Complex Events in the Evolution of the Haptoglobin Gene Cluster in Primates*

Susan M. McEvoy and Nobuyo Maeda‡
From the Department of Genetics, University of Wisconsin, Madison, Wisconsin 53706

(Received for publication, April 8, 1988)

Southern blot analyses of genomic DNA show that new world monkeys have only one haptoglobin gene but that chimpanzees, gorillas, orangutans, and old world monkeys have three. Humans have two: haptoglobin (Hp) and haptoglobin-related (Hpr). These observations suggest that a triplication of the haptoglobin locus occurred after the divergence of the new world monkeys, followed by a deletion of one locus in humans. To investigate these events, we have cloned the haptoglobin gene cluster in chimpanzee. The organization of the Hp and Hpr genes in chimpanzees is the same as in humans, including a retrovirus-like sequence in the first intron of Hpr. The third gene, which we name Hpp for haptoglobin primate, is 16 kilobases downstream of Hpr. A second copy of the retrovirus-like sequence occurs between Hpr and Hpp. The nucleotide sequence of the chimpanzee Hpp gene suggests that it may code for a functional protein, but the chimpanzee Hpr gene has a single base deletion in exon 5 that causes a frameshift. Comparison of the human and chimpanzee sequences suggests that the human Hpr gene was generated by a homologous unequal crossover between ancestral Hpr and Hpp genes. The crossover point lies within a 1.3-kilobase region containing exon 5 and 500 nucleotides 3' to the genes, but the exact point is obscured by a subsequent gene conversion event.

Haptoglobin is a hemoglobin-binding protein found in the blood plasma of vertebrates. In 1955, investigations by Smithies and Walker revealed the existence of inherited variations of the haptoglobin proteins in normal humans. Subsequent studies showed that most of the variations are due to differences in the α chains of the four-chain haptoglobin protomer (αβ)₄. There are three common alleles in humans, designated Hp¹, Hp¹₂, and Hp³.¹

In 1962, Smithies et al. proposed that the Hp² allele was formed as the result of an intragenic duplication originating from a nonhomologous unequal crossover between two Hp¹ alleles. Maeda et al. (1984) confirmed this proposal at the DNA level and showed that the nonhomologous crossover was within the fourth intron of an Hp¹ allele and the second intron of an Hp¹ allele. The resulting Hp² gene has seven exons rather than the five exons present in each of the original genes, Hp¹ and Hp¹².

The same study revealed that the entire Hp gene is duplicated in the human genome; there is a haptoglobin-related gene, Hpr, 2.2 kb downstream of the haptoglobin gene, Hp. The nucleotide sequence of the Hpr gene (Maeda, 1985) suggests that it is able to code for a structurally normal protein. However, although in vivo transfection experiments indicate that the Hpr promoter is active and cell-specific (Oliviero et al., 1987), no hemoglobin-binding protein of the expected structure has been detected (Bensi et al., 1985).

Sequence comparison of the Hpr gene with the Hp¹ gene shows an overall difference of 6.4%, counting each length difference and base substitution as a single difference. The first introns of Hp and Hpr differ most (by 8.9%). Using this datum from the region of greatest difference, one of us (Maeda, 1985) estimated that the duplication of the ancestral Hp gene occurred about 30 million years ago. The nucleotide sequences of the human gene pair, Hp and Hpr, were consistent with the hypothesis that the duplication was the result of a nonhomologous chromosomal breakage and reunion event. However, the region between the Hp and Hpr genes contains 600 bp of DNA not found anywhere else in the cloned 33-kb region. The 3'-half of this 600 bp of DNA is an Alu family repeat. We were, therefore, cautioned that the true history of the initial duplication event might have been masked by later events.

