Rhesus Fetal Globin Genes

CONCERTED GENE EVOLUTION IN THE DESCENT OF HIGHER PRIMATES*

(Received for publication, January 21, 1988)

Jerry L. Slightom§, Ben F. Koop§, Peilin Xu§, and Morris Goodman§

From the §Molecular Biology Division, The Upjohn Company, Kalamazoo, Michigan 49007 and the §Department of Anatomy and Cell Biology, Wayne State University School of Medicine, Detroit, Michigan 48201

The comparison of the nucleotide sequences of closely linked duplicated genes of higher eukaryotes has been important in the identification of molecular events that shape the evolution of mammalian genes, most notably recombinational events such as unequal crossovers and gene conversions. Toward this goal we have been comparing the nucleotide sequences of the paired $\gamma^+$- and $\gamma^+$-globin genes from species of catarrhine primes. Previous comparisons document that, within each great ape species as in humans, the paired $\gamma$-genes have been involved in gene conversion events. We now extend our analysis to the catarrhine superfamily Cercopithecoida by obtaining the nucleotide sequence of the paired $\gamma^+$- and $\gamma^+$-genes of rhesus monkey (Macaca mulatta). The rhesus $\gamma^+$- and $\gamma^+$-genes diverge less from each other than from human, chimpanzee, gorilla, and orangutan $\gamma^+$- or $\gamma^+$-genes. This finding indicates that a species-specific gene conversion occurred between rhesus $\gamma^+$- and $\gamma^+$-genes. This $\gamma$-gene conversion (labeled C14 in our series) involved at least 1898 base pairs, extending across the complete transcriptional region of the rhesus $\gamma$-genes. C14 could have resulted from a single large conversion or several short conversion events which may have involved the (TG), repetitive sequence element. Parsimony analysis of the emerged $\gamma$-gene sequence data also strengthened the evidence for the 14 previously suggested $\gamma$-gene conversion events: labeled C2, C3, and C4 in Homo; C5, C6, and C7 in Pan; C8, C9, and C10 in Gorilla; C11, C12, C13 in Pongo; C1 in the stem to Homininae (the subfamily of Homo, Pan, and Gorilla) and C0 in the stem of Hominidae (the family of Pongo and Homininae).

During the evolution of eukaryotic genomes, genes have duplicated via nonhomologous and homologous chromosomal breakage and rejoicing. In time, the duplicate genes have diverged as a result of base pair substitutions, deletions, and insertions. These processes resulted in the establishment of multigene families that encode protein products of similar, but not necessarily identical, functions. Diverging members of gene families have become in some cases pseudogenes if their function is not needed and, in other cases, as a result of positive selection, have become founders of new gene types which encode proteins that have acquired new functions. Recently duplicated genes share identical nucleotide sequences, thus their divergence from each other may not be independent because of their involvement in recombinational events such as unequal crossing over and gene conversion. Unequal crossing over events can increase or decrease the number of genes, whereas gene conversion events can decrease, as in the case of concerted evolution (Maeda and Smithies, 1986; Walsh, 1987), or increase, as in the case of antigenic variation in trypanosomes (Pays et al., 1983), divergence between duplicated genes.

Nucleotide sequence analysis of gene members of the $\alpha$- and $\beta$-globin gene families has yielded many examples of both unequal crossover events (Zimmer et al., 1980; Shen et al., 1981; Hill et al., 1986) and gene conversion events (Slighet al., 1980; Michelson and Orkin, 1983; Scott et al., 1984; Slighet al., 1985, 1987). Additional examples of recombinational events have been identified in other multigene families such as haptoglobin genes (Maeda et al., 1984), human and mouse immunoglobulin genes (Bentley and Rabbits, 1983; Flanagan et al., 1984; Hayashida et al., 1984), and mouse histocompatibility genes (Schulze et al., 1983; Weiss et al., 1983).

The $\beta$-type globin gene clusters of catarrhine primates are highly conserved, consisting of six paralogously related genes which are arranged (5' to 3') in the order of their developmental expression: 5'-(embryonic), duplicated $\gamma^+$ and $\gamma^+$ (fetal), inactive $\gamma^+$ (a pseudogene), and $\delta$ and $\beta$ (adult) (Efstratiadis et al., 1980; Barrie et al., 1981; Zimmer, 1981; Scott et al., 1984; Goodman et al., 1984; Chang and Slighet, 1984; Slighet al., 1985, 1987). The linked $\gamma^+$- and $\gamma^+$-genes are physically located within a 10-kb DNA region which arose as a result of a 5-kb duplication (Shen et al., 1981). This duplication occurred about 25 to 35 million years ago in the stem of the Catarrhini branch after divergence of Platyrrhini but before the divergence of superfamilies Cercopithecoida (old world monkeys) and Hominoidae (Homo, Pan, Gorilla, Pongo, and Hylobates) (Barrie et al., 1981; Koop et al., 1986a, 1986b; Zimmer, 1981).

In order to determine the type of molecular events which took place between the members of a duplicated gene pair, we have isolated and sequenced the closely linked $\gamma$-globin gene pair from the major group of higher primate species. We have previously obtained the nucleotide sequence of the $\gamma^+$- and $\gamma^+$-genes of human (Homo sapiens), chimpanzee (Pan troglodytes), gorilla (Gorilla gorilla), and orangutan (Pongo pygmaeus) and used the parsimony method (Goodman et al., 1984) to identify, within the paired $\gamma$-genes of each hominoid species, homologous regions which have been affected by gene

* This study was supported by National Institutes of Health Grant HL33940 (to M. G. and J. L. S.), National Science Foundation Grant BSR 83-07930 (to M. G.), and Alfred P. Sloan Foundation Award (to M. G.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J03938.

