

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 *Bst*NI-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 repetitious 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 repetitious 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 LAMBIE 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 repetitious 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 repetitious structures. This family consists of six tandemly linked genes, which encode the complex array of PRPs found in human saliva (AZEN *et al.* 1984; MAEDA *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 *Bst*NI restriction sites (*Bst*NI-type or *PRB* genes). The remaining two genes (*PRH1* and *PRH2*) do not hybridize as strongly to the probe and contain multiple *Hae*III restriction sites (*Hae*III-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 multi-gene family (AZEN *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 (AZEN *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 *Eco*RI, 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 *Hinf*I fragment from the plasmid pPRPII2.2RP, which contains the 2.2-kb *Eco*RI/*Pst*I fragment from PRP1 (AZEN *et al.* 1984) subcloned into pAT153. The 980 bp *Hinf*I 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 *Bam*HI or *Hind*III (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 *Eco*RI, 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 pro-

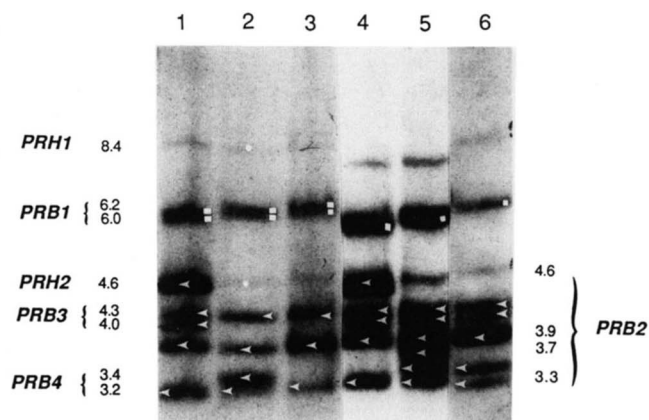


FIGURE 1.—Southern blot hybridization analysis of PRP gene patterns in six unrelated individuals. DNAs were digested with *Eco*RI. The 980-bp *Hinf*I fragment from pPRPII2.2RP was used as a probe. Assignment of length variants to specific loci was performed as described in the text. Alleles of *PRB1* 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.

TABLE 1

Incidence of polymorphic alleles in 44 unrelated individuals

Locus	Size (kb) of <i>Eco</i> RI fragment	No. of alleles in population	Observed/Expected ^a No. of individuals heterozygous at locus
<i>PRB1</i>	6.2	44	18/23.1
	6.0	36	
	5.8	7	
	5.6	1	
<i>PRB2</i>	4.6	2	5/4.8
	3.9	83	
	3.7	2	
	3.3	1	
<i>PRB3</i>	4.4	1	18/16.6
	4.3	67	
	4.2	2	
	4.0	18	
<i>PRB4</i>	3.4	6	10/11.7
	3.3	4	
	3.2	75	
	3.0	3	

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

vided by the University of Wisconsin Genetics Computer Group (DEVEREUX, HABERLI and SMITHIES 1984).