To better understand the evolutionary history of the haptoglobin gene family and the details of the various molecular events involved, the present characterization of the haptoglobin genes in primates was undertaken. We have analyzed by Southern blot hybridization (Southern, 1975) genomic DNA from the great apes: chimpanzee (Pan troglodytes), gorilla (Gorilla gorilla), and orangutan (Pongo pygmaeus); the old world monkey: rhesus monkey (Macaca mulatta); and the new world monkeys: cebus (Cebus albifrons) and spider monkey (Ateles geoffroyi). We have also cloned the entire haptoglobin gene cluster of chimpanzees and have characterized it by mapping with restriction enzymes, by hybridization to various probes, and by partial nucleotide sequencing. Here we present evidence that a triplication of the haptoglobin locus occurred after the divergence of the new world monkeys from other primates and that a deletion of one of the loci occurred in humans after their divergence from the apes. Thus, the history of the haptoglobin gene cluster is inherently complex as a result of multiple recombinational events and insertions of two retroviral sequences and several Alu repeats.
EXPERIMENTAL PROCEDURES

DNA—Genomic DNA samples from a gorilla and a chimpanzee were provided by Dr. E. Zimmer (Stanford University). DNA from an orangutan was given by Dr. A. Wilson (University of California, Berkeley), and DNA samples from a rhesus monkey (isolated from kidney) and a spider monkey (from fibroblast culture) were gifts from Dr. K. Chen and Dr. R. Spritz, respectively, of the University of Wisconsin, Madison. DNA isolated from blood samples of 10 unrelated chimpanzees and 6 unrelated orangutans were given by Dr. J. Murray at the University of Iowa.

Probes—The hp2α probe and the hpβ probe are derived from a cDNA clone and, respectively, cover the regions coding for the hp2α and hpβ polypeptides and their leader amino acids. The LTR probe includes the 5′-long terminal repeat sequence of the retrovirus-like element RTVL-I (Maeda et al., 1985) that occurs in the first intron of Hpr. The LTR probe was isolated as an 800-bp BstEI fragment. The RTVL probe is a 400-bp BamH1/XbaI fragment subcloned from the middle of the retrovirus-like sequence. Total genomic DNA isolated from human fibroblast cells (563, a gift from Dr. R. DeMars, University of Wisconsin) was nick-translated and used to probe for the Alu family of repeated sequences. Blotting and hybridization were as described by Southern (1975) and Jeffreys and Flavell (1977).

Cloning—Complete EcoR1, BglII, and BamH1 digests of chimpanzee genomic DNA were ligated into Charon 32 or Charon 35 vector arms (Loenen and Blattner, 1983) and packaged in vitro using the technique of Horn (1979). The phage libraries were screened without amplification with the hp2α probe, the promoter probe, and the RTVL probe.

DNA Sequencing—The chimpanzee haptoglobin genes were partially sequenced by the methods of Maxam and Gilbert (1977) and of Sanger et al. (1977). Sequences were analyzed using software provided by the University of Wisconsin Genetics Computer Group (Devereux et al., 1984).

RESULTS AND DISCUSSION

Southern Blot Analysis of Primate Haptoglobin Genes—We have characterized the haptoglobin gene complex of six primate species by Southern blotting genomic DNA digested with EcoR1, BamH1, BglII, HindIII, XbaI, and SacI. Each blot was hybridized to one or more of the hp2α, hpβ, promoter, LTR, and RTVL probes. Some of the results are illustrated in Fig. 1. Fig. 1A compares human, gorilla, chimpanzee, and orangutan DNAs after digestion with BamH1 and hybridization to the promoter probe. The samples labeled H2 and H1 are from two human individuals who are homozygous for the haploglobin alleles Hp2 and Hp1, respectively. The DNA from the Hp2/Hp2 individual shows 9.2- and 5.0-kb bands, and the DNA from the Hp1/Hp1 individual shows 7.5- and 5.0-kb bands as described in previous studies (Maeda et al., 1986). The chimpanzee, orangutan, and gorilla genomic DNAs each have three BamH1 bands, 12.4, 7.5, and 4.4 kb, hybridizing to the promoter probe. The presence of a 7.5-kb BamH1 fragment in DNA from these primates and in the DNA from Hp2/Hp1 homozygous humans suggests that the primate Hp gene complex includes a region comparable to Hp1.

Fig. 1B shows EcoR1 digests of chimpanzee, gorilla, and rhesus monkey DNA after hybridization to the hp2α probe. In this blot, three bands are again detected in each of the primate DNAs. The gorilla and chimpanzee DNAs are most alike; they share two bands which are 10.8 and 9.8 kb in length. The third bands in the gorilla and chimpanzee DNAs differ, being 2.7 and 7.8 kb, respectively. The rhesus monkey DNA has a very different banding pattern with 9.5-, 7.2-, and 6.9-kb bands.