§Present address: Dept. of Plant Pathology, University of Wisconsin, Madison, WI 53706.

1 The abbreviations used are: kb, kilobase pairs; bp, base pairs.
conversion events. Thus far we have identified conversion regions accumulatively labeled C0 to C13 (Slightom et al., 1980; Scott et al., 1984, Slightom et al., 1985, 1987). Many of these gene conversion events appear to have been initiated at the boundaries of a stretch of simple-sequence DNA, element, located in intron 2. This (TG)n element may provide a "hot spot" for strand breakage and transfer between related homologous regions in the genes (Slightom et al., 1980; Kilpatrick et al., 1984). Interestingly, our previous analyses indicate that conversion directions are not random; most γ-gene conversions, except C5 in chimpanzee (Slightom et al., 1985), have γ'-gene sequences (donor sequence) superimposed onto γ'-gene sequences (acceptor sequence) (see Slightom et al., 1987).

In this report we present the nucleotide sequence of the γ1- and γ2-globin gene of a rhesus monkey (Macaca mulatta) and, on the basis of this additional sequence data on γ-globin genes, extend our reconstruction of the evolutionary history of the duplicated catarrhine γ-genes to the two superfamilies that comprise the infraorder Catarrhini, Cercopithecoidea, and Hominoidea. The single γ-gene sequences from spider monkey (Giebel et al., 1985), brown, and dwarf lemur (Harris et al., 1986) were used as out groups of the catarrhine γ-gene sequences. Comparison of the rhesus γ1- and γ2-genes shows that they are nearly identical over their complete transcriptional region and therefore appear to have been involved in an extensive (over 1985 bp) gene conversion event (or a series of smaller conversions) labeled C14.

**EXPERIMENTAL PROCEDURES**

**Materials**—Restriction endonucleases were obtained from either Bethesda Research Laboratories or New England Biolabs and used as specified by the vendor. Polynucleotide kinase was obtained from U. S. Biochemicals and bovine alkaline phosphatase was obtained from Boehringer Mannheim. Radioactive nucleotides were from Amersham Corp. T4 ligase was obtained from Collaborative Research and nitrocellulose filter membranes (BA-85) were obtained from Schleicher & Schuell. Chemicals used for DNA sequencing were obtained from Boehringer-Mannheim. 3'-End labeling kit was obtained from Amersham Corp. T4 ligase was obtained from Collaborative Research and nitrocellulose filter membranes (BA-85) were obtained from Schleicher & Schuell. Chemicals used for DNA sequencing were obtained from Boehringer-Mannheim. X-ray film (Kodak X-AR-351) and DNA sequencing gel stands and safety cabinets were obtained from Fotodyne, Inc. Intensifying screens (Quanta III) were obtained from Du Pont-New England Nuclear. Wedge spacers (60 cm long and thickness ranging from 0.2 mm (top) to 0.4 cm (bottom)) for forming bell-bottom gel molds were obtained from C. B. Scientific Co. Inc.

**Rhesus γ-Gene Cloning and DNA Isolation**—Total DNA from rhesus monkey (M. mulatta) was a gift from Dr. Sandy Martin (University of California, Berkeley, CA) and its isolation has previously been described (Martin et al., 1980). Purified rhesus DNA was partially digested with either MboI or EcoRI, and DNA fragments in the range of 15 to 20 kb were size-fractionated on 5 to 20% NaCl gradients (Slightom et al., 1980). Size-fractionated rhesus DNAs were cloned into the appropriate λ vector arms; MboI-cut fragments were cloned into BamHI-cut Charon 30 arms (Rimm et al., 1980), and EcoRI-cut fragments were cloned into EcoRI-cut Charon 32 arms (Loen and Blattner, 1983). Recombinant DNA phases were packaged into phage capsids using the in vitro phage-packaging procedure described by Hohn (1979), and both recombinant phage libraries contained between 1 to 2 x 109 phage. These libraries were plated on *E. coli* host (Charon 30 on DP50supF and Charon 32 on the recA ED8767), and the resulting phage plates were analyzed in situ by blotting with nitrocellulose filters as described by Benton and Davis (1977). Nitrocellulose filters were hybridized against 32P-labeled *EcoR*I fragment (245 bp) cut from the human γ-gene clone pWJ15 (Wilson et al., 1978). Recombinant clones containing the regions of the rhesus β-globin gene cluster were then used to pick out the linked γ1- and γ2-globin genes, which contain the rhesus γ-gene linkage region, as determined by sequence analysis and comparison with this region of the sequenced human γ-genes (Yen et al., 1981). The 8.8-, 2.6-, and 10.0-kb EcoRI fragments containing the rhesus γ1-, γ2-, and γ*-globin genes, respectively, were subcloned in the vector pBR322 (see Fig. 1). The resulting clones were designated pMu13.1-R8.8, pMu14.7-R2.6, and pMu14.7-R10.0 (see Fig. 1). Plasmid DNAs were purified using the ultracentrifugation method described by Birnboim and Doly (1979), followed by two CsCl-ethidium bromide gradient bandings in a type 70.1 rotor (fixed angle rotor) to ensure removal of contaminating small RNAs.