RESULTS

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

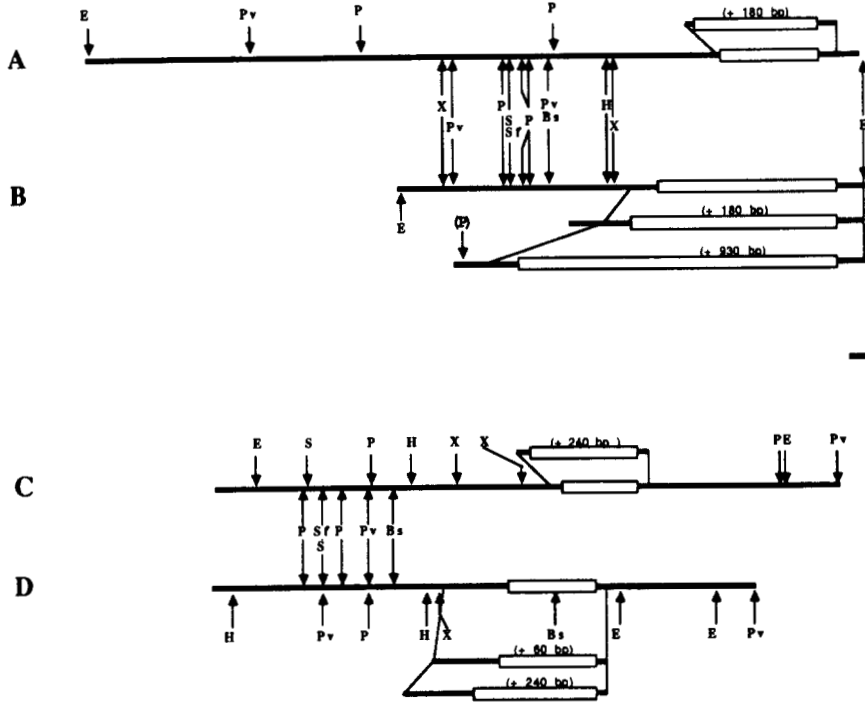


FIGURE 2.—Maps of the four human *PRB* loci illustrating their two subtypes and their allelic variants. A, *PRB1*; B, *PRB2*; C, *PRB3* and D, *PRB4*. The restriction sites indicated are *BstEII* (Bs), *EcoRI* (E), *HpaI* (H), *PstI* (P), *PvuII* (Pv), *SstI* (S), *SfiI* (Sf), and *XbaI* (X). Restriction sites unique to a particular gene are indicated by single-headed arrows. Restriction sites in identical positions in *PRB1* and *PRB2* or in *PRB3* and *PRB4* are indicated by double-headed arrows. The regions of length variation were localized and are shown above or below the respective complete maps for the shortest allele at each locus. The approximate length difference between the shortest allele at each locus and the other allele(s) is indicated in bp above the map of each length variant. The restriction maps of alleles at each locus are identical except for the *PstI* site indicated in parentheses that is present in *PRB2^{VL}* but not in the other alleles of *PRB2*. Open boxes represent the regions of hybridization to the 980-bp *HinfI* probe.

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 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

A *PRB1 and PRB2*

overall consensus repeat: CCTCCTCCAGGAAAGCCACAAGGACCACCCCAAGGAGGCAACAAGCCCAAGGTCCCCA

PRB1 consensus repeats:

CCTCCTCCAGGAAAGCCACAAGGACCACCCCAAGGAGGCAACAAGCCCAAGGTCCCCA

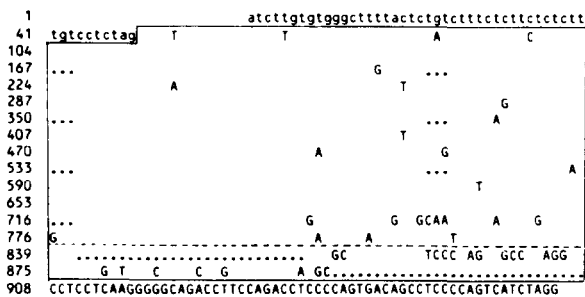
con (B1)	T	C
con (B2)	...	A T
con (B3)	...	A ... T G A G

PRB2 consensus repeats:

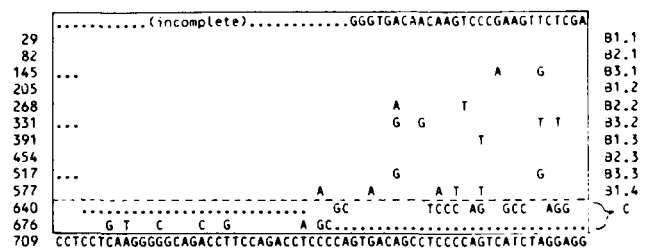
CCTCCTCCAGGAAAGCCACAAGGACCACCCCAAGGAGGCAACAAGCCCAAGGTCCCCA

con (B1)	T	C
con (B2)	...	A T
con (B3)	...	T G A G

PRB1^M



PRB2L (incomplete)



