Length Polymorphisms in Human Proline-Rich Protein Genes Generated by Intragenic Unequal Crossing Over

Karen M. Lyons, James H. Stein and Oliver Smithies

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

Manuscript received September 25, 1987
Revised copy accepted May 16, 1988

ABSTRACT

Southern blot hybridization analysis of genomic DNAs from 44 unrelated individuals revealed extensive insertion/deletion polymorphisms within the BstNI-type loci (PRB1, PRB2, PRB3 and PRB4) of the human proline-rich protein (PRP) multigene family. Ten length variants were cloned, including alleles at each of the four PRB loci, and in every case the region of length difference was localized to the tandemly repetitive third exon. DNA sequences covering the region of length variation were determined for seven of the alleles. The data indicate (1) that the PRB loci can be divided into two subtypes, PRB1 plus PRB2, and PRB3 plus PRB4, and (2) that the length differences result from different numbers of tandem repeats in the third exons. Variant chromosomes were also identified with different numbers of PRP loci resulting from homologous but unequal exchange between the PRB1 and PRB2 loci. The overall data are compatible with the observed length variants having been generated via homologous but unequal intragenic exchange. The results also indicate that these crossover events are sensitive to the amount of homology shared between the interacting DNA strands. Allelic length variants have arisen independently at least 20 times at the PRB loci, but only one has been detected at a PRH locus. Comparison of the detailed structures of the repetitive regions in PRB and PRH loci shows that the repeats in PRB genes are very similar to each other in sequence and in length. The PRH genes contain fewer repeats, which differ considerably in their individual lengths. These differences suggest that the larger number of length variants in PRB genes is related to their greater ease of homologous but unequal pairing compared to PRH genes.

STUDIES of the evolution of families of repetitive DNAs have revealed that these sequences show greater homogeneity within species than between species. SMITH (1976) constructed computer models that showed repeated unequal sister chromatid exchange can generate and maintain this homogeneity. Other investigators (SCHERER and DAVIS 1980; NAGYLAKI and PETES 1982) have considered the role of gene conversions in similar processes of homogenization.

In addition to these model-building studies, reciprocal recombination and gene conversion between equally and unequally paired repeated sequences have been directly demonstrated in yeast. Unequal sister chromatid exchange has been detected in the rDNA repeats during mitosis and meiosis (SZOSTAK and WU 1980; PETES 1980). Experiments involving dispersed repeats on the same chromosome have demonstrated that homologous but unequal recombination occurs in both meiosis (KLEIN and PETES 1981; KLEIN 1984; JACKSON and FINK 1985; JINKS-ROBERTSON and PETES 1986) and mitosis (JACKSON and FINK 1981; ROEDER, SMITH and LAMBE 1984). Gene conversions between repeated genes have also been reported (AMSTUTZ et al. 1985; JINKS-ROBERTSON and PETES 1985; JACKSON and FINK 1981).

The same types of recombinational events that occur in yeast between repeated sequences may also occur within genes that have internally repetitive structures. An increasing volume of literature documents the existence of such genes (see, for example, ALLISON et al. 1985; MUSKAVITCH and HOGNESS 1982; SHIN et al. 1985; WARREN, COROTTO and WOLBER 1986).

The human proline-rich protein (PRP) multigene family (reviewed by AZEN 1988) provides a particularly interesting opportunity to examine the types of recombinational events that can occur within genes with internally repetitive structures. This family consists of six tandemly linked genes, which encode the complex array of PRPs found in human saliva (AZEN et al. 1984; MÆDEA et al. 1985; LYONS, AZEN, GOODMAN and SMITHIES 1988).

