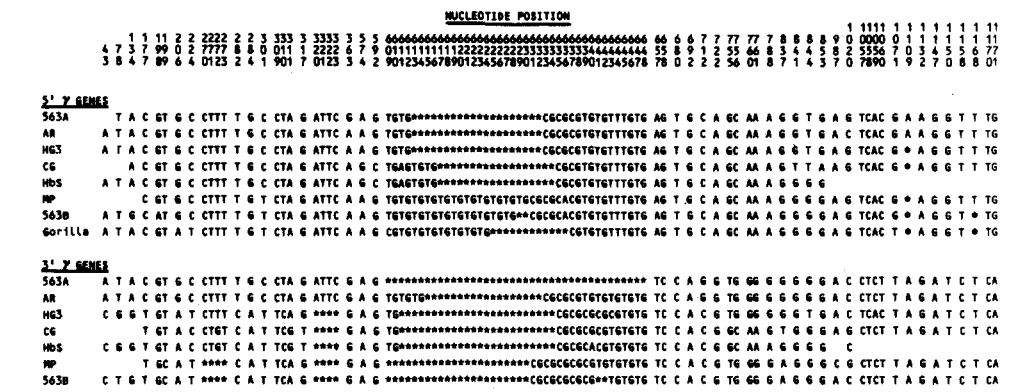

FIGURE 2.—Summary of the comparison of the γ gene sequences. The nucleotide differences between the 14 DNA sequences studied are tabulated. Numbers in vertical orientation give the nucleotide coordinates as shown in Figure 1. The DNA sequences of both γ genes from chromosomes *563A, 563B, AR* and *HG3* extend from position 1 to 1176. The DNA sequences of the 5′ G_γ gene from chromosome *CG* extend from position 120 to 1176. The DNA sequences of both γ genes from chromosome *MP* and the 3′ G_γ gene from chromosome *CG* extent from position 146 to 1176. The DNA sequences of both γ genes from chromosome *HbS* extend from position 1 to 975. Gaps, represented by the symbol *, were introduced to maximize the alignment of each sequence. Positions left blank represent those at which the DNA sequence was not determined.

sions of one γ gene by another or could be the result of mutations which mimic gene conversions.

Evolutionary tree of the seven human chromosomes: In order to analyze these data in more detail, we first constructed an evolutionary tree for the seven human chromosomes under investigation, using the combined 5′ and 3′ γ DNA sequences (Figure 3). We used the following system to score the number of mutational events, and we constructed the tree in order to minimize this score. Each nucleotide substitution was scored as a single event, except where a reasonable case could be made for several nucleotides being changed at once, such as at positions 1057–1060 where all the 5′ γ genes have the contiguous nucleotides TCAC which differ from the nucleotides CTCT at the same positions in most of the 3′ γ genes. In the region of simple sequence, which extends from positions 591 through 658, there are a few instances of single nucleotide substitutions (*e.g.*, at position 640 in the 3′ γ gene of chromosome *HbS*), but the majority of differences are changes in the numbers of $(TG)_m$ or $(CG)_n$ doublets. Each change in either m or n was scored as an independent event unless more than one change could be achieved by a single mutation, such as a deletion. Other deletions or insertions of contiguous nucleotides were scored as single events. The sequences of the gorilla G_γ and A_γ genes (SCOTT *et al.* 1984) were used in determining the root of the tree and in assigning the most likely sequence for the last common ancestor of the human chromosomes.

Using the ANCSTR program supplied by WALTER FITCH of the University of Wisconsin, we obtained two different trees, each with the same minimum number of mutational events. These two trees differ only in the placement of

