Am. J. Hum. Genet. 50:842–851, 1992

# *PRB2/1* Fusion Gene: A Product of Unequal and Homologous Crossing-over between Proline-rich Protein (PRP) Genes *PRB1* and *PRB2*

E. A. Azen,* P. O'Connell,† and H.-S. Kim‡

*Departments of Medicine and Medical Genetics, University of Wisconsin, Madison; †Howard Hughes Medical Institute, University of Utah Medical Center, Salt Lake City; and ‡Department of Pathology, University of North Carolina, Chapel Hill

## Summary

**The *PRB2/1* fusion gene is produced by homologous and unequal crossing-over between *PRB1* and *PRB2* genes that code for basic salivary proline-rich proteins (PRPs). To determine the molecular basis for the *PRB2/1* fusion gene, the DNA sequence was determined for the *PRB2/1* gene and was compared with those of the *PRB1* and *PRB2* genes. From these comparisons, the crossing-over is postulated to occur in a 743-bp region of identity, with only 1-bp mismatch between the *PRB1* and *PRB2* genes, in the third intron outside the coding region of the two genes. This region of virtual complete identity is the largest found between any of the six closely linked *PRB* genes and may facilitate recombination. Since the coding region of *PRB1* is completely absent from the *PRB2/1* gene, salivas from two white *PRB2/1* homozygotes were studied to determine which polymorphic PRPs were missing from the salivas. Polymorphic PRPs Pe, PmF, PmS, and Ps were found to be missing from the salivas. However, a white individual lacking the same salivary PRPs is a *PRB2/1* heterozygote with one *PRB1* allele. The explanation for the missing salivary proteins in this individual is unknown. The *PRB2/1* gene is relatively frequent in several populations of unrelated individuals, including American blacks ($n = 41$), American Utah whites ($n = 76$), and mainland Chinese ($n = 131$), with gene frequencies of .22, .06, and .09, respectively. Evidence for the occurrence of *PRB1/2* heterozygotes is also presented.**

## Introduction

The human salivary proline-rich proteins (PRPs) are determined by six closely linked genes on chromosome 12p13.2, genes that can be divided into two major families on the basis of the sensitivities of the DNA sequences to the restriction enzymes *Hae*III and *Bst*NI (Maeda 1985; Maeda et al. 1985; Azen and Maeda 1988). The *Hae*III-type genes, *PRH1* and *PRH2*, code for acidic PRPs (Maeda 1985; Maeda et al. 1985; Kim and Maeda 1986; Azen et al. 1987). Four *Bst*NI-type genes can be subdivided into two subgroups: *PRB1* and *PRB2*, which code for basic PRPs, and *PRB3* and *PRB4*, which code for heavily glycosylated basic PRPs (Maeda 1985; Maeda et al. 1985; Lyons et al. 1988*a*; Azen et al. 1990). Several salivary proteins have been assigned to the different *PRP* genes (Lyons et al. 1988*a*; Minaguchi and Bennick 1989; Kauffman et al. 1991). Multiple PRPs can be produced by a single *PRP* gene through mechanisms of allelic variations, posttranslational proteolytic cleavages, and differential RNA processing (Maeda et al. 1985). The six *PRP* genes have been physically linked, and the *PRP* gene cluster spans at least 700 kbp, with the most likely order from 5′ to 3′ being *PRB2*, *PRB1*, *PRB4*, *PRH2*, *PRB3*, and *PRH1* (Kim et al. 1990). Frequent intragenic homologous and unequal crossovers within the tandemly repeated sequences of the third exon of *PRP* genes result in frequent-length polymorphisms (Azen et al. 1984; Lyons et al. 1988*b*).

Evidence of presumed intergenic crossing-over between *PRB1* and *PRB2* was first found among mem-

Received July 10, 1991; final revision received December 6, 1991.

Address for correspondence and reprints: Edwin A. Azen, M.D., Departments of Medicine and Medical Genetics, 2435 Medical Sciences Center, University of Wisconsin, 1300 University Avenue, Madison, WI 53706.