**DNA Sequencing**—Chemical sequencing was performed as described by Maxam and Gilbert (1980). Generally, 10 to 20 μg of plasmid or 50 μg of λ phage DNAs were digested with the appropriate restriction enzyme and either 5'- or 3'-end-labeled. The conditions for end labeling and chemical sequencing, to ensure long DNA sequence reads (greater than 600 bp), have been described previously (Chang and Slightom, 1984). A detailed description and procedure for the extraction of DNA from rhesus, brown, and dwarf lemurs (Harris et al., 1986) was used as an outgroup and a guide to the 100-cm gel gradient sequencing gels have been presented by Slightom et al. (1987). Most DNA sequences were determined from both DNA strands by either sequencing the strands from opposite directions or by doing both 5' and 3'-end labelings of the same restriction enzyme site. Even though the sequencing strategy (see Fig. 1) shows an arrow pointing in only one direction, sequence information was obtained from both strands. Nucleotide sequences that differ among the aligned γ-globin gene sequences (see Fig. 2) were closely analyzed to ensure the accuracy of the rhesus γ-gene sequences present here. This alignment provides an excellent position-by-position test for judging the accuracy of any new primate γ-gene sequence and each nucleotide of these rhesus γ-gene sequences (presented in Fig. 2) has passed this rigorous test.

**Gene Conversion and Evolutionary Analyses**—Using the principle of minimum change or parsimony as described in previous studies (Scott et al., 1984; Goodman et al., 1984; Slightom et al., 1985, 1987), we aligned the nucleotide sequences of 14 fetal globin genes (see Fig. 2) and then determined independently, at each variable position, the most parsimonious branching arrangement. At each variable position where two γ-genes of a catarrhine taxon shared a sequence identity, which the parsimony solution depicted as derived rather than primitive, the result was taken as evidence for a putative conversion encompassing that position in the y1- to γ2-gene pair. Conversely, absence of sequence identity could be taken as evidence against conversion for that gene pair at that position, depending on the parsimonious branching arrangement. On synthesizing the results of all the separate parsimonious solutions, we identified gene conversion events by minimizing the number of conversion events needed to account for the overall most parsimonious solutions.

**RESULTS AND DISCUSSION**

**Isolation, Nucleotide Sequence, and Derived Amino Acid Sequence**—The λ and phage DNAs were hybridized to isolate and sequence the rhesus γ-globin genes are shown in Fig. 1. Rhesus γ1- and γ2-genes are located on 8.8- and 2.6-kb EcoRI fragments, respectively, an organization similar to that found in other higher primates (Barrie et al., 1981; Zimmer, 1981; Slightom et al., 1980; Scott et al., 1984; Slightom et al., 1985, 1987). The λ clones shown in Fig. 1 do not physically link the two rhesus γ-genes, but λ clone MmuCh32-4 does contain a fragment that links λ clones MmuCh30-13.1 and MmuCh32-14.7. The nucleotide sequence of this fragment is orthologous (data not shown) to the nucleotide sequence of the linkage region between human γ-g- and γ2-genes (Shen et al., 1981). In addition, our sequence analysis of the 3'-flanking DNAs (Fig. 2) clearly indicate that these two rhesus γ-genes are not alleles. The organization of the rhesus γ-genes presented here is similar to that found in a closely related species, the yellow baboon (*Papio cynocephalus*) (Barrie et al., 1981; Zimmer, 1981). Fig. 1 shows the strategy used to obtain the nucleotide sequence of the rhesus γ1- and γ2-genes; in Fig. 2 the γ-gene
sequences from rhesus and other species are aligned against one another.

Amino acid sequence analysis of the \( \gamma \) chains isolated from fetal hemoglobin of a rhesus monkey revealed no amino acid heterogeneity at any of the 146 chain positions (Mahoney and Nute, 1980). However, the derived amino acid sequences of the two \( \gamma \)-genes (Fig. 2) differ due to the nucleotide replacement substitution at position 1487 (C/T), which results in an amino acid change in residue 117 (His/Tyr). The amino acid sequence determined by Mahoney and Nute (1980) indicates that a \( \gamma^2 \) allele which encodes His at this position apparently also exists in the rhesus monkey population; however, the predominant allele remains to be determined. The \( \gamma \) chains encoded by rhesus and human differ at two amino acid positions, 77 and 104, where rhesus shares amino acid residues Asn and Arg with orangutan and spider monkey, see Fig. 2 and Table I. Both rhesus \( \gamma \)-genes encode Gly at position 136 (consistent with the findings of Mahoney and Nute (1980)) as do both \( \gamma \)-genes of orangutan and the single \( \gamma \)-gene of spider monkey (Table I).

Conservation of \( \gamma \)-Gene Structure and Regulatory Elements—From the 14 aligned nucleotide sequences of \( \gamma \)-genes, obtained from eight different primate species (Fig. 2), it is apparent that all share a high degree of nucleotide sequence conservation of DNA elements known to be important for \( \gamma \)-gene expression and mRNA processing. For example the transcriptional promoter elements, CCAAT and AATAAA, are located within stretches of sequence which are virtually invariant. The imperfect direct repeats that resulted in the placement of two consecutive CCAAT elements (Efstratiadis et al., 1980) are present in all of these genes, except brown lemur, which supports the suggestion of an ancient origin for the duplicated CCAAT-containing sequence (Efstratiadis et al., 1980; Harris et al., 1986). The position and sequence of the 3'-regulatory poly(A) signal (Proudfoot and Brownlee, 1976) is also invariant. All of these genes contain GU-rich and U-rich sequence elements (positions 1669 to 1679 and 1686 to 1695, respectively) which may be essential for mRNA 3'-end formation (Gil and Proudfoot, 1987).