A Southern blot of rhesus monkey and orangutan DNA digested with BglII (Fig. 1C) shows that in both species three bands again hybridize to the hp2α probe, although they differ in size: 16, 11.8, and 7.7 kb in rhesus monkey, and 18, 11.5, and 10.2 kb in orangutan.

Thus, in general, the blots of genomic DNA from chimpanzees, gorillas, orangutans, and rhesus monkeys show three distinctive bands hybridizing to the haptoglobin probes. (We have observed some exceptions to this three-band generalization with particular digests, but all appear to be explicable in terms of overlapping, more intense bands.) We have analyzed DNA from 8 unrelated orangutans, 11 unrelated chimpanzees, and 2 unrelated rhesus individuals. No variations in the hybridization patterns within any of these species were detected (the blots were hybridized to the hp2α probe). This observation allows us to exclude the possibility that the three bands seen in these primates are due to allelic differences, such as those seen in humans. Our overall conclusion from these Southern blotting experiments is that there are three genes in the haptoglobin complex of the chimpanzee, gorilla, orangutan, and rhesus monkey, in contrast to the two in humans.

Our data are, at first sight, in conflict with the results of Lavareda de Souza et al. (1986), who reported the finding of
only two genes in the haptoglobin gene cluster of chimpanzees but three in baboons. Their conclusion stems from observing two incompletely resolved bands of 3.4 and 3.2 kb in the blot of a HindIII digest of chimpanzee genomic DNA hybridized to a haptoglobin probe. We, however, observed three completely resolved bands in HindIII digests of chimpanzee DNA that are 3.5, 3.4, and 3.2 kb in length.

Fig. 1D shows cebus and spider monkey genomic DNAs digested with EcoRI and hybridized to the hp2o and the promoter probes. In both species only one band hybridizes, a 10.5-kb band in the spider monkey and a 9.0-kb band in the cebus monkey. Blots of the same DNAs digested with other enzymes (BglII, BamHI, HindIII, and XbaI) also show single bands. We conclude that cebus and spider monkey have only one haptoglobin gene.

The presence of a single haptoglobin gene in cebus and spider monkeys strongly suggests that the locus became more complex in old world monkeys and apes subsequent to their divergence from the new world monkeys. This suggestion is supported by our earlier deduction from sequence comparisons that the approximate time of the gene duplication which made the Hp and Hpr gene pair is 30 million years ago (Maeda, 1985). Preliminary Southern blot hybridization experiments (data not shown) suggest that mice and pigs also have a single haptoglobin gene in their genomes. Thus the simplest explanation for the presence of three loci in the old world monkey, orangutan, gorilla, and chimpanzee genomes is that a triplication of the haptoglobin gene occurred in a common ancestor of these species at some time after their divergence from the new world monkeys. The presence of only two loci in humans then implies that one of the loci was deleted in humans at some time after their divergence from the other primates.

The Organization of the Chimpanzee Hp Genes—In order to define more clearly the event which left humans with two haptoglobin genes, we have cloned and characterized the haptoglobin genes in chimpanzees, the species most recently diverged from humans (Sibley and Ahlquist, 1984; Miyamoto et al., 1987; Maeda et al., 1988). Three λ phage libraries were screened with the hp2o, hpβ, promoter, LTR, and RTVL probes to obtain, overlapping clones of a 48-kb region that contains the three chimpanzee haptoglobin genes. Inserts of these clones were mapped with restriction enzymes and hybridized to the above probes and to nick-translated total human DNA. From these results we determined the structural organization of the chimpanzee haptoglobin gene cluster as shown in Fig. 2. The clones, in the lower part of Fig. 2, are labeled with the letters E, B, or Bg, referring to the EcoRI, BamHI, or BglII libraries from which they were isolated, and with numbers according to their sizes in kilobases. The relative positions of exons, retrovirus-like sequences, and Alu family repeats were determined by the restriction mapping, by hybridization to the probes, and by partial nucleotide sequencing. We expected to find a 4.0 kb fragment (represented by a dashed line in the figure) but failed to find it despite several attempts. This fragment would contain a substantial portion of the retrovirus-like sequence that occurs in this position in both the human and chimpanzee Hp genes. Past experience in our lab and in other laboratories shows that the cloning of retrovirus-like sequences can be a very difficult task (Mager and Henthorn, 1984). Because this clone is missing from our collection, we have no overlapping fragment to link the 9.8E and 5.8E clones. However, because of the very great similarity of this chimpanzee region to its human counterpart, it is virtually certain that the 9.8E and 5.8E clones occur end to end.