PRB1^S

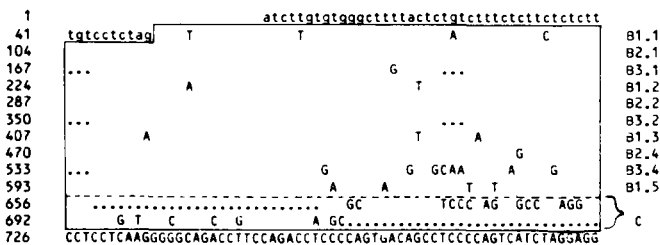


Figure 3.—Nucleotide sequences from portions of eight alleles of the four PRB loci: *PRB1*, *PRB2*, *PRB3* and *PRB4*. The DNA sequences shown in bold type within heavy boxes are the overall consensus sequences for the repeats present in exon 3 from alleles of *PRB1* plus *PRB2* (A) and *PRB3* plus *PRB4* (B). These overall consensus sequences are also given, for alignment purposes, above each locus. The consensus sequences for the B1, B2 and B3 repeats in *PRB1* and *PRB2*, and for the B4 repeats in *PRB3* and *PRB4* are displayed next in the heavy boxes above each locus with only differences from the overall consensus sequences being indicated. The DNA sequence from each allele is lightly boxed. Each repeat has been given a separate designation, indicated to the right of the sequence, that indicates the repeat type. For example, repeat B1.4 in *PRB1^M* is a B1-type repeat, and differs from the consensus B1 repeat for *PRB1* by the single nucleotide T that is indicated. The DNA sequences above the dashed line include repeats that show few differences from the consensus sequences derived for the specific locus (indicated in light type within heavy boxes). DNA sequences below the dashed line (designated "C") show substantial differences from these consensus repeat sequences, and in this case the indicated differences are from the overall consensus sequence (indicated in bold type within heavy boxes). Lower case DNA sequences represent intronic regions. Dots indicate gaps introduced to maintain alignment of repeats.

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 *Pst*I 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

B

PRB3 and PRB4

overall consensus repeat: **CCTC^C_GTCC^A_GGGAAAGCCAGAAGGACCACCC^CCCACAAGGAGGAACCAGTCCCAAGGTCCCCA**

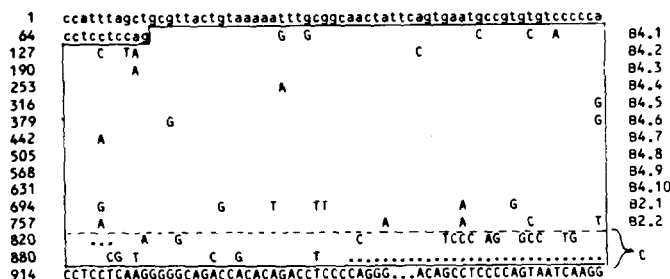
PRB3 consensus repeat:

CCTC^C_GTCC^A_GGGAAAGCCAGAAGGACCACCC^CCCACAAGGAGGAACCAGTCCCAAGGTCCCCA
G G

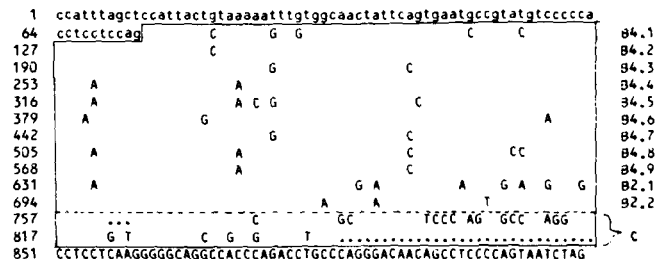
PRB4 consensus repeat:

CCTC^C_GTCC^A_GGGAAAGCCAGAAGGACCACCC^CCCACAAGGAGGAACCAGTCCCAAGGTCCCCA
C A

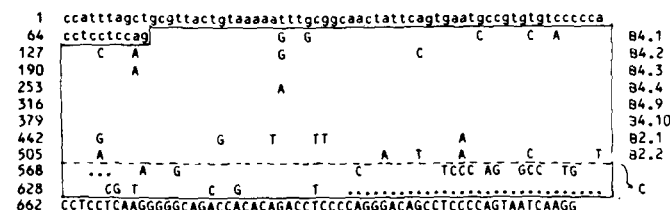
PRB3L



PRB4L



PRB3S



PRB4M



PRB4S

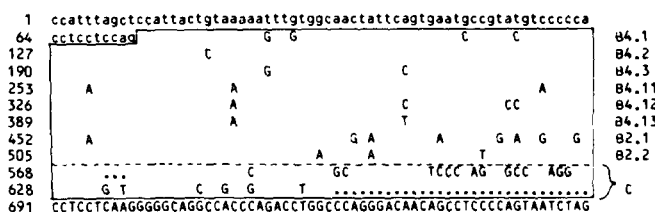


FIGURE 3.