The six PRP genes are arranged on chromosome 12p (MAMULA et al. 1985) in the order 5' PRB2-PRB1-PRB4-PRH2-PRB3-PRH1 3' (H.-S. KIM, unpublished data); the cluster spans a physical distance of approximately 600 kb with approximately 70 to 180 kb separating adjacent genes. The six genes are of two types that can be distinguished by their ability to hybridize to a probe derived from PRB1 (AZEN et al.
1984; Maeda 1985). Four of the genes (PRB1, PRB2, PRB3 and PRB4) hybridize strongly to the probe and contain multiple BstNI restriction sites (BstNI-type or PRB genes). The remaining two genes (PRH1 and PRH2) do not hybridize as strongly to the probe and contain multiple HaeIII restriction sites (HaeIII-type or PRH genes). Nucleotide sequence analyses (Maeda et al. 1985; Kim and Maeda 1986) have demonstrated that the PRPs contain a series of proline-rich tandem repeats which vary in length from 16 to 21 amino acids (48–63 bp).

Initial molecular studies of the human PRP multigene family (Azem et al. 1984) demonstrated that four of the six genes show frequent insertion/deletion length polymorphisms. We have extended these studies by analyzing at the DNA sequence level ten alleles from these four loci. This analysis reveals that the length differences are due to changes in the numbers of the proline-rich tandem repeats. We consider several models for the generation of these length differences and find that intragenic homologous but unequal exchange provides the most simple explanation among those examined. We also find that the number of PRP loci has been altered in some individuals by unequal exchanges between different loci.

MATERIALS AND METHODS

DNA preparation: High molecular weight DNA was prepared from peripheral blood leukocytes of 44 unrelated individuals, including three individuals examined in a previous study (Azem et al. 1984) by the method of Poncz et al. (1982).

Southern blot hybridizations: Human genomic DNA (5.0 µg per lane) was digested with EcoRI, electrophoresed in 0.8% agarose gels, and transferred to nitrocellulose according to the method of Southern (1975) with modifications as described by Wahl, Stern and Stark (1979). Filters were hybridized to a nick-translated 980-bp Hinfl fragment from the plasmid pPRPII2.2RP, which contains the 2.2-kb EcoRI/PstI fragment from PRP1 (Azem et al. 1984) subcloned into pAT153. The 980 bp Hinfl fragment used as the probe contains the tandemly repetitious exon 3 of PRB1. Hybridization conditions were as described in Vanin et al. (1983).

Construction and screening of phage libraries: Insert DNA was prepared by digesting human genomic DNA to completion with BamHI or HindIII (New England Biolabs). Bacteriophage λ Charon 35 (Loenen and Blattner 1983) arms were prepared, and phage libraries were constructed essentially as described by Maniatis, Fritsch and Sambrook (1982). The phage libraries were plated onto Escherichia coli strain K802recA.

Subcloning the polymorphic PRP fragments: DNA isolated from small recombinant bacteriophage growths was digested with EcoRI, subcloned into pAT153, and transformed into E. coli strain K802recA.

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). Nucleotide sequences were determined for both strands. The sequences were analyzed using software provided by the University of Wisconsin Genetics Computer Group (Devereux, Habetli and Smithies 1984).

RESULTS

Length polymorphisms: Previous studies (Azem et al. 1984) showed that the nick-translated 980-bp Hinfl fragment from PRB1 detected polymorphisms in human genomic DNA samples from three individuals.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Size (kb) of EcoRI fragment</th>
<th>No. of alleles in population</th>
<th>Observed/Expected*</th>
<th>No. of individuals heterozygous at locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRB1</td>
<td>6.2</td>
<td>44</td>
<td>18/23.1</td>
<td></td>
</tr>
<tr>
<td>PRB2</td>
<td>4.6</td>
<td>2</td>
<td>5/4.8</td>
<td></td>
</tr>
<tr>
<td>PRB3</td>
<td>4.4</td>
<td>1</td>
<td>18/16.6</td>
<td></td>
</tr>
<tr>
<td>PRB4</td>
<td>3.4</td>
<td>6</td>
<td>10/11.7</td>
<td></td>
</tr>
</tbody>
</table>

* Calculated from observed allele frequencies, assuming Hardy-Weinberg mating.