The three genes in the chimpanzee haptoglobin cluster all occur in the 48-kb cloned region. The first two genes have the same organization as the human Hp and Hpr genes as judged by our Southern blot analyses and by the partial nucleotide sequences presented below (see Fig. 2, upper part), and we have named them accordingly. The chimpanzee Hpr gene is situated approximately 2.5 kb downstream of the chimpanzee Hp gene and has a retrovirus-like insertion in its first intron, as does the human Hpr gene. In addition, the chimpanzee Hp and Hpr genes both have Alu family repeat sequences in the same positions as do the human genes, just 3′ to the promoter region in both genes, just 5′ to the first exon of Hpr, and in the fourth intron of Hpr.

We name the third gene in chimpanzees Hpp, for haptoglobin-primate. This gene is located about 16 kb downstream of the last exon of Hpr. Hpp is structurally very similar to the other two genes. It contains five exons. Two Alu repeats are present in Hpp at positions equivalent to those in Hpr, one in the first intron and the other in the fourth intron. The first of these Alu repeats is also present in Hp, but the second is not. There is an additional Alu sequence present about 3 kb downstream from the end of the Hpp gene. This Alu sequence is also present 3 kb downstream of the human Hpr gene. It is not present downstream of the chimpanzee Hpr gene.

The tandem arrangement of the three genes in the chimpanzee haptoglobin gene cluster strongly suggests that the single gene in the ancestral chromosome was duplicated initially and that a triplication followed. The initial recombination was most likely an unpredictable and rare event, but once the gene became duplicated, a homologous but unequal crossover between the duplicated region could occur more readily (Maeda and Smithies, 1986).

Insertion of the Alu Repeats—Because all three genes have Alu repeats in their first intron, the original single haptoglobin gene must have had this Alu repeat before the gene was subsequently duplicated. The Alu repeat in the fourth intron, on the other hand, does not occur in any Hp genes, although this repeat occurs at exactly the same positions in the chimpanzee Hpr and Hpp genes. These data thus indicate that the 3′ member of a pair of ancestral duplicated haptoglobin genes acquired an Alu repeat in the fourth intron prior to the triplication event. We also infer that a homologous unequal crossover between the first and second of the ancestral duplicated gene must have taken place somewhere 3′ to this Alu repeat.

The lengths of the Alu repeats in the fourth intron of the chimpanzee Hpr and Hpp genes are different although they occur at exactly the same position in the two genes. The nucleotide sequences (see Fig. 3) show that the Alu repeat in the chimpanzee Hpr gene is only about half the length of the usual primate Alu repeats, as is the equivalent repeat in the human Hpr gene. The Alu repeat in the chimpanzee Hpp gene, on the other hand, has the normal (dimeric) structure shared by most primate Alu sequences. The positions of these Alu repeats and their nucleotide sequences thus confirm the hypothesis proposed by one of us (Maeda, 1985) that the short Alu sequence in the human Hpr gene was formed by deletion of a portion of an ancestral longer dimeric form. The present data establish that the deletion of the internal part of the Alu sequence in Hpr must have occurred after the triplication of the haptoglobin locus, but at some time before the divergence of chimpanzees and humans.