0002-9297/92/5004-0021$02.00

bers of Utah reference pedigrees (O'Connell et al. 1987). From genomic Southern analysis, Lyons et al. (1988*b*) determined that one individual was apparently a homozygote for a fusion gene. Comparison of the restriction map of this fusion gene with the restriction maps of the *PRB*1 and *PRB*2 genes showed that the 5′ region of the fusion gene is identical to the 5′ region of *PRB*2 and includes the third exon of *PRB*2. The restriction map of the region 3′ to exon 3 in the fusion gene is identical to that of *PRB*1. Therefore, this fusion gene was named *PRB*2/1.

To determine the molecular basis for the presumed intergenic crossing-over event, we cloned, sequenced, and compared the *PRB*2/1 gene from the same individual (8136) studied by Lyons et al. (1988*b*) with the *PRB*1 and *PRB*2 genes that were previously sequenced (H.-S. Kim, unpublished data). In order to assign specific protein products to the *PRB*1 gene, we studied salivas from two *PRB*2/1 homozygotes to determine which polymorphic PRPs were missing from the salivas, since the coding region of *PRB*1 is absent from the *PRB*2/1 gene. We also determined the approximate gene frequencies of *PRB*2/1 in American black, American Utah white, and mainland Chinese populations. Evidence for the occurrence of *PRB*1/2 heterozygotes is also presented.

## Material and Methods

### *Cloning and Sequencing the* PRB2/1 *Fusion Gene*

The *PRB*2/1 gene from the white *PRB*2/1 homozygote (8136) was cloned as a 17-kbp *Hin*dIII fragment from a size-fractionated library in lambda bacteriophage Charon 40 (Dunn and Blattner 1987). The *PRB*1 and *PRB*2 genes were previously (H.-S. Kim, unpublished data) cloned as 17-kbp and 12-kbp *Hin*dIII fragments, respectively, into bacteriophage Charon 35 (Loenen and Blattner 1983). Recombinant bacteriophage clones were detected by hybridization to the *Hin*fI 980 PRP probe (from the tandemly repeated sequences of exon 3 of *PRB*1 [Azen et al. 1984]), which hybridizes to exon 3 regions of all six *PRP* genes (Maeda 1985). The *Bam*HI/*Hin*dIII fragments containing the entire *PRB*1, *PRB*2, and *PRB*2/1 genes were then subcloned into the plasmid pUC18 (Messing 1983) (plasmid clones are bracketed in fig. 1*b*) for subsequent DNA sequencing by the method of Maxam and Gilbert (1977). For the three genes, 100% of both strands was sequenced, and sequencing was done from multiple sites by using different restriction endonucleases, so that sequences overlapped all cleavage sites. The sequence of the *PRB*2/1 gene was determined from the 5′ *Bam*HI site to the *Eco*RI site just 5′ to exon 4, as shown in figure 1*b* (striped bar). The sequences of the *PRB*1 and *PRB*2 genes were previously determined (H.-S. Kim, unpublished data) from the 5′ *Bam*HI site to positions just 3′ to exon 4, as shown in figure 1*b* (striped bars). Sequence data were analyzed using software provided by the University of Wisconsin Genetics Computer Group (Devereux et al. 1984).

### *Southern Blot Analysis of Genomic DNAs for PRP Genes*

DNA samples isolated from peripheral blood leukocytes were completely digested with *Eco*RI, *Bam*HI,