FIGURE 1.—Southern blot hybridization analysis of PRP gene patterns in six unrelated individuals. DNAs were digested with EcoRI. The 980-bp Hinfl fragment from pPRPII2.2RP was used as a probe. Assignment of length variants to specific loci was performed as described in the text. Alleles of PRB1 are identified by white squares. Alleles of PRB2, PRB3 and PRB4 are identified by white arrows placed on the center, right, and left, respectively, of the hybridizing band. Bands corresponding to PRH1 and PRH2 are identified by white dots (shown only in lane 2). The lengths of polymorphic alleles are indicated in kb.
We have extended these observations by examining EcoRI digests of genomic DNA from an additional 41 unrelated individuals. An illustrative set of results is presented in Figure 1. The polymorphisms detected in these EcoRI digests were also detectable with other restriction enzymes, demonstrating that the variants result from insertions or deletions of DNA rather than from point mutations.

The bands could be assigned to their respective loci by Southern blot analysis using several restriction enzymes (data not shown). Most of the polymorphic bands appear as doublets in Figure 1, with the length differences between the bands in a doublet ranging from approximately 100 to 300 bp (see, for example, the doublets at 4.3/4.0, 3.9/3.7, and 3.4/3.2 kb in individual 5). Several of the polymorphic bands in Figure 1 are not associated with obvious doublets because they result from a greater length difference. For example, individuals 1 and 4 in Figure 1 carry a 4.6-kb band associated with PRB2, and individual 2 carries a 3.3-kb band associated with PRB2.

Table 1 lists the sizes and frequencies in 44 unrelated individuals of the allelic length variants detected by Southern blot analysis of genomic DNA digested with EcoRI. Four different length variants were identified at each of the four PRB loci, but none were detected at either of the PRH loci. However, one length variant undetectable in Southern blots of the type used in our present study has been identified at the PRH1 locus and is discussed below. We also identified individuals with a decreased or an increased intensity of the EcoRI fragments associated with PRB1. These individuals are discussed in more detail below. No alterations in the intensities of bands associated with any other locus were evident.

In total, 26 different DNA patterns were identified among these 44 individuals, and 33 of the individuals were heterozygous for at least one allelic length variant. The expected number of heterozygotes was calculated for each locus from the observed frequencies among these 44 individuals, and 33 of the individuals were heterozygous for at least one allelic length variant. The expected number of heterozygotes was calculated for each locus from the observed frequencies of the different alleles assuming Hardy-Wienberg proportions (Table 1). The observed and expected values do not differ significantly.

**Restriction site map analysis of allelic length variants:** In order to investigate the nature of these frequent polymorphisms, ten different length variants were cloned, including examples from each of the four PRB loci. The length of each cloned fragment was confirmed directly by Southern blotting experiments of genomic DNA in order to insure that the cloned and uncloned genomic alleles were identical in length.

The restriction maps of the EcoRI fragments containing PRB loci are compared in Figure 2. PRB1 and PRB2 share many restriction enzyme recognition sites 5' to the region of length variation. Similarly, PRB3 and PRB4 have sites in common 5' to the region of length variation. Thus, PRB1 and PRB2 form one subtype, and PRB3 and PRB4 form a second subtype of the PRB loci.

The regions of length difference among the alleles were localized by restriction mapping (Figure 2). In all alleles, the region of length variation hybridized strongly to the 980-bp HinfI probe, suggesting that
the length differences occur in exon 3. The restriction sites flanking the regions of length variation were identical among sets of alleles except for a single *PstI* site present in the longest allele of *PRB2* that is absent in the other two alleles. This difference probably reflects a simple restriction site polymorphism. The restriction mapping experiments indicate that most of the polymorphic alleles varied in length by a multiple of about 60 bp, the approximate length of the repeats which comprise exon 3 (Figure 2).

DNA sequence analysis of polymorphic alleles: Nucleotide sequences for the regions of length variation were determined for seven of the ten alleles cloned in this study and are presented in the appendix. These sequences are compared in Figure 3. Although no alleles of *PRB2* were sequenced in our present study, the sequence of an incomplete cDNA from this locus was determined by MAEDA et al. (1985) and is included in Figure 3 for comparative purposes.