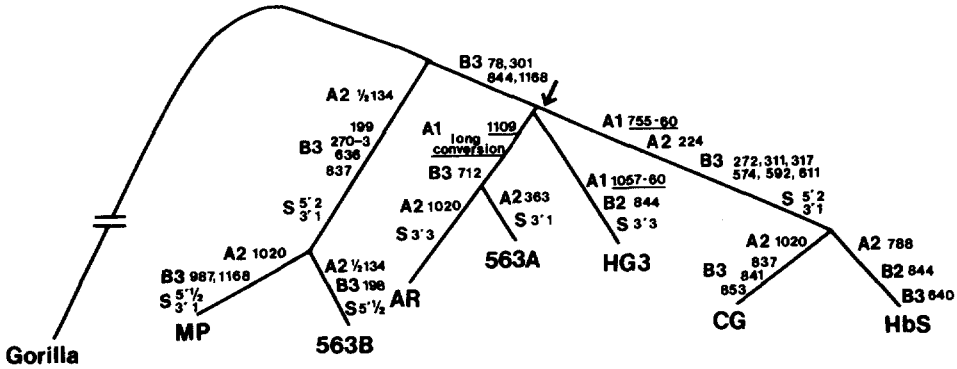

FIGURE 3.—Evolutionary tree of the seven chromosomes. An evolutionary tree was constructed using the human DNA sequence data shown in Figure 2. The nucleotide sequences from the gorilla G_γ and A_γ genes (SCOTT *et al.* 1984) were used to help establish the root of the tree. The mutational events, including gene conversions, single nucleotide changes and length differences, were scored as discussed in the text. The length of the branches of the tree are proportional to the number of mutational events that were placed along each branch. The approximate placement of each mutational event was recorded on the tree. Each mutational event resulting in a nucleotide substitution is identified by its category (see Figure 4) and by its location in the same nucleotide coordinates, as in Figures 1 and 2. Length differences in the region of simple sequence for each gene are identified by the symbol S, 5′ if it occurred in the 5′ γ gene or 3′ if it occurred in the 3′ γ gene, and the minimum number of mutational events necessary to account for each length difference. Those mutational events that could be placed in one of two branches were recorded in both branches as half an event. The mutational events of category A1 are underlined. An arrow marks the alternative placement (discussed in the text) of the branch leading to chromosome *HG3*.

chromosome *HG3*. Figure 3 shows the tree that placed chromosome *HG3* on the branch leading to chromosomes *AR* and *563A*. The other tree places chromosome *HG3* on the branch leading to chromosomes *HbS* and *CG*. This alternative placement of chromosome *HG3* is shown by an arrow in Figure 3; it does not alter any of our conclusions.

The evolutionary tree presented in Figure 3 records the approximate placement of each of the observed mutational events in the 14 kbp of DNA sequence considered in the study of these seven chromosomes. A few instances were encountered during construction of the tree where a mutation at either of two positions was equally compatible with the data. In such cases we scored the event as half an event at each of the two locations. A minimum of 50 mutational events have occurred since the last common ancestor of these seven human chromosomes. The tree shows no evidence for the occurrence of any single crossover events in the region between the two genes in any of the chromosomes, in that the same basic tree is obtained if the 5′ and 3′ γ genes are treated independently.

Categories of mutational events: We divided the types of mutational events into a total of six categories, comprised of two main divisions each containing three subdivisions, as shown in Figure 4. The two main divisions are: (A) mutational events at positions where the human 5′ and 3′ γ genes usually have different sequences and (B) mutational events at positions where the genes usually have the same sequence. The subdivisions are: (1) mutations that can