The length of intron 1 varies by 1 bp among these \( \gamma \)-genes; a T-nucleotide (position 255, Fig. 2), which is present in the lower primate \( \gamma \)-genes and orangutan \( \gamma \)-gene, is absent from orangutan \( \gamma \)- and Homininae \( \gamma \)- and \( \gamma^2 \)-genes. Intron 1 5' donor and 3' acceptor splice sequences are highly conserved;
Nucleotide Sequence of Rhesus Fetal Globin Genes
Nucleotide Sequence of Rhesus Fetal Globin Genes
**Fig. 2.** Nucleotide sequence alignment of primate γ-globin genes. From a single human the αγ- and α2γ-genes from chromosome "A" are denoted Hsa α2γ and Hsa αA, respectively, and the γ-gene from chromosome "B" is denoted Hsa BA (Slighlom et al., 1980; Shen et al., 1981). gorilla, Gg (Scott et al., 1984); chimpanzee, Ptr (Slighlom et al., 1985); orangutan, Ppy (Slighlom et al., 1987); spider monkey, Age (Giebel et al., 1985); and lemur, Lfu, and dwarf lemur, Lld (Harris et al., 1986). The numbering system used was set by the overall alignment of the γ-gene sequences; asterisks indicate gaps which were used to increase sequence identities among these genes. For any position where one sequence differs from another, the nucleotide (or asterisk) for that position is given for each gene sequence. Note nucleotide sequence elements that may have biological importance: promoter elements CCAAT and AATAAA and poly(A) addition signals are overlined. The γ-gene amino acid sequences are shown below the lower counting line, and amino acid differences are printed below the appropriate codon. The translation initiator codon is the first Met, and the terminator codon is designated TER. Vertical arrows show the location of exon-intron boundaries, all of which conform to the GT/AG rule (Breathnach et al., 1978). Horizontal arrows indicate the sizes expected for the 5'- and 3'-untranslated regions; however, note that the length of both these regions has only been substantiated for the human genes (see Slighlom et al., 1980) and the 3'-region of the spider monkey gene (Giebel et al., 1985).
TABLE I

Amino acid replacements encoded in primate γ-globin gene nucleotide sequences, amino acid residue number (nucleotide position)

| Primate | 1 | 2 | 5 | 10 | 18 | 33 | 47 | 51 | 65 | 73 | 75 | 77 | 86 | 104 | 117 | 118 | 135 | 136 | 142 |
|---------|---|---|---|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|
| Hsa γ_1  | Gly | His | Glu | Thr | Val | Val | Asn | Ala | Lys | Asp | Ile | His | Ala | Lys | His | Phe | Thr | Gly | Ser |
| Aγ_1     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |     |   |   |   |   |   |
| Ppy γ_1  |   |   |   |   | Leu |   | Asp | Asn |   |   |   |   |   |   |     |   |   |   |   |   |
| Aγ_2     |   |   |   |   | Leu | Leu | Asp | Val | Asn |   |   |   |   |   |     |   |   |   |   |   |
| Mmu γ_1  |   |   | Val | Val |   |   | Asp | Ile | Asn |   |   |   |   |   |     |   |   |   |   |   |
| Aγ_2     |   |   | Val | Val |   |   | Asp | Ile | Asn |   |   |   |   |   |     |   |   |   |   |   |
| Age γ    | Ser | Asn | Ala | Ala | Val | Ser | Pro | Val | Glu | Ile | Asn | Gly | Arg | Leu | His | Ala |   |   |   |

Only one difference is observed in the lemur sequences, a transversion (G to T/C) in position 7 of the donor consensus sequence (C or A-A-G/g-t-a or g-a-g-t; Mount, 1982). Similarly, variation in intron 1 acceptor sequences (from the consensus sequence (c or t),n-c or t-a-g/G) is found only in the lemur sequence which have a transition just after the splice junction (position 268 in Fig. 2).

In contrast, intron 2 sequences show a considerable amount of variation in both length and shared nucleotide sequence identity (Fig. 2). The shortest intron 2 sequence (849 bp) is found in dwarf lemur (Harris et al., 1986) and the largest is in the gorilla γ- (906 bp) (Scott et al., 1984); rhesus intron 2 sequences contain 895 and 887 bp for the γ_1- and γ_2-genes, respectively. However, intron 2 donor and acceptor splice sequences show few substitutions, and the observed substitutions retain nucleotide identity with the consensus sequences, e.g. the A or G nucleotides in positions 2 and 6 of the donor sequence (Fig. 2). The intron 2 acceptor sequences are invariant in and near the splice junction; their only differences, of which many are transitions, are found in the pyrimidine-rich region which does not alter the nucleotide composition of the intron acceptor sequences. These comparative observations are consistent with the intron splice sequence requirements found by van Santen and Spritz (1985), in which nucleotide sequences adjacent to the donor (3' 6 bp) and acceptor (5' 20 bp) were found to be important components of the splicing mechanism and thus are strictly conserved. Despite the nearly 20% divergence among the primate γ-gene intron 2 sequences presented in Fig. 2, the mechanism for splicing out these introns probably remains unaltered, following the excision order described by Lang et al. (1985). Mutations in the essential splice donor and/or acceptor nucleotides or mutations which create new internal donor or acceptor splice sites can have devastating consequences, such as that resulting in β-thalassemia (Orkin and Kazazian, 1984; Dobkin and Bank, 1985).