The DNA sequences presented in Figure 3 cover
Length Polymorphisms in PRP Genes

PRB3 and PRB4

the tandemly repetitious third exons. The sequences have been aligned to allow comparisons of individual repeats within each allele. Note that three types of repeated element (B1, B2 and B3) comprise most of exon 3 from PRB1 and PRB2 (Figure 3A) while a single type of repeated element (B4) comprises most of exon 3 from PRB3 and PRB4 (Figure 3B) (MAEDA et al. 1985). The overall consensus sequences that can be derived for the repeats in the two subtypes of the PRB loci are indicated in heavy boxes and bold type at the tops of the comparisons in Figures 3A and 3B. This sequence is also given above the sequence of each allele, for alignment purposes. The consensus sequences that can be derived for the B1, B2, B3 and B4 repeats in each locus are indicated next in heavy boxes and regular type, with only the differences from the overall consensus sequences for each subtype being shown. The actual nucleotide sequences from each allele are presented next within the light boxes, again with only differences from the repeat consensus sequences for the B1, B2, B3 and B4 repeats being indicated.

Comparisons of the DNA consensus sequences for the tandem repeats derived from each of the PRB loci (Figure 3) confirm that the PRB genes can be divided into two subtypes. PRB1 and PRB2 form one subtype

---

FIGURE 3.
and share a repeating array of B1, B2 and B3 repeats in exon 3. The consensus sequences of the B1 and B2 repeats are identical, and the consensus sequences for the B3 repeats have only a single difference: the B3 repeat in \( PRB1 \) contains a 3-bp gap that is not present in the consensus repeat from \( PRB2 \).

\( PRB3 \) and \( PRB4 \) form a second subtype. The third exons of these genes are composed largely of a single repeat type, B4, and the consensus sequences for the B4 repeats from \( PRB3 \) and \( PRB4 \) differ at only two positions. This subdivision is further supported by the observation that the 3' portions (region "C") of \( PRB1 \) and \( PRB2 \) contain a gap of 27 bp that is not found in \( PRB3 \) or \( PRB4 \).

Figure 3B shows that the two sequenced alleles of \( PRB3 \) differ in the number of B4 repeats that they contain. \( PRB3L \) (Short) has six copies of this repeat, while \( PRB3H \) (Long) has ten copies. Similarly, each of the three alleles of \( PRB4 \) contain different numbers of the B4-type repeat: \( PRB4L \) (Short) has six copies, \( PRB4M \) (Medium) has seven copies, and \( PRB4H \) (Long) has nine copies. The DNA sequences from the two alleles of \( PRB1 \) presented in Figure 3A demonstrate that these alleles differ in length by three repeats in such a way that the array of B1, B2 and B3 repeats is maintained. Although nucleotide sequences were not determined for the three alleles of \( PRB2 \) cloned in this study, their restriction maps indicate that these alleles also vary in length by multiples of three repeats: \( PRB2L \) (Short) appears to be three repeats (\( 3 \times 63 \) bp \( \approx 180 \) bp) shorter than \( PRB2H \) (Long) while \( PRB2M \) (Very Long) appears to be 15 repeats longer than \( PRB2L \) (15 \( \times 63 \) bp \( \approx 900 \) bp). Thus the evidence suggests that alleles of \( PRB1 \) and \( PRB2 \), with the three-repeat array, differ in length only by multiples of three repeats.

**Analysis of base substitutions:** Inspection of the individual sequences presented in Figure 3 reveals that most repeats within a locus vary from the consensus B1, B2, B3 or B4 repeat for that locus by fewer than five base substitutions. This sequence conservation might reflect selection for the maintenance of a strictly repeated amino acid sequence and/or homogenization of the sequences due to repeated unequal exchange.