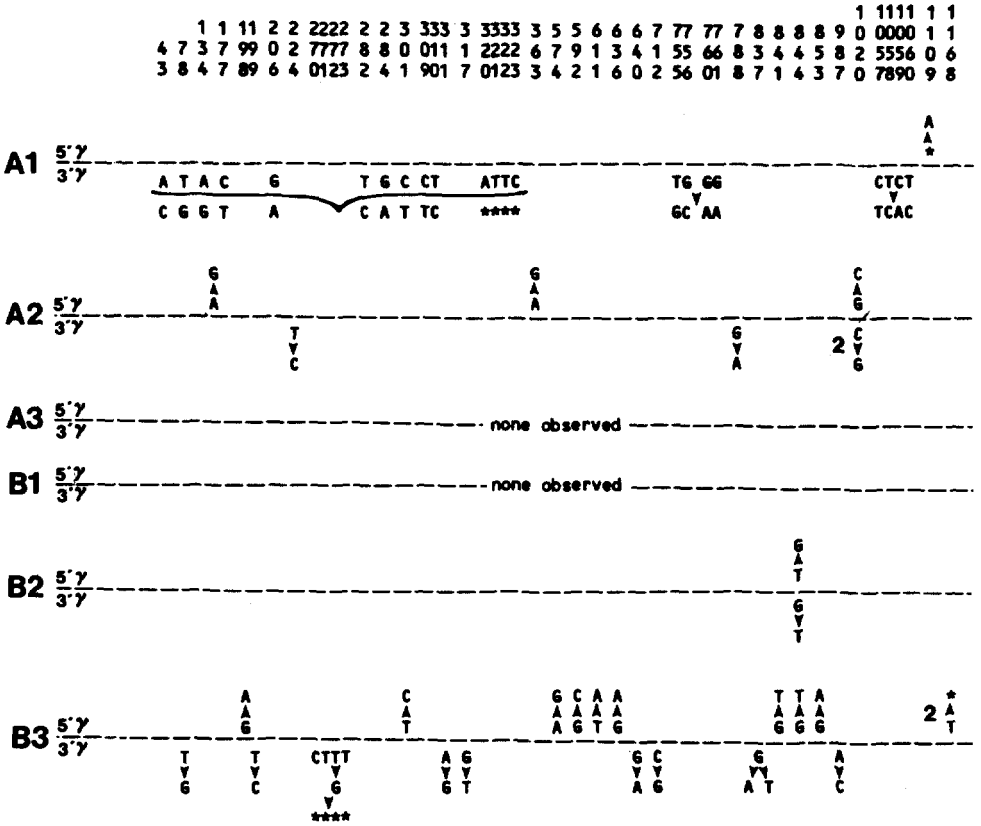

FIGURE 4.—Categories of mutational events. The nucleotide replacements within the DNA sequences of the 14 γ genes are divided into two main divisions: (A) mutational events at positions where the 5′ and 3′ γ genes usually have different sequence; (B) mutational events at positions where the 5′ and 3′ γ genes usually have the same sequence. Three subdivisions were made of the events within the two main divisions (A and B): (1) mutations that are most easily explained as the result of gene conversions, (2) mutations that are either the result of gene conversions or of single nucleotide changes, and (3) mutations that cannot be explained as the result of gene conversions. The nucleotide coordinates of each mutational event are given at the top of the figure in vertical orientation. The six categories of nucleotide replacements are shown. In each category, the mutational events occurring in the 5′ γ genes are shown above the dashed line, and those occurring in the 3′ γ genes are shown below the dashed line. For each mutational event, the original nucleotide(s) is shown closer to the dashed line and the substituted nucleotide(s) is shown further from the dashed lines, with the direction of the nucleotide change designated by an arrow. Because the A2 mutational event at position 1020 and the B3 mutational event at position 1168 each occurred twice, they are marked with the number 2. The large horizontal bracket encloses the 14 linked nucleotide substitutions associated with the long gene conversion described originally by SLIGHTOM, BLECHL and SMITHIES (1980).

be interpreted to be the result of a transfer of nucleotides between two γ genes by gene conversion, but cannot be accounted for by either a single nucleotide change or by the deletion or addition of a random number of nucleotides (*i.e.*, by a number different from that already observed at the same position); (2) mutations that can be accounted for either by a gene conversion

or by a single nucleotide change; and (3) mutations that cannot be accounted for by a gene conversion between any human γ-globin genes of known sequence.

We observed four mutational events in the A1 category, which consists of those cases where the data are difficult to explain except by a gene conversion that has transferred sequences from one γ gene to another. One of the four is the long gene conversion (nucleotide positions 1 through 323), first described in the A_γ gene of chromosome *563A* by SLIGHTOM, BLECHL and SMITHIES (1980). We found a second example of this gene conversion in the 3′ A_γ gene of chromosome *AR*.

The second A1-category event occurred in the common ancestor to chromosomes *HbS* and *CG* and resulted in the apparent transfer of the nucleotide pairs GC (positions 755–756) and AA (positions 760–761) from a G_γ gene to the A_γ gene of this common ancestor. Other interrelated changes are present in this region of chromosomes *HbS* and *CG* that make it impossible to give a completely unambiguous description of the mutational history of the region. The nucleotide at position 1020 in the 3′ γ gene of chromosome *CG* is a G, which is characteristic of G_γ genes, whereas the nucleotide at position 788 is G, which is characteristic of A_γ genes. Similarly, the 3′ A_γ gene of chromosome *HbS* contains the G_γ-like nucleotide A at position 788 and the A_γ-like nucleotide C at position 1020. The data can be explained by the occurrence of at least one A1-category gene conversion, extending a minimum of seven nucleotides (from positions 755 through 761) or a maximum of 344 nucleotides (from positions 713 through 1056), plus the substitution of at least one additional nucleotide either at position 788 in the 3′ G_γ gene of chromosome *CG* or at position 1020 in the 3′ A_γ gene of chromosome *HbS*. The additional nucleotide substitutions at position 788 or position 1020 are both category A2 events; that is, they could be due either to a gene conversion or to a nucleotide substitution that mimics a gene conversion.

The third A1-category event resulted in the apparent transfer of the nucleotides TCAC at positions 1057–1060, normally characteristic of G_γ genes, to the A_γ gene of chromosome *HG3*. Chromosome *HG3* was selected for study because it has the polymorphic *Hin*dIII site in the IVS2 of both of its γ genes, an arrangement that makes it a candidate for having a gene conversion. This *Hin*dIII (+ +) haplotype is the result of the replacement of a G by a T at position 844 in the 3′ γ gene, and it is located 210 bp away from the sequence TCAC at positions 1057–1060. The introduction at the T at position 844 and of the sequence TCAC at positions 1057–1060 could both be the result of a single event. If this is true, then it was probably a "patchy" gene conversion, because the 3′ γ gene in chromosome *HG3* has a C at position 1020; therefore, it is still an A_γ gene. Alternatively, the event leading to the T at position 844 could have occurred independently of the event leading to the sequence TCAC at positions 1057–1060. In either case, the sequence TCAC in the A_γ gene of chromosome *HG3* is extremely difficult to explain, except as the result of a gene conversion involving a minimum of four nucleotides (positions 1057–1060) and a maximum of 282 nucleotides (positions 789 through 1070).