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 se-

quences 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

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. *PRB3^S* (Short) has six copies of this repeat, while *PRB3^L* (Long) has ten copies. Similarly, each of the three alleles of *PRB4* contains different numbers of the B4-type repeat: *PRB4^S* (Short) has six copies, *PRB4^M* (Medium) has seven copies, and *PRB4^L* (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: *PRB2^S* (Short) appears to be three repeats (3×63 bp ≈ 180 bp) shorter than *PRB2^L* (Long) while *PRB2^{VL}* (Very Long) appears to be 15 repeats longer than *PRB2^S* (15×63 bp ≈ 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 *PRB1^M*, *PRB2^L*, *PRB3^L*, and *PRB4^L* (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 an 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

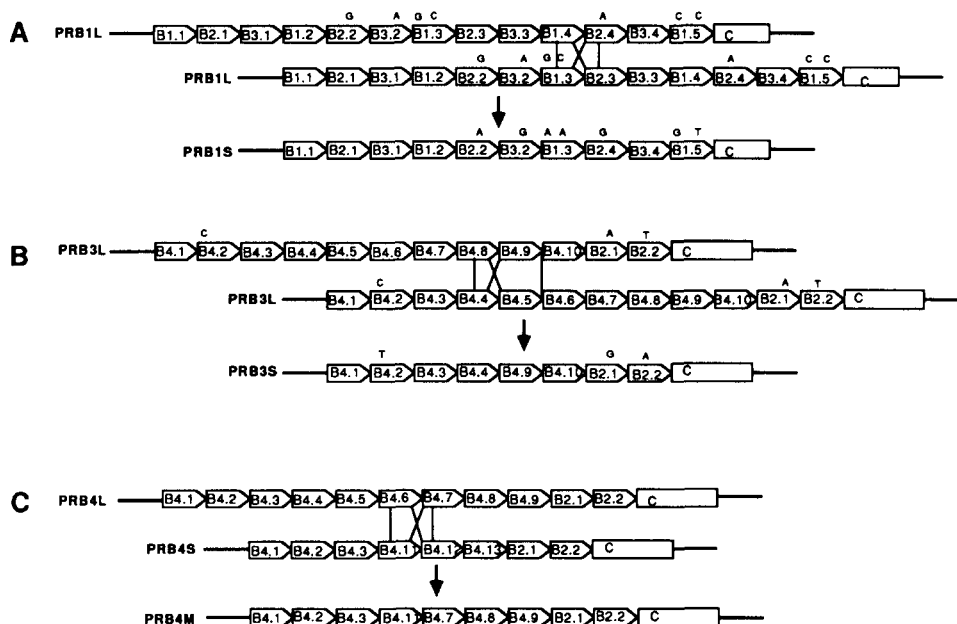


FIGURE 4.—Specific models for the generation of allelic length variants at the *PRB* loci. Each open arrow represents one repeat. The names within each arrow are the same as those used to identify the repeats in Figure 3. Crossover zones are indicated by vertical lines surrounding a cross, and nucleotide differences between allelic variants are identified above the repeat in which they occur. The exact locations of these base differences can be determined by reference to Figure 3.

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 crossover events required. Attempting to account for each of the sequenced alleles as being derived from events involving the other sequenced allele(s) is clearly artificial, because the true histories are likely to involve alleles that we have not studied. Nevertheless, the models allow us to delineate the *types* of events that may have occurred. In constructing the models, the locations of the postulated crossover events were assigned so as to mini-

mize 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 *PRB4^M*, for example, can be entirely accounted for by a single unequal crossover between *PRB4^L* and *PRB4^S*. 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 *PRB3^S* from *PRB3^L*, one crossover and three nucleotide substitutions were required. Similarly, the generation of *PRB1^S* from *PRB1^M* 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 *HinfI* 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 - β 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 repetitious third exon. The restriction maps of the regions 3' 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 repetitious 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