The potential contribution of selection for a particular amino acid sequence was investigated by comparing the DNA sequence of each repeat to the repeat consensus sequence that it most resembles. The number of base pair changes that result in amino acid replacements and the number that are silent changes were determined for each repeat in \( PRB1M \), \( PRB2L \), \( PRB3L \), and \( PRB4H \) (in total, 145 replacement substitutions and 60 silent substitutions). This comparison reveals that the observed ratio of replacement to silent substitutions (2.4:1) is nearly identical to that expected on the assumption that the three possible base substitutions occur with equal probability (2:1:1).

This result implies near neutrality of amino acid changes. Thus, although there is some evidence for selection for the production of this group of proteins (all gaps are 3 bp in length, and no nonsense substitutions have been observed), there is no evidence for selection of a particular amino acid sequence. In fact, the observation of individuals with deletions involving the \( PRB1 \) and \( PRB2 \) loci (see below) suggests that there is not strong selection for the production of a particular PRP. Many of the repeats in each allele share the same base substitutions from the repeat consensus sequence suggesting that some mechanism has resulted in the distribution of base substitutions across more than one repeat unit.

**Generation of allelic length variants:** There are several mechanisms that could generate changes in the numbers of repeats in an array. Intergenic unequal crossovers could produce not only alterations in the number of repeats within a single allele, but would also change the number of loci on the chromosome. Gene conversion events involving unequal numbers of repeats on the interacting strands, or intragenic unequal crossovers, would not change the number of loci on the chromosome.

We can rule out single homologous but unequal exchanges between different loci as the mechanism responsible for the generation of the seven sequenced length variants, since these variants were not associated with changes in the intensity of the bands corresponding to any other PRP loci.

A pairwise comparison of all repeats permits inference of the types of events that generated the sequenced length variants. The analysis was performed using the COMPARE program (DEVEREUX, HABERLI and SMITHIES 1984). Gaps were introduced where necessary to maximize alignments. All gaps were found to be 3 bp in length, and were scored as single differences.

Different mechanisms for the generation of length variants predict different outcomes from the pairwise comparisons. For example, an exchange of repeats between different loci resulting from a gene conversion event could be detected by the existence of fewer differences between repeats from a different locus (the donor locus) than between repeats from an allele of the same locus (the recipient locus). On the other hand, unequal exchange between alleles would be indicated by the existence of fewer differences between repeats from alleles at the same locus than between repeats from different loci.

The pairwise comparisons indicate that in all the sequences examined there are as few or fewer differences between the repeats from alleles than between repeats from different loci. For example, no evidence
was found for the introduction of B1, B2, or B3 repeats from PRB1 or PRB2 into PRB3 or PRB4. Similarly, no B4 repeats were found in any alleles of PRB1 or PRB2. Thus, there is no evidence for the occurrence of gene conversions between PRP genes of different subtypes.

In the comparisons made between alleles of PRB1 and PRB2, and between alleles of PRB3 and PRB4, some of the repeats from the two loci were found to have as few differences as the repeats from within a single locus. Such sequence similarities between the loci within each subtype may reflect either their recent divergence from an ancestral gene or recent gene conversions, which can under favorable circumstances be distinguished (see, for example, SLIGHTOM et al. 1980). The sequence differences between the members of the PRB subtypes are, however, so few that this distinction is not possible with the present data for either the PRB1 and PRB2 loci or the PRB3 and PRB4 loci. Short gene conversions between the members of these two pairs of loci that do not result in the generation of length variation may have occurred.

The pairwise comparisons facilitated the construction of specific models able to account for the generation of each of the variants studied at the PRB loci (Figure 4). The models were constructed so as to minimize the number of additional base substitutions that were also needed to account for the data.