The fourth A1-category event is the insertion of an A at position 1109 in the G_γ gene of the common ancestor of chromosomes *563A* and *AR*, making these two G_γ genes like A_γ genes at this position. This addition is readily accounted for as the consequence of a gene conversion of up to 60 nucleotides in length (positions 1072 through 1131). However, because the molecular mechanisms responsible for the addition or deletion of single nucleotides are not completely understood, we cannot exclude the possibility that local DNA sequences may have facilitated the addition of this nucleotide in some unrecognized way.

Seven examples of mutational events of category A2 were observed, *i.e.*, the replacement of a single nucleotide in one γ gene by the nucleotide that is present at the corresponding position in the linked γ gene. Three of these nucleotide substitutions occurred at position 1020, in the 5′ A_γ gene of chromosome *AR* and in the two 3′ G_γ genes of chromosomes *CG* and *MP*. These three transversions were selected by us because of our choice of chromosomes. The remaining four A2-category nucleotide substitutions at positions 134, 224, 363 and 788, all transitions, were not selected by us. We have no direct means of determining whether these replacements are due to gene conversions or to nucleotide changes that mimic gene conversions. However, we present below statistical arguments indicating that the latter explanation is very unlikely to account for all seven A2-category nucleotide substitutions.

No mutational events of category A3 were found, *i.e.*, no instances were found where a nucleotide at a position where the 5′ and 3′ γ genes differ was replaced by a nucleotide different from the one at the same position in either gene. There are at least two possible explanations for our finding seven nucleotide substitutions in category A2 and none in category A3. It is possible that the only viable nucleotides at the relevant positions are the two already observed in either the G_γ or A_γ genes. The three examples of nucleotide substitutions at position 1020 within the coding region may be, in part, explicable in this way. This explanation is unconvincing, however, for the other four positions at which we have identified A2 mutational events. All of these positions lie well within the large intervening sequence and, thus, are unlikely to be subject to much selection. The other possibility is that short gene conversions between the two fetal globin genes actually occur more frequently than do random nucleotide substitutions. This latter explanation is favored by the statistical arguments presented below.

We found no unequivocal examples of category B1 mutational events, *i.e.*, examples of putative gene conversions at positions where the G_γ and A_γ genes usually have the same sequence. This is not surprising because, in order to meet the criteria that we have set, the putative gene conversions would have to include at least two positions. Furthermore, to be scored in category B1, these substitutions would have to occur in a branch of the tree at such a position that the event would be recognized, and not merely cause the tree to be rearranged slightly.

Two examples of category B2 mutational events were observed, *i.e.*, nucleotide substitutions at positions where the G_γ and A_γ genes are usually alike that

might be due to gene conversions or to nucleotide substitutions mimicking gene conversions. One of the two examples was preselected by our initial choice of chromosomes; it is a change from G to T at position 844 that results in the presence of a polymorphic *Hin*dIII site in the A_γ gene of chromosome *HG3*. The other B2 mutational event was not preselected, but is also a change at position 844 that causes the T at position 844 in the 5′ γ gene of the common ancestor of chromosomes *563A, AR, HG3, HbS* and *CG* to be replaced by a G in chromosome *HbS*.

The coincidence of our finding an unselected B2 event in the G_γ gene of chromosome *HbS* at position 844, where we had already observed several other preselected events, requires comment. We have found three short A1-category gene conversions, four unselected A2-category mutations and the unselected B2-category mutation involving position 844. The probability that one of these eight possible gene conversions would overlap a specific nucleotide is >0.1 if the average length of the gene conversions is only 16 nucleotides. Thus, the coincidence of finding a possible unselected gene conversion at position 844, where we also observed preselected events, is not unexpected. The DNA sequence surrounding this particular B2 mutational event provides additional support for the suggestion that this nucleotide substitution arose as the consequence of a gene conversion. The closely related 5′ G_γ genes of chromosomes *HbS* and *CG* differ by three residues within a region of only 13 nucleotides. A single gene conversion of either of these G_γ genes by an unidentified donor γ gene could account for this cluster of differences.