Extensive Gene Conversion between Rhesus γ-Globin Genes—Fig. 3 presents the conversion regions for the duplicated catarrhine γ-genes which are suggested by the position-by-position parsimony analysis over the 1994 aligned nucleotide positions of the 14 primate γ-genes shown in Fig. 2. A “dot” at a variable site indicates that the positional paralogues of γ_1- and γ_2-genes of the same species are closer genealogically to each other than to positional orthologues, therefore supporting an hypothesized conversion. In contrast, an open “square” indicates the reverse and supports the absence of conversion. Eight rows of these intraspecies γ_1- to γ_2-gene comparisons are presented; two of the rows are for the human genes (consisting of sequenced γ_1- or γ_2-gene from chromosome A, γ_2- or γ_2-gene also from chromosome A, and the γ_2-gene from a second chromosome, B). Four rows are for the paired γ_1- and γ_2-genes of chimpanzee, gorilla, orangutan, and rhesus monkey. The seventh row is for the deduced ancestral hominine (human, chimpanzee, and gorilla; HCG in Fig. 3) paired genes, and the eighth row is for the deduced ancestral hominid (hominine and orangutan; HCGO in Fig. 3) paired genes. Each of the eight rows shows a highly nonrandom distribution of dots and squares which identify the same three distinct regions suggested by our earlier analysis (Sligh- tom et al., 1987). The distribution of dots indicates that the γ-gene 5’-regions (the approximately 650 bp extending from the 5’-flanking region through exon 1, intron 1, and exon 2; positions -153 to 495 in Fig. 2) are prone to gene conversions, see region 1, Fig. 3. In contrast, the 3’-regions (3’-untranslated and flanking DNAs, labeled region 3 in Fig. 3; positions 1582 to 1839 in Fig. 2) are not prone to conversions. Between these 5’- and 3’-regions is located region 2 (consisting of intron 2 and exon 3, approximately 1000 bp; positions 496 to
The distribution of dots can be accounted for by the minimun number of hypothesized gene conversions, if we consider two or more consecutive dots as part of the same conversion region. Each hypothesized gene conversion region is allowed to spread until it meets the first of two consecutive squares in rows representing paired hominid genes but essentially only dots in the row for the paired rhesus y-genes.

The aligned y'- and y2-genes (5'- and 3'- genes, respectively) of human (H), chimpanzee (C), gorilla (G), orangutan (O), and rhesus (R), plus the hypothetically derived y-gene ancestors of human, chimpanzee, and gorilla (HCG) and of HCG and orangutan (HCGO) are used to locate gene regions that may have been involved in conversion events. The converted regions are suggested above the position by position parsimony analysis results, the series of open boxes and dots. These regions were derived using the single y-gene sequence alignment of spider monkey, brown lemur, and dwarf lemur (as shown in Fig. 2) to identify where intraspecific similarity of the duplicated y1- and y2-genes (paralogous positions) is greater than the interspecific similarity of y' or y2 (orthologous positions). The open boxes indicate positions where interspecific similarity between y' (or y2)-genes is greater than intraspecific similarity between y1- and y2-genes. Two consecutive dots indicate greater intraspecific similarity and suggest a conversion event (Slightom et al., 1987), labeled C9 to C14. These suggested conversion regions are shown above the lines containing position by position analysis, represented as boxes or dots. Below the map of y' and y2-genes are indicated the three y-gene regions which show different evolutionary histories, and using regional parsimony analysis the overall phylogenetic order for these regions has been constructed. The complex nature of region 2 necessitated breaking up sequences into small subregions that appear to reflect patchy conversion regions (see discussion in Slightom et al., 1987).

In the 5'-direction rhesus conversion C14 extends beyond the y-gene sequences aligned in Fig. 2. Alignment of the presently available 200 bp of additional 5'-flanking DNA sequence from the two rhesus y-genes (Fig. 4) shows a continuation of sequence identity with the first difference being located at position -157 and a second difference being located 56 bp farther upstream, at position -214. This extensive nucleotide sequence identity between the rhesus y1- and y2-
5'-flanking DNAs extends up to position -282, after which three closely spaced sequence differences are found (Fig. 4). The original γ-gene duplication encompasses about 5 kb (Shen et al., 1981), and its 5'-boundary is located about 1200 bp 5' farther than the sequences shown in Fig. 4. The degree of divergence between the 5'-most 100 bp of these rhesus γ-genes presented in Fig. 4 is only about 4%, much below the expected level of 14 to 15% for noncovered γ-gene regions (Shen et al., 1981). Thus, position -281 is unlikely to be the termination point for all conversion events which extended into these 5'-flanking regions, but it may represent the 5'-termination point for the most recent conversion. If we assume that the 5'- and 3'-boundaries of converted sequence in the rhesus γ-genes are at positions -281 (Fig. 4) and position 1696 (Fig. 2), conversion 14 involved at least a total of 1898 bp.