A Second Retrovirus-like Sequence Occurs in the Chimpanzee Gene Cluster—We were surprised to find a second retrovirus-like element in the intergenic region between the chimpanzee Hpr and Hpp genes. This insertion was detected
because the 10.8E clone hybridized strongly to the LTR probe, which contains the 5'-long terminal repeat sequence from the RTVL-I insertion in human Hpr. A 1-kb EcoRI fragment mapping to the middle of the 12.4B clone also hybridized to the LTR probe. The hybridization of this fragment was weak, but it was detectable under normally hybridization conditions. The distance between these two hybridizing regions, about 8 kb, is only 1 kb shorter than the length of the retrovirus-like insertion in the Hpr gene, which indicates that another retrovirus-like sequence related to RTVL-I is present in this region. We have confirmed this by sequencing a small portion of the region just downstream from the DNA fragment that hybridizes the LTR probe. The sequenced region shows strong homology to the human RTVL-I sequence (data not shown).

It is unusual for two related retrovirus-like elements to be found in one short region of the genome. This is even more remarkable when one considers that genomic Southern blots hybridized to the RTVL-probe indicated that there are no more than 15 copies of the RTVL-I elements in either the human or the chimpanzee genome.2 We have no evidence at this time that bears on the event(s) which led to either insertion, or on what their effects may be on the expression of the genes in this region.

Nucleotide Sequence Analyses of the Chimpanzee Genes—
The nucleotide sequence of the chimpanzee Hpp gene was determined except for about 1 kb in the first intron. (98% of the sequence was determined on both strands.) It shows that there are no frameshift or nonsense mutations in the coding regions and that the exon-intron splicing junctions have no defects. Thus the chimpanzee Hpp gene appears to be structurally capable of producing a functional protein, as is the human Hpr gene (Maeda, 1985; Bensi et al., 1985).

The nucleotide sequence was determined for a region of the chimpanzee Hpr gene extending from about 150 bp 5' to its second exon to about 1.0 kb into the 3' flanking region. In the coding portion of the sequenced region we found a single base deletion (position 16809 in Fig. 3) that causes a frameshift mutation in exon 5. The chimpanzee Hpr is therefore not capable of producing a functional protein, which indicates that the Hpr gene is dispensable in chimpanzees.

We also determined a provisional nucleotide sequence for about 3.2 kb of the chimpanzee Hpp gene, starting from intron 4 to about 1 kb 3' to the termination codon of the gene. (About 70% of its sequence was determined on only one strand.)

In our previous work, we have shown that about 600 bp at the 3' end of the human Hp and Hpr genes are identical, and we suggested that this identity resulted from a gene conversion event between the two genes (Maeda, 1985). The present sequence data now allow us to establish when this gene conversion occurred, and which of the two genes was donor and which was recipient. In Fig. 3 we compare the chimpanzee Hpp, Hpr, and Hp sequences with the corresponding portions of the human Hpr and Hp genes. These comparisons show that in the region where the human Hp and Hpr sequences are identical (nucleotides 17385–18004), the human Hpr sequence differs from the chimpanzee Hpr sequence at 11 positions. There are, on the other hand, only 5 differences between the human Hpr and chimpanzee Hpp genes in the same region. The fact that the human Hpr sequence is more similar to the Hp gene than Hpr gene in chimpanzees in this region indicates that the nucleotide sequences in this 600-bp region of the human genes were transferred between the Hp
Evolution of the Haptoglobin Gene Cluster in Primates

FIG. 3. Nucleotide sequence comparison of the haptoglobin genes in chimpanzee and humans. The complete sequence of the human Hpr gene in the fourth intron through the region 9' to the gene is given. The numbers on the right margin correspond to nucleotide positions given in Fig. 2 of Maeda (1980). Hyphens are used to represent where the chpp, chpr, hpp, and hpr genes all share nucleotides identical to those in hpr. Where at least one sequence has a different nucleotide, all the nucleotides at the position are given. Asterisks indicate gaps inserted to maximize the identity of sequences. The splicing acceptor site at the beginning of exon 5 is marked by a downward arrowhead, the stop codon by a bracket, and an AATAAA sequence by a wavy underline. The direct repeats flanking the Alu repeat (boxed) in the fourth intron are overlined. The 9' end of the intron is in the human Hp and Hpr gene as a result of a gene conversion event is indicated by two bent arrows. Positions where human Hpr shares features only with chimpanzee Hpr, with chimpanzee Hpp, and with human Hp are marked by solid dots, by solid diamonds, and by solid triangles, respectively. Stretches of differences unique to one sequence are denoted by dashed overlines.
Evolution of the Haptoglobin Gene Cluster in Primates

and Hpr genes at some time after the divergence of humans and chimpanzees, and that the transfer was most likely from the human Hp gene to the human Hpr gene.