Category B3 is by far the largest category, *i.e.*, nucleotide substitutions or length changes, occurring at positions where the G_γ and A_γ genes usually have the same sequence, that cannot be interpreted as gene conversions. There are 22 such events; ten are transversions, nine are transitions and three are length differences. Two different nucleotide substitutions occur at one position (position 837 of the 3′ γ genes) in two different branches of the evolutionary tree. A coincidence of this type is to be expected ($P = 0.16$) if 19 nucleotide substitutions are distributed randomly among the not more than 962 positions (see discussion below) at which replacements are likely to occur without negative or positive selection [$1/962 + 2/962 + \cdots + 18/962 = 0.16$]. Also included in category B3 are two instances of a length difference of a single nucleotide at position 1168. The 5′ γ genes of chromosome *563B*, of gorilla (SCOTT *et al.* 1984) and of chimpanzee (J. SLIGHTOM, personal communication), all lack the T at position 1168. It appears that there is no T at position 1168 in the ancestral 5′ γ gene. Hence a T must have been independently inserted in the 5′ γ gene of chromosome *MP* and in the 5′ γ gene of the common ancestor of chromosomes *563A, AR, HG3, CG* and *HbS*. This particular length difference results in a run of either seven or eight T residues; therefore, it is not surprising to find it occurring more than once.

Statistical considerations of the category A2 events: We investigated statistical arguments to help decide whether the observed events of category A2 are due to nucleotide substitutions that mimic gene conversions or to *bona fide* gene conversions. First, we estimated the number of positions not likely to be

subject to appreciable selection. Of the 1176 nucleotides considered, 19 nucleotides from the second exon and 129 nucleotides from the third exon are coding sequences and are likely to be under some selection. The donor and acceptor RNA splicing sites of the human globin genes probably contain a maximum of ten and 30 selected nucleotides, respectively (R. SPRITZ, personal communication). The 3′ untranslated region probably contains only six selected nucleotides, the sequence AATAAA (EFSTRATIADIS *et al.* 1980; FITZGERALD and SHENK 1981; WOYCHIK *et al.* 1984). The average length of the simple sequences (from position 591 through 658) is about 20 nucleotides less than the maximum length that we used in our numbering system. A region of 210 nucleotides (from positions 364 through 573) contains no differences in any of the 14 human genes considered here, suggesting that it might be under some selection (see also SLIGHTOM, BLECHL and SMITHIES 1980). However, both the gorilla and chimpanzee γ genes have several differences within this region (SCOTT *et al.* 1984; J. SLIGHTOM, personal communication), which suggests the contrary. Because of this uncertainty, the following calculations were made in two ways: (1) assuming that these 210 nucleotides are subject to selection and (2) assuming that they are not. The total number of positions not likely to be restricted by selection is either 962 [1176−(19 + 129 + 10 + 30 + 6 + 20)] or 752 [962 − 210].

Among the 962 (or 752) positions within the region under examination that can reasonably be considered as not under any appreciable selection, there are 20 positions at which most of the G_γ and A_γ genes usually differ by a single nucleotide; thus, there are 20 positions at which category A2 and A3 mutations can occur. Category A2 mutations were observed at four of these 20 positions. There are 942 (or 732) neutral positions at which the G_γ and A_γ genes do not usually differ and at which, therefore, category B2 and B3 mutations can occur. We observed 21 B3-category mutations and two B2-category mutations. If the observed nucleotide substitutions occur randomly among all of the 962 (or 752) neutral positions, we would expect about 47 = (942/20) [or 37=(732/20)] times as many B2 and B3 events as A2 and A3 events. The fact that, in regions not under selection, we observed two B2 events, 21 B3 events, four A2 events and no A3 events, suggests that the A2 events are about nine times more frequent than would be expected by chance. Furthermore, the A2 mutational events not only must occur at positions where the 5′ and 3′ γ genes differ, but they must also substitute the nucleotide that is already present in the linked γ gene. If it is equally likely that any one of the three nucleotides could be substituted at each of these positions, then only one-third of the nucleotide substitutions at positions where the G_γ and A_γ genes usually have different sequences would lead to an A2 event. Consequently, the A2 mutational events may actually be up to 27 times more frequent than would be expected by chance. The high frequency of A2 mutational events led us to examine more closely each event, in an attempt to determine whether there might be selection in favor of specific nucleotide replacements at the positions at which the 5′ and 3′ γ gene already differ.

A careful examination of the local DNA sequence surrounding each position

where A2- and B2-category mutational events have occurred did not reveal the presence of any identifiable DNA sequence feature, such as a palindrome (RIPLEY and GLICKMAN 1982) or direct repeat (EFSTRATIADIS *et al.* 1980; ALBERTINI *et al.* 1982), that might promote specific nucleotide substitutions. Nor can any of these recurring mutational events be explained as the consequence of deamination of the 5-methylcytosine residue found in some CpG dinucleotides (BIRD 1980; MCCLELLAND and IVARIE 1982). Nor does an examination of the nucleotides at the corresponding positions in the fetal globin genes of gorilla (SCOTT *et al.* 1984) and chimpanzee (J. SLIGHTOM, personal communication) suggest a predisposition in favor of these particular mutational events. Since we could find no features of the DNA sequence that suggest a predisposition to these particular nucleotide substitutions, we conclude that at least some of the category A2 replacements are more likely to be the result of gene conversion than the result of random nucleotide substitutions that mimic gene conversions.