Is C14 the Result of a Single or Multiple Conversion Events?—If the converted region of the rhesus γ-genes labeled C14 occurred as a single conversion event, which included the (TG)₁ elements, no significant variation in the degree of divergence over this region would be expected. However, if the divergence pattern within C14 shows disparity among regional comparisons, C14 may actually be the result of more than one conversion event. Uniformity of divergence within C14 can be tested by analyzing divergence within specific gene regions; we selected genic regions with natural boundaries (exons, introns, untranslated, and flanking) and subdivided intron 2 by using the 5'- and 3'-ends of the (TG), element as arbitrary boundaries. Both γ-genes are active; thus, coding sequences were examined separately from noncoding because exons are subject to selective pressures. The divergence values are as follows: 5'-flanking (3/281 = 1.0%); 5'-untranslated + exon and intron 1 + exon 2 (0/491 = 0%; these regions were grouped because they show no divergence); 5'-end of intron 2 to 5'-end of (TG), element (4/578 = 0.7%); 3'-end of (TG), element to 3'-end of intron 2 (3/248 = 1.2%); exon 3 (1/136 = 0.7%); and 3'-untranslated region (3/86 = 3.5%). This analysis of divergence clearly shows disparity among the various regions of C14; in intron 2 the degree of divergence 3' and 5' of the (TG), element differs by almost a factor of 2, which supports the possibility that the conversion region designated C14 may be the result of more than one conversion event. The most recently converted part of C14 is located 5' of the (TG), elements, a possibility supported by the fact that the two rhesus (TG), elements are identical in their 5'-regions but differ in length (4 bp) and composition in their 3'-regions (Fig. 2). Two possible histories of conversion events could account for the disparity in the divergence of the regions of C14. Frame A of Fig. 5 shows a history of conversion events in which two conversions were initiated at the 5'- and 3'-boundaries of the (TG), element and they extend in their respective 5'- and 3'-directions. To account for the disparity of divergence found, the 3'-conversion (labeled C14a, in frame A) would predate the 5'-conversion (labeled C14b, in frame A). Frame B of Fig. 5 shows another history of conversion.
of identity. These rhesus (TG), elements are considerably smaller than their homologous counterparts in the orangutan and gorilla species. The (TG), elements are flanked by short repeats. After this proto-(TG), element appeared, its development probably continued as a result of mispairings during exchanges between homologous in the single γ-locus of the early simians. After γ-gene duplication in the stem of catarrhines (Barrie et al., 1981; Shen et al., 1981), the developed (TG), element could also initiate exchanges between paralogous γ1- and γ2-loci, thus reducing divergence between these genes and increasing their chance of evolving in concert.

The (TG), elements present in the rhesus γ-genes are among the largest found thus far, with n = 24 and m = 22 for γ1- and γ2-genes, respectively (Fig. 2). Moreover, the DNA sequences flanking these (TG), elements share a high degree of identity. These rhesus (TG), elements are considerably larger than those found in the orangutan γ-genes, which have n values of 11 and 13 (Slightom et al., 1987). Our analysis of the orangutan γ-genes suggests that this pair of genes has experienced a low number of gene conversions, none of which appear to have been initiated by the (TG), elements. This low level of conversions within the orangutan γ-genes correlates with its short (TG), elements and lack of shared identity between nucleotide sequences adjacent to the (TG), elements. In contrast, the γ-genes of rhesus appear to have experienced a recent and extensive gene conversion. The long length of the rhesus (TG), elements and high degree of identity shared between sequences adjacent to the (TG), elements (especially 5' of the element) support our hypothesis that these elements play an important role in some of the γ-gene conversions (Slightom et al., 1987). This hypothesis is also supported by analysis of the role that (TG), elements play in recombinations between SV40 molecules (Stringer, 1985) and the reduced level of conversion in Ascobolus immenus as a result of mispairing between exchanging DNA strands (Nicola and Rossignol, 1983). The occurrence of a recent and extensive conversion between the rhesus γ-genes obscures any evidence of more ancient conversions in the ancestry of these genes; thus, it is impossible to determine if the ancestral gene pairs were frequently involved in conversion events. However, the fact that the rhesus γ-genes not only contain the elements favoring the initiation and extension of conversion events but also have been involved in a recent conversion suggests that these genes may frequently have engaged in gene conversions. Thus, an important extension to our hypothesis is that if the (TG), element is long and its flanking DNA matches, then frequent conversions can be expected. In other words, the lack of divergence between a gene pair indicates their involvement in frequent conversion events, which is consistent with the gene conversion model proposed by Walsh (1987).

Evolution and Polarity of γ-Gene Conversion—With the addition of the rhesus sequence to the γ-gene alignment (Fig. 2), our parsimony analyses continue to support the previously identified intraspecific γ-gene conversion regions for the four hominid species: human conversions C2, C3, and C4; chimpanzee conversions C5, C6, and C7; gorilla conversions C8, C9, and C10; and orangutan conversions C11, C12, and C13 (Slighom et al., 1987). Also, inclusion of the paired rhesus γ-genes in the parsimony analyses supports the somewhat patchy ancestral hominoid conversion CI and provides evidence for an ancestral hominoid conversion C0, which continues over the entire (TG), segment starting in the 5'-flanking DNA region and terminating in the 3'-untranslated region, Fig. 3). As can be seen in the evolutionary trees constructed by the parsimony method for regions 1, 2, and 3 of the aligned γ-gene sequences, only region 3 (3'-untranslated and flanking DNA) shows complete absence of gene conversions.