The Human Hpr Gene Was Formed by a Fusion between Ancestral Hpr and Hpp Genes—Our restriction enzyme mapping data show that the region flanking the human Hpr gene on its 3’ side is very similar to the corresponding region 3’ to the chimpanzee Hpp gene both in the presence of particular restriction sites (two HindIII sites and one BglII site occur at equivalent positions 3’ to the human Hpr and chimpanzee
Hpp genes, but not 3' to the chimpanzee Hpr gene) and in the occurrence of an Alu repeat 3 kb downstream from the two genes. On the other hand, the 5' region of the human Hpr gene looks very similar to the corresponding 5' region of the chimpanzee Hpr gene, both in the presence of the same retrovirus-like insertion in their first introns and in the occurrence of a shortened form of an Alu repeat in their fourth introns. Thus, one side of the human Hpr is like the chimpanzee Hpr gene, but the other side is like the chimpanzee Hpp gene. These reciprocal similarities are readily accounted for if the human Hpr locus is a hybrid formed from the 5' and 3' regions of two ancestral genes that are represented at the present time by the chimpanzee Hpr and Hpp genes. This hypothesis can be tested, and the region of crossing over can be located, by comparing the human Hpr gene sequence to the chimpanzee Hpr and Hpp gene sequences.

In the upper part of Fig. 4 we compare diagrammatically the nucleotide differences and similarities between these three genes. As can be seen from the figure, the human Hpr gene (hHpr in the diagram) differs by only 1.7% in sequence from the chimpanzee Hpr gene (cHpr) upstream of its exon 5, yet it differs from the chimpanzee Hpp gene (cHpp) by 8.2% in the same region. In the region 3' to exon 5, these relationships invert; hHpr differs by 6.1% from cHpr (0.5 kb compared) but by only 1.5% from cHpp (1.9 kb compared).

Thus, the comparison indicates the crossover point is located somewhere within a 1.3-kb region extending from the beginning of exon 5 to 500 bp on its 3' side. In this 1.3-kb region the relationships between the different sequences are not clear cut. For example, there are 2.8% differences between the human Hpr and chimpanzee Hpp sequences, 2.2% differences between the human Hpr and chimpanzee Hpr sequences, and the chimpanzee Hpr and Hpp genes differ by 2.3%. The similarities and the small magnitudes of these differences may be explained in part by selection at the protein level (since exon 5 represents a substantial part of the region of similarity), or it might be the consequence of some recombinational event(s).

![Diagram of sequence relationships](image)

**Fig. 4.** Summary of sequence relationships of the human Hpr gene to the chimpanzee Hpp (slashed line) and the chimpanzee Hpr (solid line) gene. The human and chimpanzee Hp genes are represented by wavy and beaded lines, respectively. Exons 2 through 5 are marked as E2-E5. The human Hpr sequence is divided into three regions and the percentage differences between hHpr and each of the other sequences are given. A bracket on the top indicates the region compared in Fig. 3. The length differences in the fourth intron caused by an Alu insertion in the common ancestor of Hpr and Hpp, and a deletion within the Alu sequence in the ancestral Hpr are illustrated by gaps with parenthesis. Nucleotide positions where hHpr shares features with cHpp, cHpr, or hHpp are marked by solid diamonds, solid dots, and solid triangles, respectively. Stretches of sequence difference which are unique to one of the sequences are marked by open boxes. G.C. refers to a gene conversion.