It must be emphasized that the proportion of *gene conversions* that we can detect is considerably less than the proportion of *nucleotide substitutions* that we can detect, because we cannot observe gene conversions in regions where the two genes do not differ. Despite this inate inability to score all gene conversion events, our data provide evidence for at least four and up to 13 short gene conversions and one long gene conversion within about a 1-kbp region in the 14 fetal globin genes that have been examined. Our statistical analysis suggests that the lower estimate of the number of gene conversions is very unlikely.

A complication in making an accurate estimate of the number of gene conversions stems from one of the assumptions made in drawing the evolutionary tree depicted in Figure 3. The tree is one of two that minimize the total number of mutational events. Different trees can be drawn that, in exchange for having a greater number of mutational events, can avoid one or other of the gene conversions. For example, the A1 gene conversion event at positions 755–761 can be eliminated at the price of requiring four B3-category mutations at positions 78, 301, 844 and 1168 to have occurred twice: once in the branch leading to chromosomes *HbS* and *CG* and once in the branch leading to chromosomes *AR, 563A* and *HG3*. We found that each such elimination of a gene conversion required either a different tree or the further escalation of the number of additional repeated mutations. We conclude, therefore, that our estimate is reasonable, although it could be modified if more repeated mutations were accepted.

Simple sequence length differences are generated by intrachromosomal events: Although we could find no evidence for interchromosomal crossing-over events within the seven fetal globin gene regions, as judged by a failure to find sequences juxtaposed from two different branches of our evolutionary tree, it is apparent that crossing over between closely related fetal globin regions could easily result in recombinants that we would be unable to detect. Some of our data are relevant to this possibility. In particular, we observed a minimum of 15 length differences in the number of (TG) or (CG) doublets in

the regions of simple sequence. Since SLIGHTOM, BLECHL and SMITHIES (1980) suggested that this region was a "hotspot" for recombination, we first thought that these length differences might have been generated by unequal crossing over between homologous chromosomes. However, in each case where closely related chromosomes differ within the region of simple sequence (*e.g.*, the 5′ G_γ genes of chromosomes *563B* and *MP* and the 3′ A_γ genes of chromosomes *563A* and *AR*), the two chromosomes still share characteristic nucleotide substitutions on *both* the 5′ and 3′ sides of the region of simple sequence. Thus, if these length differences are due to unequal crossing over, the participants most likely were sister chromatids, rather than homologous chromosomes. An alternative possibility is that the length differences are due to some type of intramolecular slipping during replication (*cf.* EFSTRATIADIS *et al.* 1980) and not to crossing over.

Gene conversion between 5′ and 3′ γ genes or between alleles?: All four of the category A2 gene conversions and most of the category A2 and B2 events that we have detected involve the exchange of DNA sequences between the 5′ and 3′ (nonallelic) γ genes. However, two of the category A2 and one of the category B2 mutational events could alternatively be interpreted as being the result of gene conversions between allelic genes. For example, the nucleotide difference at position 363 in the 5′ γ genes of chromosomes *AR* and *563A* is either the result of an A2 mutational event in the 5′ G_γ gene of chromosome *563A* involving a nonallelic 3′ γ gene as a donor or it is the result of an A2 mutational event in the 5′ A_γ gene of chromosome *AR* involving an allelic 5′ γ gene as a donor. Although the number of cases is small where gene conversions between alleles can explain our findings, it is still possible that gene conversions between allelic genes could occur equally or even more frequently than gene conversions between nonallelic genes. The much greater level of sequence identity between alleles, compared to the level between nonalleles, may have prevented us from identifying these interallelic gene conversions in an unequivocal way.

DISCUSSION

In summary, we have identified between four and 13 new examples of gene conversion between the human γ-globin genes. With one exception, all of the gene conversions identified in this study are short, less than 300 bp in length. Most of the detected gene conversions are likely to be due to the exchange of DNA sequence between nonallelic genes. Only three of these potential gene conversion, each involving a single nucleotide substitution, are possibly the result of gene conversions between allelic genes.