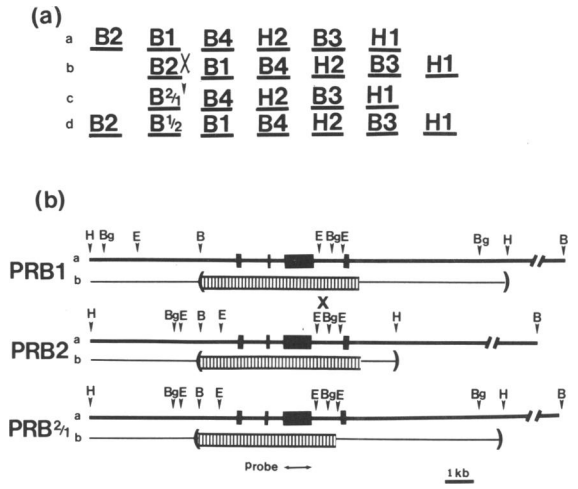


**Figure 1** Postulated crossing-over event in the PRP gene complex, and restriction maps of *PRB*1, *PRB*2, and *PRB*2/1 genes. (*a*), Postulated crossing-over event in the *PRP* gene complex, leading to *PRB*2/1 and *PRB*1/2 fusion genes, shown schematically. B1, B2, B3, and B4 represent *PRB*1, *PRB*2, *PRB*3, and *PRB*4 genes, respectively. H1 and H2 represent *PRH*1 and *PRH*2 genes, respectively. Lines a and b represent the crossing-over between DNA strands in the region of the *PRP* gene complex. Lines c and d show the *PRB*2/1 and *PRB*1/2 fusion gene chromosomes, respectively. (*b*), Restriction maps, cloning, and sequencing strategy for *PRB*1, *PRB*2, and *PRB*2/1 genes. The postulated crossing-over region is indicated by "X". The *Hin*fI 980 PRP probe is shown below the *PRB*2/1 map. Lines a show restriction maps. H = *Hin*dIII; Bg = *Bgl*II; E = *Eco*RI; and B = *Bam*HI. The broken lines between the 3′ *Hin*dIII and 3′ *Bam*HI sites indicate that this region has not been mapped. Also, in order to facilitate presentation of the map, the length of this region, although known, is not in scale. From genomic mapping experiments, the hybridizing *PRB*1 and *PRB*2 *Bam*HI fragments were determined to be 22 kbp and 15 kbp, respectively. Exons are denoted by blackened boxes. Lines b show phage clones (thin lines) spanning the *PRP* genes and plasmid clones (in parentheses). Striped bars show the regions sequenced.

and/or *Hin*dIII, and the Southern blots were hybridized to the *Hin*fI 980 PRP probe according to methods described elsewhere (Vanin et al. 1983; Azen et al. 1984). Three populations of unrelated individuals were studied. These included 76 American whites (Utah samples), 131 Chinese (mainland China), and 41 American blacks.

### *Electrophoretic Analysis of Salivary PRP Polymorphisms*

Parotid salivary proteins were electrophoresed in several gel systems and were typed for PRP polymorphisms as described elsewhere (methods are summarized by Azen 1989). In brief, acidic PRPs were typed in isoelectric focusing gels, pH 3.5–5.2 (Azen and Denniston 1981); basic and glycosylated PRPs were typed in 10% SDS and acid/Tris-citrate polyacrylamide gels (Azen and Yu 1984*a*, 1984*b*; Azen 1989) and acid/lactate polyacrylamide gels, pH 2.4 (Azen et al. 1979; Azen and Denniston 1980).

## Results

### *Molecular Analysis of the PRB2/1 Fusion Gene*

The postulated crossing-over event between *PRB2* and *PRB1* genes in the *PRP* gene complex leading to the *PRB2/1* and *PRB1/2* fusion genes is schematically shown (fig. 1*a*). The *PRB2/1* chromosome has deletions of both the normal *PRB1* and *PRB2* genes (fig. 1*a*, line c), whereas both these genes plus the *PRB1/2* fusion gene are present on the *PRB1/2* chromosome (fig. 1*a*, line d). The comparison of the DNA sequences of *PRB1* and *PRB2* genes (H.-S. Kim, unpublished data) showed several interesting features (schematically shown in fig. 2). Exons 1 and 2 of *PRB1* and *PRB2* genes are identical, and the 500-bp region upstream from exon 1 differs by a single base. However, exon 3 of *PRB1* and *PRB2* genes shows 21 (2.3%) nucleotide differences among 913 bp compared. Intron 3 of the two genes shows the largest region of homology wherein the crossover is postulated to occur (homologous region is indicated by interrupted horizontal lines in fig. 2). Within this 743-bp homologous region there is only 1-bp mismatch.