The most simple explanation for the generation of each of the allelic length variants is that it is the product of a single homologous but unequal crossover event between alleles or sister chromatids. The generation of PRB4M, for example, can be entirely accounted for by a single unequal crossover between PRB4L and PRB4S. For the remaining pairs of alleles, base substitutions must be assumed in order to fully explain the generation of one allelic length variant from another. Thus, in order to fully account for the generation of PRB3L from PRB3S, one crossover and six nucleotide substitutions were required. Similarly, the generation of PRB1S from PRB1M requires one crossover and six nucleotide substitutions. It is, of course, possible that some of the unsequenced length variants would require fewer nucleotide substitutions in order to account for the generation of the sequenced length variants. In spite of this uncertainty, the models postulating unequal crossovers between alleles or sister chromatids require fewer events in order to account for the generation of allelic length variants than do any other models we tested, including models involving unequal gene conversions (or unequal double crossovers).

Unequal crossovers between loci: Examination of genomic DNAs by blot hybridization revealed that several individuals had bands corresponding to PRB1 of greater or less than normal intensity. In one individual, the EcoRI fragment containing the PRB1 gene was completely absent and was not replaced by any alternate fragment (Figure 5A, lane 3). This individual appears to be homozygous for a chromosome carrying a deletion involving the PRB1 locus. We have
also observed two individuals in our sample of 44 with a band corresponding to PRB1 which appears to be half as intense as the band found in most individuals, suggesting that they are heterozygous for a deletion-bearing chromosome (data not shown). Three individuals had a band of greater than normal intensity at the PRB1 locus. The 6.2-kb EcoRI fragment in lane 2 of Figure 5A, for example, has a greater intensity than the fragments associated with PRB1 in lanes 1 and 4, suggesting that this individual carries a duplication involving PRB1. In addition, one individual with three different length variants of PRB1 was detected in our sample (data not shown). These observations show that duplications and deletions involving PRB1 are relatively common in the studied population.

In order to investigate the nature of a chromosome carrying a presumptive deletion, genomic DNA (kindly provided by P. O'CONNELL) from the individual homozygous for this rearrangement was digested with three different restriction enzymes and analyzed by agarose gel electrophoresis (lane 3, Figure 5, A–C). The Southern blot was hybridized to the 980 bp HindIII probe. With all three enzymes a band corresponding to either PRB1 or PRB2 is absent. These data and the results of digestion with several other restriction enzymes (data not shown) permit the derivation of a map of restriction sites for the PRP gene region on this variant chromosome and a description of its history.

Comparison of the maps of the normal PRB1 and PRB2 regions (Figure 6, A and B, respectively) shows that although the positions of many restriction enzyme sites are identical in the two regions, several restriction sites or fragment lengths are unique to either the PRB1 or the PRB2 gene region. Comparison of the restriction map of the fusion gene (reminiscent of the Lepore δ-δ gene) on the deletion chromosome (Figure 6C) with the maps of the normal PRB1 and PRB2 loci shows that the restriction map of the fusion gene region is identical to that of the PRB2 gene region 5’ to and including the tandemly repetitive third exon. The restriction maps of the regions 5’ to exon 3 in the fusion gene and the PRB1 gene are identical. These results are consistent with the generation of the variant chromosome via a homologous but unequal crossover event between PRB1 and PRB2 as illustrated in Figure 6, with the region of crossover being localized to a 700-bp region defined by the BstEII and EcoRI sites in the repetitive third exon of PRB1 (assuming that there are no unrecognized restriction site polymorphisms that alter these sites).

Genomic DNA from the individual having an EcoRI fragment associated with PRB1 of greater than normal intensity (lane 2 in Figure 5A) was also digested with BamHI and HindIII (lane 2 in Figure 5, A–C). With all three enzymes, the band corresponding to either

---

**FIGURE 5.—Southern blot hybridizations of restriction digests of genomic DNA from four individuals (lanes 1 to 4). Individual 3 has a deletion detected by the HindIII 980 bp probe. Individual 2 carries a band of increased intensity. Individuals 1 and 4 are normal. The locations of the missing bands in individual 3 and the bands of increased intensity in individual 2 are indicated respectively by dots and squares. Their sizes are indicated in kb.**