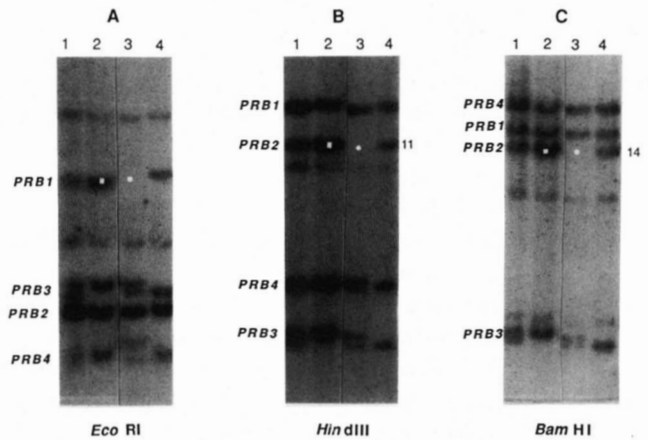


FIGURE 5.—Southern blot hybridizations of restriction digests of genomic DNA from four individuals (lanes 1 to 4). Individual 3 has a deletion detected by the *HinfI* 980 bp probe. Individual 2 carries a band of increased intensity. Individuals 1 and 4 are normal. The locations of the missing bands in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

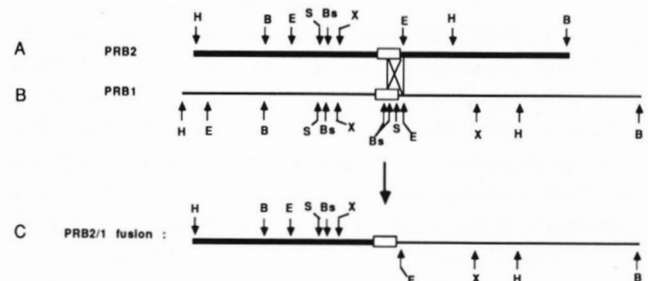


FIGURE 6.—Formation of a PRB fusion gene. A, Map of the *PRB2* gene region; B, map of the *PRB1* gene region; C, map of the *PRB2/1* fusion gene region. Only sites pertinent to mapping with the *HinfI* 980-bp probe are shown. In C, sites indicated above the line are identical to those in *PRB2*; sites 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 repetitious 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* digests, 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 in spite of the 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 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% ho-

mologous. 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 recombination 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