Figure 3 shows a comparison of the *PRB2/1* fusion gene and *PRB1* and *PRB2* genes in the region immediately surrounding and including the postulated crossing-over. The 5′ region of *PRB2/1* is identical to that of *PRB2* (with the exception of a single nt difference at *PRB2* nucleotide position (nt) 4664, indicated by an asterisk) until nt 4964. The DNA sequences of *PRB1*, *PRB2*, and *PRB2/1* genes are identical (with the exception of the 1-bp mismatch mentioned above) from *PRB2* nt 4221 to *PRB2* nt 4963 (this region is boxed in fig. 3). The DNA sequence of *PRB2/1* further 3′ to *PRB2* nt 4963 is identical to that of *PRB1*. Since the fusion gene shows the same "C" nucleotide as *PRB1* (and not the "T" nucleotide of *PRB2*) at position 4664, the crossing-over most likely occurred in *PRB2* nt 4221–4663. However, it could have occurred in *PRB2* nt 4664–4963, if the single-nucleotide difference at position 4664 arose as a mutation after the crossing-over event.

### *The PRB2/1 Fusion Gene Is Common in Three Different Populations*

It is normal for the six *PRP* genes to be represented by hybridizing bands at six different locations, after a Southern blot analysis of *Eco*RI-, *Hin*dIII-, and *Bam*HI-digested DNAs by the *Hin*fI 980 PRP probe (Maeda 1985). The bands at these different locations are often seen as closely spaced doublets that represent heterozygosity for frequent length polymorphisms (see examples in figs. 4 and 5). From the diagram of the crossover between *PRB1* and *PRB2* genes (fig. 1*a*), as

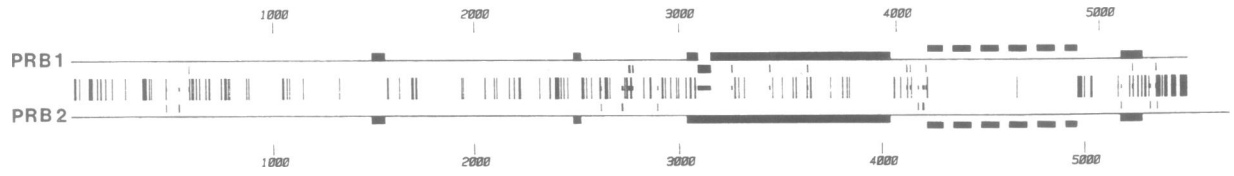


**Figure 2** Diagrammatic comparison of aligned *PRB1* and *PRB2* genes, to emphasize differences and identities. The nucleotide sequences of the two aligned genes are represented by horizontal lines, and their exons are shown as black bars (H.-S. Kim et al., unpublished data). Nucleotide differences between the genes are denoted by long vertical lines between the horizontal lines. Length differences are denoted by short vertical lines between the two horizontal lines, as well as by second small vertical lines close to the genes. The gap in exon 3 of *PRB1* represents a length difference compared with *PRB2*. The postulated 743-bp region of crossing-over in intron 3 (also the largest region of identity), which results in the *PRB2/1* and *PRB1/2* genes, is indicated by horizontal dashed lines.