The observation of between four and 13 gene conversions within the region under examination and the absence of any detectable crossing over within any of the seven chromosomes studied suggests that the gene conversions may have arisen as the consequence of some general process other than that of crossing over. Yet, any mechanism that leads to gene conversions must involve the physical interaction between homologous sequences. It is likely that such physical interactions occur with regularity in the germ line of higher eukaryotes

only during meiosis. Biochemical studies of the zygotene and pachytene stages of meiosis suggest that single-stranded DNA is produced from some regions along each chromosome during zygotene (HOTTA, CHANDLEY and STERN 1977; STERN and HOTTA 1977; HOTTA, TABATA and STERN 1984). These regions of single-stranded DNA could, in principle, invade homologous double-stranded DNA molecules. It may be that such interactions between single-stranded DNA molecules and intact duplexes occasionally lead to short gene conversions. We have proposed elsewhere (SMITHIES and POWERS 1985) that local DNA strand invasions form the molecular basis for the recognition of homologous chromosomes during meiosis. Specifically, we suggested that single-stranded regions of DNA extruded from sites along chromosomes invade nearly intact DNA duplexes and, perhaps under the influence of a *RecA*-type protein, scan the invaded duplex for homologous sequences. Stable heteroduplexes could thus be formed between homologous sequences. If a stable heteroduplex is formed in a region of homology between homologous chromosomes, then other nearby invasions will have formed stable heteroduplexes. Cooperation between such nearby heteroduplexes will lead consequently to pairing between homologous chromosomes. On the other hand, if a stable heteroduplex is formed as the consequence of a homologous sequence being found at a nonhomologous chromosomal location, then nearby invasions will not have led to stable heteroduplexes; thus, pairing between nonhomologous chromosomes will not be achieved. The relics of some of the nonhomologous invasions, which normally lead only to abortive pairing, may be what is being observed as short gene conversions. Since the proximity of sequences will most likely influence the probability that they will meet, nonallelic gene conversions between closely linked genes may be seen more often than conversions between genes at widely separated loci.

In conclusion, we propose that the short gene conversions observed between the human fetal globin genes may be the relics of abortive pairing between the closely related but nonallelic fetal globin genes and that short gene conversions of this type will prove to be very common within duplicated regions in the genomes of higher eukaryotes.

We are indebted to T. H. J. HUISMAN and his colleagues for providing DNA and tissue from the individuals *A.R.*, *C.G.* and *M.P.* We thank FRANK GROSVELD for the cosmid clone *HG3*, and BERNARD FORGET for tissue from the individual P.P. We are especially grateful to JERRY SLIGHTOM for helpful comments during the course of this research and for generously sharing unpublished data. We also thank SHI-HSIANG SHEN for sharing unpublished data. We thank SYLVIA ROBERTSON and WALTER FITCH for assisting us in determining the evolutionary tree, using their ANCSTR program. This work was supported by National Insititutes of Health grants GM20069 and AM20120 to O.S. In addition, P.A.P. was supported by National Institutes of Health training grant T32GM07133. This is paper 2784 from the Laboratory of Genetics, University of Wisconsin, Madison.

LITERATURE CITED

ALBERTINI, A. M., M. HOFER, M. P. CALOS and J. H. MILLER, 1982 On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. Cell **29:** 319–328.

BENTLEY, D. L. and T. H. RABBITTS, 1983 Evolution of immunoglobulin V genes: evidence indicating that recently duplicated human V_k sequences have diverged by gene conversion. Cell **32:** 181–189.

BIRD, A. P., 1980 DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. **8:** 1499–1504.

BLATTNER, F. R., A. E. BLECHL, K. DENNISTON-THOMPSON, H. FABER, J. E. RICHARDS, J. L. SLIGHTOM, P. W. TUCKER and O. SMITHIES, 1978 Cloning human fetal γ-globin and mouse α-type globin DNA: preparation and screening of shotgun collections. Science **202:** 1279–1284.

DEVEREUX, J., P. HABERLI and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

EFSTRATIADIS, A., J. W. POSAKONY, T. MANIATIS, R. M. LAWN, C. O'CONNELL, R. A. SPRITZ, J. K. DERIEL, B. G. FORGET, S. M. WEISSMAN, J. L. SLIGHTOM, A. E. BLECHL, O. SMITHIES, F. E. BARALLE, C. C. SHOULDERS and N. J. PROUDFOOT, 1980 The structure and evolution of the human β-globin gene family. Cell **21:** 653–668.

FITZGERALD, M. and T. SHENK, 1981 The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell **24:** 251–260.

GROSVELD, F. G., H.-H. M. DAHL, E. DEBOER and R. FLAVELL, 1981 Isolation of β-globin related genes from a human cosmid library. Gene **13:** 227–237.

HARDISON, R. C. and J. B. MARGOT, 1984 Rabbit globin pseudogene ψβ2 is a hybrid of δ- and β-globin gene sequences. Mol. Biol. Evol. **1:** 302–316.

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

HOTTA, Y., A. C. CHANDLEY and H. STERN, 1977 Meiotic crossing-over in lily and mouse. Nature **269:** 240–242.

HOTTA, Y., S. TABATA and H. STERN, 1984 Replication and nicking of zygotene sequences; control by a meiosis-specific protein. Chromosoma **90:** 243–253.