- ALLISON, L. A., M. MOYLE, M. SHALES and C. J. INGLES, 1985 Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* **42**: 559-610.
- AMSTUTZ, H., P. MUNZ, W.-D. HEYER, U. LEUPOLD and J. KOHLI, 1985 Concerted evolution of tRNA genes: intergenic conversion among three unlinked serine tRNA genes in *S. pombe*. *Cell* **40**: 879-886.
- AYARES, D., L. CHERKURI, K.-Y. SONG and R. KUCHERLAPATI, 1986 Homology requirements for intermolecular recombination in mammalian cells. *Proc. Natl. Acad. Sci. USA* **83**: 5199-5203.
- AZEN, E. A., 1988 Genetic protein polymorphisms of human saliva. In: *Clinical Chemistry of Human Saliva*, Edited by J. TENOVUO, CRC Press, Inc., Cleveland, Ohio (in press).
- AZEN, E. A., P. GOODMAN, S. FLYNN, N. MAEDA and H.-S. KIM, 1987 Alleles at the *PRH1* locus coding for the human salivary acidic proline-rich proteins (PRPS) Pa, Db, and PIF. *Am. J. Hum. Genet.* **41**: 1035-1047.
- AZEN, E., K. M. LYONS, T. MCGONIGAL, N. L. BARRETT, L. S. CLEMENTS, N. MAEDA, E. F. VAVIN, D. M. CARLSON, and O. SMITHIES, 1984 Clones from the human gene complex coding for salivary proline-rich proteins. *Proc. Natl. Acad. Sci. USA* **81**: 5561-5565.
- DEVEREUX, J., P. HABERLI and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**: 387-395.
- JACKSON, J. A., and G. R. FINK, 1981 Gene conversion between duplicated genetic elements in yeast. *Nature* **292**: 306-311.
- JACKSON, J. A., and G. R. FINK, 1985 Meiotic recombination between duplicated genetic elements in *Saccharomyces cerevisiae*. *Genetics* **109**: 303-332.
- JINKS-ROBERTSON, S., and T. D. PETES, 1985 High-frequency meiotic gene conversion between repeated genes on nonhomologous chromosomes in yeast. *Proc. Natl. Acad. Sci. USA* **82**: 3350-3354.
- JINKS-ROBERTSON, S., and T. D. PETES, 1986 Chromosomal translocations generated by high-frequency meiotic recombination between repeated yeast genes. *Genetics* **114**: 731-752.
- KIM, H.-S., and N. MAEDA, 1986 Structures of two *HaeIII*-type genes in the human salivary proline-rich multigene family. *J. Biol. Chem.* **261**: 6712-6718.
- KLEIN, H. L., 1984 Lack of association between intrachromosomal gene conversion and reciprocal exchange. *Nature* **310**: 748-753.
- KLEIN, H. L., and T. D. PETES, 1981 Intrachromosomal gene conversion in yeast. *Nature* **289**: 144-148.
- LISKAY, R. M., A. LETSOU and J. L. STACHELEK, 1987 Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. *Genetics* **115**: 161-167.
- LOENEN, W. A. M., and F. R. BLATTNER, 1983 Lambda Charon vectors (Ch 32, 33, 34 and 35) adapted for DNA cloning in recombination-deficient hosts. *Gene* **26**: 171-179.
- LYONS, K. M., E. A. AZEN, P. A. GOODMAN and O. SMITHIES, 1988 Many protein products from a few loci: assignment of human salivary proline-rich proteins to specific loci. *Genetics* **120**: 255-265.
- MAEDA, N., 1985 Inheritance of the human salivary proline-rich proteins: a reinterpretation in terms of six loci forming two subfamilies. *Biochem. Genet.* **23**: 455-464.
- MAEDA, N., H.-S. KIM, E. A. AZEN and O. SMITHIES, 1985 Differential splicing and post-translational cleavages in the human salivary proline-rich protein gene system. *J. Biol. Chem.* **260**: 11123-11130.
- MAMULA, P. W., N. A. HEEREMA, C. G. PALMER, K. M. LYONS and R. C. KARN, 1985 Localization of the human salivary protein complex (SPC) to chromosome band 12p13.2. *Cytogenet. Cell Genet.* **39**: 279-284.
- MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning (A Laboratory Manual)*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- MAXAM, A. M., and W. GILBERT, 1977 A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74**: 560-654.
- MUSKAVITCH, M. A. T., and D. S. HOGNESS, 1982 An expandable gene that encodes a Drosophila glue protein is not expressed in variants lacking remote upstream sequences. *Cell* **29**: 1041-1051.
- O'CONNELL, P., G. M. LATHROP, M. LAW, M. LEPPERT, Y. NAKAMURA, M. HOFF, E. KUMLIN, W. THOMAS, T. ELSNER, L. BALLARD, P. GOODMAN, E. AZEN, J. E. SADLER, G. Y. LAI, J.-M. LALOUEL and R. WHITE, 1987 A primary genetic linkage map for human chromosome 12. *Genomics* **1** (in press).
- NAGYLAKI, T., and T. D. PETES, 1982 Intrachromosomal gene conversion and the maintenance of sequence homology among repeated genes. *Genetics* **100**: 315-337.
- PETES, T. D., 1980 Unequal meiotic recombination within tandem arrays of yeast ribosomal DNA genes. *Cell* **19**: 765-774.
- PONCZ, M., D. SOLOWIEJCZYK, B. HARPEL, Y. MORY, E. SCHWARTZ and S. SURREY, 1982 Construction of human gene libraries from small amounts of peripheral blood: analysis of β -like globin genes. *Hemoglobin* **6**: 27-36.
- ROEDER, G. S., M. SMITH and E. J. LAMBIE, 1984 Intrachromosomal movement of genetically marked *Saccharomyces cerevisiae* transposons by gene conversion. *Mol. Cell. Biol.* **4**: 703-711.
- RUBNITZ, J., and S. SUBRAMANI, 1984 The minimum amount of homology required for homologous recombination in monkey cells. *Mol. Cell. Biol.* **4**: 2253-2258.
- SCHERER, S., and R. W. DAVIS, 1980 Recombination of dispersed repeated DNA sequences in yeast. *Science* **209**: 1380-1384.
- SHEN, P., and H. V. HUANG, 1986 Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* **112**: 441-457.
- SHIN, H.-S., T. A. BARGIELLO, B. T. CLARK, F. R. JACKSON and M. W. YOUNG, 1985 An unusual coding sequence from a Drosophila clock gene is conserved in vertebrates. *Nature* **317**: 445-448.