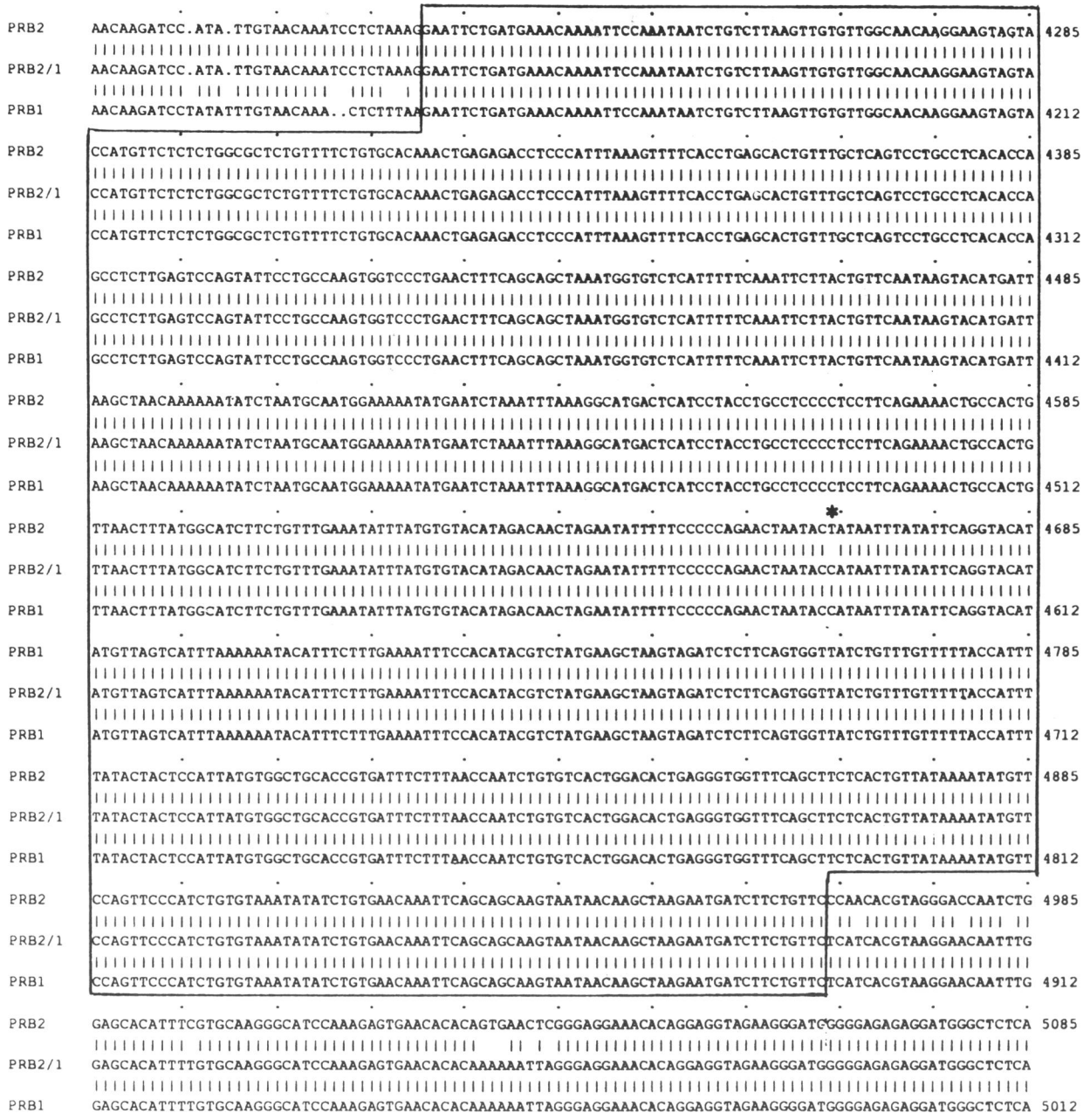


**Figure 3** DNA sequence comparisons of *PRB1*, *PRB2*, and *PRB2/1* genes. DNA sequences of aligned *PRB1*, *PRB2*, and *PRB2/1* genes in the region of the postulated crossing-over in intron 3 of *PRB1* and *PRB2* genes are shown. The sequence and numbering of nucleotides for *PRB1* and *PRB2* genes will be reported elsewhere (H.-S. Kim, unpublished data). The 743-bp region of identity of the three genes (with 1-bp mismatch, shown by an asterisk), in which crossing-over is postulated to occur, is boxed. The DNA sequence of the *PRB2/1* gene 5′ to this region is identical to that of the *PRB2* gene, but, 3′ to this region, it is identical to that of the *PRB1* gene. Identical nucleotides are tied by vertical lines.