Although identification of the converted region of the rhesus γ-gene pair was relatively straightforward, assignment of the γ-gene donor and acceptor sequences within the subregions is difficult. Not only does C14 span the complete transcriptional region of the paired rhesus genes, but so does the ancestral conversion C0 (the gene conversion hypothesized to have occurred in the hominoid stem), except for one location. A three-base position gap occurs at nucleotide positions 505 to 506 in the γ1-gene of rhesus monkey, orangutan, and gorilla and in the γ1-gene of rhesus monkey, but not in any of the other γ-genes (see Fig. 3). If this location is excluded from the hypothesis conversion C0 in the hominoid stem (i.e. C0 is considered to be a patchy conversion), this gap can be explained as resulting from a three-base deletion in the nascent γ1-locus which arose in the catarrhine stem after the γ-gene duplication. Much later in the lineage to rhesus monkey, γ2-donor sequence containing this deletion replaced, by a conversion event, the homologous γ1-acceptor sequence.
three-base deletion region indicates that for the rhesus γ-genes the region of C14 which is 5' of the (TG), element γ2-gene sequence was superimposed onto the γ1-gene. The 5'-region conversions in human (C3 and C4) and chimpanzee (C7) genes apparently involved γ-sequences superimposed onto γ2-genes (Slighotm et al., 1987), which would account for the absence of the 503- to 506-position deletion in human and chimpanzee γ-genes. The 5'-conversion regions in gorilla (C6) and orangutan (C11) genes end before reaching position 503; therefore, the deletion at positions 503 to 506 in the γ2-lineage of the gorilla and orangutan γ2-genes would be expected to remain. This postulated occurrence of a patchy-type conversion pattern in the hominid lineages has allowed us to deduce that in these lineages γ1-donor sequences were superimposed onto γ2-acceptor sequences in most conversions; only chimpanzee conversion C5 of the hominid intraspecific conversions shows the opposite polarity (Slighotm et al., 1985). From this lengthy argument we conclude that the three-base deletion at positions 503 to 506 does provide the only evidence for the polarity of at least part of C14, suggesting that like C5 in chimpanzee the rhesus γ2-donor sequence was superimposed onto the γ1-acceptor gene sequence.

Concluding Remarks—The sequence of the paired γ-globin genes from rhesus has provided us with the most striking example so far of γ-gene conversion. Conversion C14 is even more extensive than conversion C4, which occurred between the human γ-gene pair on chromosome A and involved about 1500-bp (Slighotm et al., 1980; Shen et al., 1981), because it extended across the complete length of the rhesus γ-gene transcript and into the 5'-flanking DNAs. Whether such extensive gene conversions are common to γ-gene pairs of other rhesus individudals or other Cercopithecoid species remains to be determined. Further investigation of γ-genes from Cercopithecoid lineages could be quite rewarding in view of the fact that this superfamily contains many more living species than in Hominoidae. Thus, it should be possible to determine if extensive species-specific conversions are common in cercopithecoids or if their conversions patterns will be similar to the patchy-type found in hominoids.

Walsh (1987) has recently presented a mathematical gene conversion model in which conversion events are considered to be frequent if members of a gene pair are found to be identical. Identity of the gene pair members is maintained by frequent conversions until mutations occur (base mutations, insertions, or deletions). As the members of a gene pair accumulate such mutations conversion frequencies are reduced until the individual gene members “escape” from the conversion process responsible for their concerted evolution (Walsh, 1987). The γ-gene pairs of the catarrhine primates studied so far appear to be in different stages of the gene conversion process, with some gene pairs accumulating mutations in such important elements as the (TG), and its flanking sequences and thus may be escaping from the conversion process. The human and rhesus γ-gene pairs each appear to be continually involved in conversions and thus are not accumulating the mutations necessary to escape the conversion process, whereas the orangutan and gorilla γ-gene pairs each appear to be escaping the conversion process. Evolutionary divergence between the two nonallel γ-genes is subject to selective pressures because at least one γ-globin gene must remain functional. The other γ-gene member must also remain functional if expressed, but if silenced it could become another δ-type globin pseudogene. We previously suggested that the 5' γ-gene, which in hominoids is expressed at three-fold higher levels than the 3' γ-gene, is accordingly subject to more stringent selective pressure (Slighotm et al., 1987). If the γ-gene conversion process were to become completely blocked in a species such as gorilla or orangutan, the γ2-gene in that species would have the likely fate of degenerating into a pseudogene. Alternatively, but a much rarer fate would be that the γ2-gene would acquire a new advantageous function for the species.

Acknowledgments—We are grateful to Dr. Sandy Martin for the gift of rhesus DNA and for isolating clone MmuCh30-13.1. We thank Lynn Whyle for help in type setting the sequences shown in Fig. 2.