PRB3L

1	agtgctcttattctgactcctccttagactgagagccctcaactgcttctcttcttccagcattccactccagagttc	80
81	tagagcttcaactgaaaatgcaagaatagtgctgggtcttattttgtgcatctccatttagctgcttactgta	160
161	aaaatttgcgcaactattcagtgaaatgctgctccccacctctccagGAAAGCCAGAAGGACGACGCCCAAGG	240
241	AGGAAACAGCCCAACGTACCCACCTCTCTAGGAAAGCCAGAAGGACGACCCCAAGGAGGCAACAGTCCCAAG	320
321	GTCCCCACCTCGTCCAGGAAAGCCAGAAGGACACCCCAAGGAGGAAACAGTCCCAAGGTCCTCCACCTCGTCCG	400
401	GGAAAGCCAGAAGGACACCCCAAGGAGGAAACAGTCCCAAGGTCCTCCACCTCGTCCGGAAAGCCAGAAGGACC	480
481	ACCCCAAGGAGGAAACAGTCCCAAGGTCCTCCGCTCGTCCGGAGAGCCAGAAGGACACCCCAAGGAGGAA	560
561	ACAGTCCCAAGGTCCTCCGCTCATCCGGAAAGCCAGAAGGACACCCCAAGGAGGAAACAGTCCCAAGGTCCTC	640
641	CCACCTCGTCCGGAAAGCCAGAAGGACACCCCAAGGAGGAAACAGTCCCAAGGTCCTCCACCTCGTCCGGAA	720
721	GCCAGAAGGACACCCCAAGGAGGAAACAGTCCCAAGGTCCTCCACCTCGTCCGGAAAGCCAGAAGGACACCC	800
801	CACAAGGAGGAAACAGTCCCAAGGTCCTCCACCTCGTCCAGGAAAGCCAGAAGGATCACCTTCAAGGAGGCAACAA	880
881	CCTCGAGGTCCTCCACCTCATCCAGGAAAGCCAGAAGGACACCCCAAGGAGGCAACAACTCAACGTCCCTCTC	960
961	TCCAAGAGGCCCAAGGACACCCCAAGGAGGCAATCCCAAGGAGGCAATCCCAAGGAGGTCCTCCGCTGGAAAGCC	1040
1041	GACCACCTCCACCTCCTCAAGGGGACAGACCACAGACCTCCCAAGGACAGCCTCCCAAGTAACTAAGGTTCAATGAC	1120
1121	AGGTATGA	1128

PRB4S

1	cttctagactgggctatttctctccttagactgagagccctcaacttctccttttccccagcgttccactccaga	80
81	gttctagggcttcaactgaaaatgcaagaatagtgatctgggtctcattttgtgcatctccatttagctccattac	160
161	tgtaaaatttggtgcaactattcagtgaaatgctgctccccacctctccagGAAAGCCAGAAGGACGACGCCAC	240
241	AAGGAGGAAACAGCCCAACGTCCCCACCTCTCCAGGAAAGCCACAAGGACACCCCAAGGAGGAAACAGTCC	320
321	CAAGTCCCCACCTCTCCAGGAAAGCCAGAAGGACGACCCCAAGGAGGCAACAGTCCCAAGGTCCTCCACCTCA	400
401	TCCAGGAAAGCCAGAAGGACACCCCAAGGAGGAAACAGTCCCAAGGTCCTCCACCTCTCCAGGAAAGCCAGA	480
481	GACACCCCAAGGAGGCAACAGTCCCAAGGTCCTCCACCTCTCCAGGAAAGCCAGAAGGACACCCCAAGGA	560
561	GGTAACAGTCCCAAGGTCCTCCACCTCATCCAGGAAAGCCAGAAGGACACCCCAAGGAGGAAACAGTCCCGAAG	640
641	TGCCCCGATCTCTCCAGGAAAGCCACAAGGACACCCCAAGGAGGCAACAGGTCCTCAAGGTCCTCCACCTCTCG	720
721	AGCCACAAGGCCACCCCAAGGAGGCAATCCCAAGGAGGCAATCCCAAGGAGGTCCTCCGCTGGAAAGCCCAAGG	800
801	CCACCTCTCAAGGGGACAGCCACAGACCTGCCAGGACAAACAGCTCCCAAGTAACTAG	865