well as from the restriction map comparing *PRB1*, *PRB2*, and *PRB2/1* genes (fig. 1*b*), the following characterizes the Southern blot analysis of *PRB2/1* gene homozygotes: in the *Eco*RI digest, the 6.5-kbp *PRB1* band is missing, and the *PRB2/1* band migrates to the same position as the 4.0-kbp *PRB2* band. In the *Hin*dIII and *Bam*HI digests, the 12-kbp and 15-kbp *PRB2* bands are missing, and the *PRB2/1* band mi-

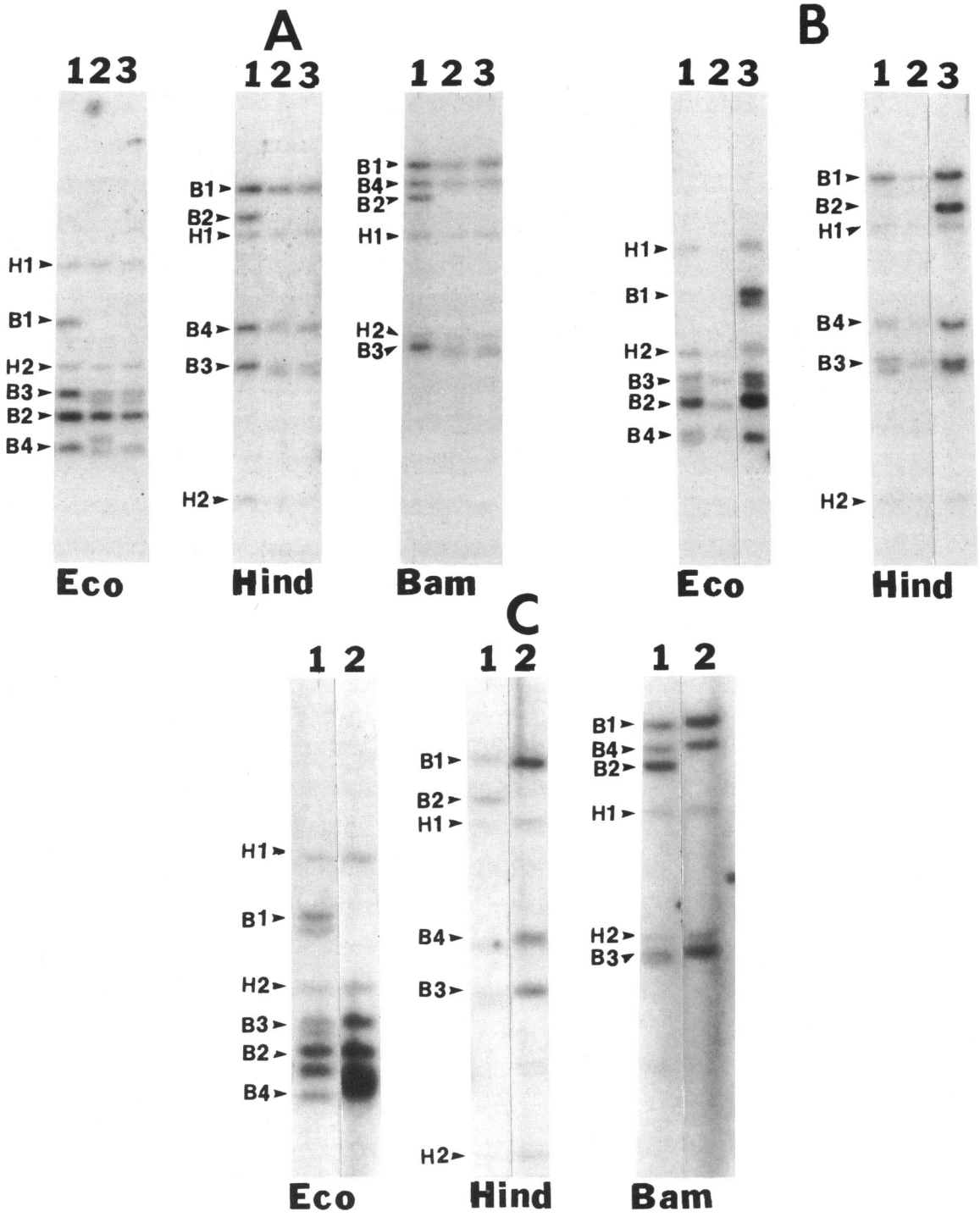


**Figure 4** Examples of *PRB2/1* types in white, black, and Chinese populations. Southern blots of endonuclease-digested genomic DNAs are hybridized to the *Hin*fI-980 PRP probe. *A*, White Utah *PRB2/1* types. Lane 1, *PRB2/1* heterozygote with one *PRB1* allele. Lane 2, *PRB2/1* homozygote (8136). Lane 3, *PRB2/1* homozygote (8483). *B*, Black *PRB2/1* homozygotes. Lane 1, *PRP2/1* homozygote (51). Lane 2, *PRB2/1* homozygote (11). Lane 3, "Normal" *PRB1* heterozygote with two different-length *PRB1* alleles (best seen in the *Eco*RI digest). *C*, Chinese *PRB2/1* homozygote. Lane 1, "Normal" *PRB1* heterozygote with two different-length *PRB1* alleles. Lane 2, *PRB2/1* homozygote (89–30). In *PRB2/1* homozygotes there is an absence of the *PRB1* band in *Eco*RI digests and the *PRB2/1* band migrates to the same position as does *PRB2*. In *Bam*HI and *Hin*dIII digests, the *PRB2* band is missing and the *PRB2/1* band migrates to the same position as does *PRB1*. Eco = *Eco*RI; Hind = *Hin*dIII; and Bam = *Bam*HI. Other symbols are as defined in fig. 1. Closely spaced doublet bands, best seen in the *Eco*RI digests, represent heterozygous length polymorphisms.

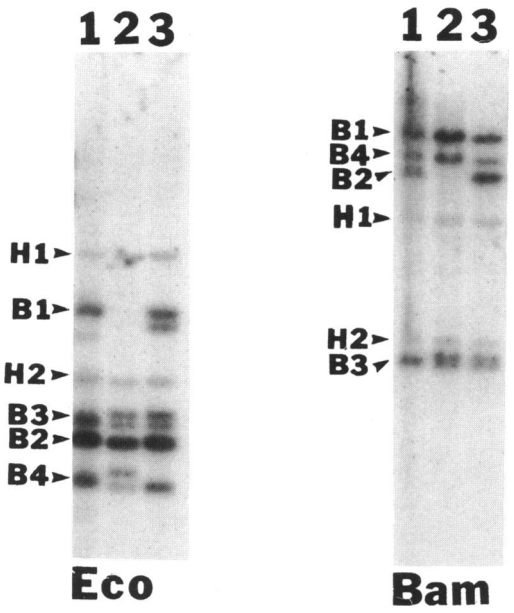


**Figure 5** Southern blot analysis of endonuclease-digested genomic DNA of a Madison white (MVO) who is a *PRB2/1* heterozygote. The Southern blot is hybridized to the *Hin*fI 980 PRP probe. Lane 1, MVO who is a *PRB2/1* heterozygote with one *PRB1* allele. Lane 2, White *PRB2/1* homozygote (8136). Lane 3, "Normal" *PRB1* heterozygote with two different length *PRB1* alleles. Closely spaced doublet bands, as seen at B1, B3, and B4 positions (best seen in the *Eco*RI digest), represent heterozygous length polymorphisms. Symbols are as defined in figs. 1 and 4.

grates to the same position as the 17-kbp and 22-kbp *PRB1* bands. Thus, *PRB2/1* homozygotes are scored by the absence of the normal *PRB1* band in *Eco*RI digests and are confirmed by the absence of the normal *PRB2* band in *Hin*dIII and/or *Bam*HI digests.