JACKSON, J. A. and G. R. FINK, 1981 Gene conversion between duplicated genetic elements in yeast. Nature **292:** 306–311.

JEFFREYS, A. J., 1979 DNA sequence variants in the G_γ-, A_γ-, δ- and β-globin genes of man. Cell **18:** 1–10.

KLAR, A. J. S. and J. N. STRATHERN, 1984 Resolution of recombination intermediates generated during yeast mating-type switching. Nature **310:** 744–747.

KLEIN, H. L., 1984 Lack of association between intrachromosomal gene conversion and reciprocal exchange. Nature **310:** 748–751.

KLEIN, H. L. and T. D. PETES, 1981 Intrachromosomal gene conversion in yeast. Nature **289:** 144–148.

LIEBHABER, S., M. GOOSSENS and Y. W. KAN, 1981 Homology and converted evolution at the α1 and α2 loci of human α-globin. Nature **290:** 26–29.

MAXAM, A. M. and W. GILBERT, 1977 A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74:** 560–654.

MCCLELLAND, M. and R. IVARIE, 1982 Asymmetric distribution of CpG in an "average" mammalian gene. Nucleic Acids Res. **10:** 7865–7877.

MELLOR, A. L., E. H. WEISS, K. RAMACHANDRAN and R. A. FLAVELL, 1983 A potential donor for the *bm1* gene conversion event in the C57BL mouse. Nature **306:** 792–795.

MESELSON, M. S. and C. M. RADDING, 1975 A general model for genetic recombination. Proc. Natl. Acad. Sci. USA **72:** 358–361.

MICHELSON, A. M. and S. ORKIN, 1983 Boundaries of gene conversion within the duplicated human α-globin genes. J. Biol. Chem. **258:** 15245–15254.

PETES, T. and G. R. FINK, 1982 Gene conversion between repeated genes. Nature **300:** 216–217.

PONCZ, M., D. SOLOWIEJCZYK, B. HARPEL, Y. MORY, E. SCHWARTZ and S. SURREY, 1982 Construction of human gene libraries from small amounts of peripheral blood: analysis of β-like globin genes. Hemoglobin **6:** 27–36.

POWERS, P. A., C. ALTAY, H. J. HUISMAN and O. SMITHIES, 1984 Two novel arrangements of the human fetal globin genes: G_γ-G_γ and A_γ-A_γ. Nucleic Acid Res. **12:** 7023–7034.

RIPLEY, L. S. and B. W. GLICKMAN, 1982 Unique self-complementarity of palindromic sequences provide DNA structural intermediates for mutation. Cold Spring Harbor Symp. Quant. Biol. **47:** 851–862.

SCOTT, A. F., P. HEATH, S. TRUSKO, S. H. BOYER, W. PRASS, M. GOODMAN, J. CZELUSNIAK, L.-Y. E. CHANG and J. L. SLIGHTOM, 1984 The sequence of the gorilla fetal human loci. Mol. Biol. Evol. **1:** 371–389.

SCHULZE, D. H., L. R. PEASE, S. S. GEIER, A. A. REYES, L. A. SARMIENTO, R. B. WALLACE and S. G. NATHENSON, 1983 Comparison of the cloned *H-2K*^{bm1} variant gene with the *H-2K*^{b} gene shows a cluster of seven nucleotide differences. Proc. Natl. Acad. Sci. USA **80:** 2007–2011.

SHEN, S., J. L. SLIGHTOM and O. SMITHIES, 1981 A history of the human fetal globin gene duplication. Cell **26:** 191–203.

SLIGHTOM, J. L., A. E. BLECHL and O. SMITHIES, 1980 Human fetal G_γ- and A_γ-globin genes: complete nucleotides sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21:** 627–638.

SMITHIES, O. and P. A. POWERS, 1985 Gene conversions and their relationship to homologous chromosome pairing. Philos. Trans. R. Soc. Lond. [Biol.] In press.

STERN, H. and Y. HOTTA, 1977 Biochemistry of meiosis. Philos. Trans. R. Soc. Lond. [Biol.] **277:** 277–294.

STOECKERT, C. J., JR., F. S. COLLINS and S. M. WEISSMAN, 1984 Human fetal globin DNA sequences suggest novel conversion event. Nucleic Acids Res. **12:** 4469–4479.

WEISS, E., L. GOLDEN, R. ZAKUT, A. MELLOR, K. FAHRNER, S. KVIST and R. A. FLAVELL, 1983 The DNA sequence 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.

WOYCHIK, R. P., R. J. LYONS, L. POST and F. M. ROTTMAN, 1984 Requirement for the 3′ flanking region of the bovine growth hormone gene for accurate polyadenylation. Proc. Natl. Acad. Sci. USA **81:** 3944–3948.

Communicating editor: B. W. GLICKMAN