---

**FIGURE 6.—Formation of a PRB fusion gene. A, Map of the PRB2 gene region; B, map of the PRB1 gene region; C, map of the PRB2/PRB1 fusion gene region. Only sites pertinent to mapping with the HindIII 980-bp probe are shown. In C, sites indicated above the line are identical to those in PRB2; sites below the line are identical to those in PRB1. The potential crossover zone generating the fusion gene is indicated by vertical lines surrounding a cross. The open boxes represent the tandemly repetitive third exon. Restriction enzymes are B, BamHI; Bs, BstEII; E, EcoRI; H, HindIII; S, SstI; X, XbaI.**

PRB1 (in the EcoRI digest) or PRB2 (in the HindIII or BamHI digests) is of greater than normal intensity. These results are consistent with the presence of a PRB1/PRB2 fusion gene, which migrates with PRB1 in the EcoRI digest and with PRB2 in the BamHI and HindIII digest, in addition to the normal PRB1 and PRB2 loci on this variant chromosome. Thus, this individual is carrying a duplication comparable to that which would be formed as the reciprocal of the event leading to the deletion illustrated in Figure 6 [reminiscent of the anti-Lepore β-δ gene (Hemoglobin Miyada)].

**DISCUSSION**

Previous cDNA sequence studies of the human PRP multigene family (MAEDA et al. 1985) demonstrated that the six loci could be classified into two groups [the HaeIII-type (PRH) and the BstNI-type (PRB) loci]
based on the sequences of the tandem repeats comprising the third exons of these genes. The present study demonstrates that the four PRB genes can be further subdivided; PRB1 and PRB2 form one subtype, and PRB3 and PRB4 form a second subtype. This observation suggests that sequences have been exchanged recently between the members of each of these two subclasses and/or that the members of each subclass share a recent common ancestor.

Although the data presented here do not allow precise determination of the evolutionary relationships among the four PRB loci, they do indicate that repeated unequal crossing over between alleles (or sister chromatids) has played a significant role in the evolution of the loci. Inspection of the sequences presented in Figure 3 reveals that the same base difference relative to the repeat consensus sequence is often found in several repeats within a given allele. This observation is consistent with SMITH's (1974) proposal that base substitutions spread to multiple copies of a tandem repeat as a consequence of repeated intragenic unequal exchange. Evidence supporting a predominant role for repeated intragenic exchanges rather than events involving alleles at different loci in the evolution of the PRB genes is our observation that the repeat consensus sequences derived for each of the PRB loci are unique. The consensus B3 repeat sequences from PRB1 and PRB2 differ, as do the consensus B4 repeats from PRB3 and PRB4.

Our study demonstrates that length variants are common at the PRB loci. We have identified 20 insertion/deletion polymorphisms at these loci. Pairwise comparisons indicate that the most simple explanation for the generation of these length differences is via homologous but unequal intragenic exchange. Although several different examples of length variants were detected at each of the four PRB loci, none were found in this study at either of the PRH loci. This fact that the third exons of PRH1 and PRH2 are composed mainly of tandem repeats (KIM and MAEDA 1986). However, AZEN et al. (1987) found that one of the three common alleles at the PRH1 locus is one repeat longer than the other two alleles. Thus, length variants appear to arise much more readily at the PRB loci than at the PRH loci.

The relative paucity of allelic length variants at the PRH loci could be due either to structural features that interfere with unequal pairing, and/or to selection against size variants. The observation that the one variant PRH1 allele is fairly common in Caucasians suggests that selection against size variants is not likely to be the major cause for this difference. Accordingly we examined the sequences of PRH and PRB genes to see if we could find some features that might account for the difference.