PRB4M

1	cttctagactgggctatttctctccttagactgagagccctcaacttctccttttccccagcgttccactccaga	80
81	gttctagggcttcaactgaaaatgcaagaatagtgatctgggtctcattttgtgcatctccatttagctccattac	160
161	tgtaaaatttggtgcaactattcagtgaaatgctgctccccacctctccagGAAAGCCAGAAGGACGACGCCAC	240
241	AAGGAGGAAACAGCCCAACGTCCCCACCTCTCCAGGAAAGCCACAAGGACACCCCAAGGAGGAAACAGTCC	320
321	CAAGTCCCCACCTCTCCAGGAAAGCCAGAAGGACGACCCCAAGGAGGCAACAGTCCCAAGGTCCTCCACCTCA	400
401	TCCAGGAAAGCCAGAAGGACACCCCAAGGAGGAAACAGTCCCAAGGTCCTCCACCTCTCCAGGAAAGCCAGA	480
481	GACGACCCCAAGGAGGCAACAGTCCCAAGGTCCTCCACCTCATCCAGGAAAGCCAGAAGGACACCCCAAGGA	560
561	GGCAACAGTCCCAAGGTCCTCCACCTCTCCAGGAAAGCCAGAAGGACACCCCAAGGAGGCAACAGTCCCAAGG	640
641	TCCCCACCTCATCCAGGAAAGCCAGAAGGACACCCCAAGGAGGAAACAGTCCCGAAGTCCCGATCTCTCCAG	720
721	GAAAGCCACAAGGACACCCCAAGGAGGCAACAGGTCCTCAAGGTCCTCCACCTCTCGGAAAGCCACAAGGCCA	800
801	CCACAGGAGGCAATCCCAAGGAGGCAATCCCAAGGAGGCAATCCCAAGGAGGTCCTCCGCTGGAAAGCCCAAGG	880
881	CAGGCCACAGACTGCCAGGACAAACAGCTCCCAAGTAACTAG	928

PRB4L

1	cttctagactgggctatttctctccttagactgagagccctcaacttctccttttccccagcgttccactccaga	80
81	gttctagggcttcaactgaaaatgcaagaatagtgatctgggtctcattttgtgcatctccatttagctccattac	160
161	tgtaaaatttggtgcaactattcagtgaaatgctgctccccacctctccagGAAAGCCACAAGGACGACGCCAC	240
241	AAGGAGGAAACAGCCCAACGTCCCCACCTCTCCAGGAAAGCCACAAGGACACCCCAAGGAGGAAACAGTCC	320
321	CAAGGTCCTCCACCTCTCCAGGAAAGCCAGAAGGACGACCCCAAGGAGGCAACAGTCCCAAGGTCCTCCACCTCA	400
401	TCCAGGAAAGCCAGAAGGACACCCCAAGGAGGAAACAGTCCCAAGGTCCTCCACCTCTCCAGGAAAGCCAGA	480
481	GCCGACCCCAAGGAGGACACAGTCCCAAGGTCCTCCACCTACTCCAGGAAAGCCGGAAGGACACCCCAAGGA	560
561	GGAAACAGTCCCAAGGTCCTCCACCTCTCCAGGAAAGCCAGAAGGACACCCCAAGGAGGCAACAGTCCCAAGG	640
641	TCCCCACCTCATCCAGGAAAGCCAGAAGGACACCCCAAGGAGGCAACAGTCCCAAGGTCCTCCACCTCTCCAG	720
721	GAAAGCCAGAAGGACACCCCAAGGAGGCAACAGTCCCAAGGTCCTCCACCTCATCCAGGAAAGCCAGAAGGACCA	800
801	CCCCACAGGAAAGGAAACAGTCCCGAAGTCCCGATCTCTCCAGGAAAGCCACAAGGACACCCCAAGGAGGCA	880
881	CAAGCTCAAGGTCCTCCACCTCTCGGAAAGCCACAAGGCCACCCCAAGGAGGCAATCCCAAGGAGGTCCTCAGG	960
961	CTCTGCTGGAAAGCCCAAGGGCCACCTCCACCTCTCAAGGGGACAGGCCACCAAGACTGCCAGGACAAACAGCT	1040
1041	CCCCAGTAACTAG	1054

FIG. 7—(Continued)