We inspected the sequences in this region in more detail, also taking the sequences of the Hpr genes in humans and chimpanzees into consideration. In Figs. 3 and 4, solid diamonds indicate positions where hHpr and cHpp share features that are not seen in the other three sequences, and solid dots indicate positions where hHpr and cHpr share features that are not in the others. These features are informative in tracing back the ancestral relationships. The solid diamonds indicate differences most likely accumulated in the Hpp precursor gene between the triplication event and the deletion crossover event. The solid dots indicate differences similarly accumulated in the Hpr precursor gene. As we can see in Fig. 4, the distribution of these particular shared features divides the region into two. All the solid dots occur 5' to nucleotide position 17105, and all but one of the solid diamonds occur 3' to position 17227. This suggests that the crossover point appears to be in the region from 17105 to 17227.

Multiple Gene Conversion Events Near the Crossover Point—Several other features within and near the region of the apparent crossover point are worth noting. In the total region compared in Fig. 3, there are 13 positions where hHpr and hHpp share features which are not found in either cHpr or cHpp. They are indicated by solid triangles in Figs. 3 and 4. Eight of them occur in the 5' part of the human Hpr gene, where a gene conversion event was detected previously. Three of them occur in the 5' part of exon 5; they create a region (16630-16868) of near identity between hHpr and hHpp. This represents another possible gene conversion events in humans. Because this region includes the position where a frameshift mutation has been found in the chimpanzee Hpr gene, it is conceivable that the human Hpr sequence also had the same mutation at one time, but it has since been corrected by gene conversion. Nucleotide sequence analyses of the Hpr genes from other primates, such as the gorilla, will be informative in this respect.

Another feature of interest is the occurrence of stretches of differences unique to one of the sequences. Three of them occur in the region and are marked in Fig. 3 by dashed overlines and in Fig. 4 by open boxes. For example, a cluster of nine differences unique to the human Hpr gene occur in a stretch of 44 base pairs within the region of apparent crossover point. Such a large number of differences in a small region and in only one of the genes is not easily explained by the random accumulation of mutations after a crossover event which took place in the human lineage no more than 5 million years ago. It rather implies that these mutations are old and that recent gene conversion events have made the three chimpanzee genes identical. These gene conversions probably partially account for the small number of percentage differences between the sequences, as described earlier, and obscure the exact point where the homologous unequal crossover occurred between the ancestral Hpr and Hpp genes.

The crossover leading from the triplicated loci to the present duplicated loci in humans has thus occurred in a region that is implicated in several other recombinational events. It is an extremely active region in recombination. Such regions have also been identified in the human 8-globin gene cluster (Chakravarti et al., 1984) and in the mouse major histocompatibility complex (Kobori et al., 1984). To date, however, the molecular basis of the properties of these recombinational "hot spots" has not been recognized.

**Evolution of the Haptoglobin Gene Cluster in Primates—** Fig. 5 summarizes the events that we hypothesize have shaped the haptoglobin gene cluster in primates. Initially, one gene was present, and this still remains in new world monkeys, such as spider monkeys and cebus monkeys. After primate
Evolution of the Haptoglobin Gene Cluster in Primates

three genes would originally have been of equal lengths. The inequality of the intergenic distances shows that the evolutionary history must have been more complex than a simple unequal crossover. Either a deletion of 3.5 kb in the region between Hp and Hpr caused a shorter distance between these genes, or a second insertion event in the region between Hpr and Hp made this distance longer. We have arbitrarily shown a deletion event in Fig. 4 to illustrate this idea.

Finally, after this divergence from the primate lineage, the human Hpr gene was formed by a fusion event resulting from a homologous unequal crossover between the ancestral Hpr and Hpp genes. Several other recombinational events have subsequently taken place in the gene cluster and have contributed to the generation of genetic variations in humans (Maeda and Smithies, 1986) including the formation of chromosomes with multiple copies of Hpr genes in the black population (Maeda et al., 1986).

In conclusion, we have shown that the haptoglobin gene cluster has expanded and contracted during the evolution of primates, and it has been shaped by gene conversions, and by retroviral and other sequence insertions. The history of this very small gene family is manifestly complex and serves to illustrate the important roles that recombinational events play in the evolution of the genes in multigene families.

Acknowledgments—We thank Drs. J. Murray, K. Chan, J. Slightom, and R. Spritz for providing us with primate DNA, and L. Becker, S. Hiller, and C. Bylstad for technical help. We thank Dr. Oliver Smithies for his support and valuable discussions.

REFERENCES