REFERENCES

Barrie, P. A., Jeffreys, A. J., and Scott, A. F. (1981) J. Mol. Biol. 149, 319-336
Bentley, D. L., and Rabbitts, T. H. (1983) Cell 32, 181-189
Benton, W. D., and Davis, R. W. (1977) Science 196, 180-182
Birnboim, H. C., and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523
Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978) Proc. Natl. Acad. Sci. U. S. A. 75, 4853-4857
Chang, L.-Y. E., and Slighotm, J. L. (1984) J. Mol. Biol. 180, 767-784
Collins, F. S., and Weissman, S. M. (1984) Prog. Nucleic Acid Res. Mol. Biol. 31, 315-435
Degnan, M., and Ehrlich, S. D. (1979) Gene (Amst.) 6, 23-28
Dobkin, C., and Bank, A. (1985) J. Biol. Chem. 260, 16332-16337
Drahtsides, A., Posakowy, J. W., Maniatis, T., Lawn, R. M., O'Connor, N., C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slighotm, J. L., Blechl, A. E., Smithies, O., Baralle, E. F., Shoulders, C. C., and Proudfoot, N. J. (1980) Cell 21, 653-661
Flanagan, J. G., Lefranc, M.-P., and Rabbitts, T. H. (1984) Cell 36, 651-668
Giebel, L. B., van Santen, V. L., Slighotm, J. L., and Spritz, R. A. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 6980-6989
Gill, A., and Proudfoot, N. J. (1987) Cell 48, 399-406
Goodman, M., Koop, B. F., Czelusniak, J., Weiss, M. L., and Slighotm, J. L. (1984) J. Mol. Biol. 180, 865-893
Hardison, R. C. (1981) J. Biol. Chem. 256, 11780-11786
Harris, S., Thackeray, J. R., Jeffreys, A. J., and Weiss, M. L. (1986) Mol. Biol. Evol. 3, 465-484
Hayashida, H., Miyata, T., Yamawaki-Kataoka, Y., Honjo, T., Wels, J., and Blattner, F. (1984) EMBO J. 3, 2047-2053
Hill, A. V. S., Bowden, D. K., Weatherall, D. J., and Clegg, J. B. (1986) Blood 67, 1611-1618
Hohn, B. (1979) Methods Enzymol. 68, 298-308
Kilpatrick, M. W., Rhyuk, J., Singleton, C. R., Zarling, D. A., Jovin, T. M., Hanish, L. H., Emlinger, B. F., and Wells, R. D. (1984) J. Biol. Chem. 259, 7268-7274
Koop, B. F., Goodman, M., Xu, P., Chan, K., and Slighotm, J. L. (1969a) Nature 219, 234-238
Koop, B. F., Miyamoto, M. M., Embury, J. E., Goodman, M., Czelusniak, J., and Slighotm, J. L. (1969b) J. Mol. Evol. 24, 94-102
Lang, K. M., van Santen, V. L., and Spritz, R. A. (1985) EMBO J. 4, 1911-1916
Loomes, W. A. M., and Blattner, F. R. (1983) Gene (Amst.) 26, 171-179
Maeda, N., and Smithies, O. (1986) Annu. Rev. Genet. 20, 81-108
Maeda, N., Yang, F., Barnett, D. R., Bowman, B. H., and Smithies, O. (1984) Nature 309, 131-135
Mahoney, W. C., and Nute, P. E. (1980) Biochemistry 19, 4436-4442
Martin, S. L., Zivver, E. A., Kian, Y. W., and Wilson, A. C. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 3563-3566
Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499-560
Michelson, A. M., and Orkin, S. H. (1983) J. Biol. Chem. 258, 15245-15254
Miyamoto, M. M., Slighotm, J. L., and Goodman, M. (1987) Science 238, 369-373
Mount, S. M. (1982) Nucleic Acids Res. 10, 459-472
Nicolas, A., and Rossignol, J.-L. (1983) EMBO J. 2, 2265-2270
Orkin, S. H., and Kazarzyn, H. H., Jr. (1984) Annu. Rev. Genet. 18, 131-171
Pays, E., Van Assel, S., Laurent, M., Davirwe, M., Vervoort, T., Van Meirenen, N., and Steinert, M. (1983) Cell 34, 371-381
Proudfoot, N. J., and Browne, G. G. (1976) Nature 263, 211-214
Rimm, D. L., Horness, D., Kureza, J., and Blattner, F. R. (1980) Gene (Amst.) 12, 301-309
Schulze, D. H., Pease, L. R., Géer, S. S., Reyes, A. A., Sarmiento, L. 12437
Nucleotide Sequence of Rhesus Fetal Globin Genes
Nucleotide Sequence of Rhesus Fetal Globin Genes

A., Wallace, R. B., and Nathenson, S. G. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 2007–2011
Scott, A. F., Health, P., Trusko, S., Boyer, S. H., Prass, W., Goodman, M., Czelusniak, J., Chang, L.-Y. E., and Slightom, J. L. (1984) Mol. Biol. Evol. 1, 371–389
Shen, S.-H. S., Slightom, J. L., and Smithies, O. (1981) Cell 26, 191–203
Slightom, J. L., Blechl, A. E., and Smithies, O. (1980) Cell 21, 627–638
Slightom, J. L., Chang, L.-Y. E., Koop, B. F., and Goodman, M. (1985) Mol. Biol. Evol. 2, 370–389
Slightom, J. L., Theisen, T. W., Koop, B. F., and Goodman, M. (1987) J. Biol. Chem. 262, 7472–7483
Stringer, J. R. (1985) Mol. Cell. Biol. 5, 1247–1259
van Santen, V. L., and Spritz, R. A. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 2885–2889
Walsh, J. B. (1987) Genetics 117, 543–557
Weiss, E., Golden, L., Zakut, R., Mellor, A., Fahrner, K., Kvist, S., and Flavell, R. A. (1983) EMBO J. 2, 453–462
Wilson, J. T., Wilson, L. B., deRiel, J. K., Villa-Komaroff, L., Efstratiadis, A., Forget, B. G., and Weissman, S. M. (1978) Nucleic Acids Res. 5, 563–581
Zimmer, E. A. (1981) Evolution of Primate Globin Genes. Ph.D. thesis, University of California, Berkeley
Zimmer, E. A., Martin, S. L., Beverley, S. M., Kan, Y. W., and Wilson, A. C. (1990) Proc. Natl. Acad. Sci. U. S. A. 77, 2168–2172