The consensus sequences for PRH and PRB repeats (KIM and MAEDA 1986) are approximately 80% homologous. No structural features, such as Chi sites and simple sequence DNA, are evident that might predispose the PRB loci to recombine. However, the organization of the third exons of the PRH loci suggests two features that might explain the infrequent generation of length variants at these loci compared to the PRB loci. First, the third exons of the two PRH genes contain only five tandem repeats, whereas the number of repeats in the PRB genes is higher (range 8-13, median 10-11), allowing a greater opportunity for unequal pairing. Secondly, the five individual repeats in the PRH loci exhibit substantial length variation, being, respectively, 48, 51, 51, 63 and 63 bp long. These length differences are very likely to affect adversely the stability of misaligned DNA strands. In contrast, the PRB repeats are in general all the same length. Consequently, we suggest that the length differences among the repeats found in PRH loci account for the differences in the frequency of generation of length variants between the PRH and PRB loci. We cannot, however, rigorously exclude the possibility that the number of repeats in the protein products encoded by the PRH loci is more constrained as a result of some critical structure needed for their physiological function.

We identified intergenic unequal crossover events involving the PRB1 and PRB2 loci, but no other PRP loci. The observation of both deletions and duplications involving these loci suggests that such intergenic events are not rare, since it is extremely unlikely that the deleted and duplicated chromosomes arose during a single meiosis and were both found in our studied population. One factor that may influence the relative frequencies of unequal recombination between loci is the chromosomal organization of the PRP multigene family. Thus, PRB1 and PRB2 are physically closer (70 kb) to each other than are any of the other loci (H.-S. Kim, unpublished data). PRB3 and PRB4, on the other hand, are separated from each other by approximately 350 kb and one of the PRH loci is located between them (H.-S. Kim, unpublished data). Unequal crossover between the PRB3 and PRB4 loci would therefore lead to the loss (or gain) of two loci and a total of 350 kb of DNA. This factor may be partly or even completely responsible for the observation that unequal crossing over between PRB3 and PRB4 has not been detected by us or by others (O'CONNELL et al. 1987).

Studies on homology requirements for recombinination in bacteria (WATT et al. 1985; SHEN and HUANG 1986) have demonstrated that efficient homologous recombination requires an uninterrupted stretch of at least 25-50 bp of perfect homology, and that even a single mismatch dramatically lowers the efficiency. Experiments in mammalian cells (RUBNITZ and SUBRAMANI 1984; AYARES et al. 1986; LISKAY, LETSOU and STACHELEK 1987) have also demonstrated a strict
homology requirement for homologous recombination.

Taken as a whole, our evidence suggests that synapsis formation during unequal pairing within the repetitious regions of the PRB loci is very sensitive to sequence differences among repeats. Thus, the strict maintenance of the B1-B2-B3 array in alleles of PRB1 and PRB2 can be explained as the consequence of the inability of the B3 repeats, which contain 3 bp gaps relative to the B1 and B2 repeats, to pair with the B1 or B2 repeats. The greater length variations of the repeats within alleles of PRH loci could account for the observation of only a single length variant at a PRH locus. Our results thus provide further evidence that homologous recombination is very sensitive to differences in the amounts of homology between interacting strands.

We thank Peter O’Connell for providing several human genomic DNAs, and Hyung-Suk Kim for permission to quote unpublished data. We also thank Patricia A. Powers, Philip Howles and James F. Crow for helpful discussions on the manuscript. This work was supported by National Institutes of Health grant GM20069.

LITERATURE CITED


B2


SLIGHTOM, J. L., A. E. BLECHL and O. SMITHIES, 1980  Human fetal \( ^{\gamma} \) and \( ^{\delta} \)-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell 21: 627-638.


Communicating editor: B. W. GLICKMAN

APPENDIX

Nucleotide sequences (Figure 7) of the Hinfl (PRB1\(^\gamma \), PRB1\(^\delta \), PRB4\(^\gamma \), PRB4\(^\delta \)) and HaellI plus HaellI/Hinfl fragments containing the tandemly repetitious third exons. Sequences from the intron proceeding the third exon are in lower case. Sequences from the third exon are in upper case.

![Figure 7](image-url)

---

**PRB1**

![Nucleotide sequences from the third exons of PRB1, PRB2, PRB3, and PRB4 (figure is continued on p. 278).](image-url)
FIG. 7—(Continued)