Examples of Southern blot analysis of all of the *PRB2/1* homozygotes found in the populations studied are shown in figure 4 (two whites, lanes *A2* and *A3*; two blacks, lanes *B1* and *B2*; and one Chinese, lane *C2*). An example of a known *PRB2/1* heterozygote (from the white Utah population) with one *PRB1* allele is shown in figure 4, lane *A1*, and examples of a "normal" *PRB1* heterozygote with two different-length *PRB1* alleles (best seen in the *Eco*RI digest) are shown in figure 4, lanes *B3* and *C1*.

*PRB2/1* gene frequencies in American black and mainland Chinese populations were estimated from the number of *PRB2/1* homozygotes. Among 41 blacks there were 2 *PRB2/1* homozygotes, and the *PRB2/1* gene frequency was estimated to be approximately .22. Among 131 Chinese, there was 1 homozygote, and the *PRB2/1* gene frequency was estimated to be approximately .09. A more accurate estimate of gene frequency for *PRB2/1* in the American Utah whites was obtained from the number of *PRB2/1* homozygotes (two) plus heterozygotes (five). After Southern blot analysis of *Eco*RI-digested DNAs, the five *PRB2/1* heterozygote samples from the American Utah whites were primarily ascertained by abnormal segregation of their *PRB1* polymorphisms in family studies. The *PRB2/1* heterozygote genotypes were then confirmed by densitometric analysis of the autoradiograms from the above Southern blots, where a single band (representing one *PRB1* allele) of approximately one-half normal intensity was seen at the *PRB1* position. In a typical family, the father shows an apparent type 1,1 (arbitrary designation for two *PRB1* alleles of the same size), and the mother is a type 1,2 (*PRB1* alleles of different sizes). However, one of the children from the mating is an apparent type 2,2. After nonpaternity is excluded by testing the samples with many other polymorphic markers, the father's correct *PRB1* type is 1,0 and the child's type is 2,0. This result indicates the segregation of a "null" in these families. This abnormal pattern of segregation was seen in five families, and one *PRB2/1* heterozygote was chosen from each family for inclusion in the population analysis.

Densitometric scans were performed on these families, and the suspected *PRB2/1* heterozygotes (with one *PRB1* allele) were compared with a "normal" *PRB1* heterozygote with two *PRB1* alleles at the *PRB1* position. After normalization to the *PRB2* gene signal as an internal loading standard, the ratio of the signal at the *PRB1* position of the "normal" *PRB1* heterozygote to each of the five *PRB2/1* heterozygotes used in the population study was 1.9, 1.8, 2.4, 1.9, and 2.4, respectively, for the five comparisons. This result supports the assigned *PRB2/1* heterozygote genotype for the five samples. The stringent method for ascertaining the *PRB2/1* heterozygotes probably leads to a conservative estimate for the *PRB2/1* gene frequency, since some *PRB2/1* heterozygotes may have been missed if informative family data were not available for some samples. Among the 76 whites, there were 2 *PRB2/1* homozygotes and 5 *PRB2/1* heterozygotes, with a *PRB2/1* gene frequency estimated to be approximately .06.

### *Evidence for the Occurrence of* PRB1/2 *Heterozygotes*

After Southern blot analysis of *Eco*RI-digested DNAs, a *PRB1/2* homozygote would be expected to show four gene copy signals (two from *PRB1* and two

from *PRB*1/2) at the *PRB*1 position (see fig. 1*a* and *b*), and the signal may be distributed over one or more polymorphic bands. We have not recognized the *PRB*1/2 homozygote genotype in the Utah population. However, if the number of *PRB*1/2 homozygotes is expected to be approximately equal to the number of *PRB*2/1 homozygotes in the same population (2/76), then failure to see the *PRB*1/2 homozygotes (0/76) may have occurred by chance.

However, we have evidence for the occurrence of the *PRB*1/2 heterozygote type. This evidence includes Southern blot analysis and densitometry of families from the Utah population and restriction analysis of a cloned candidate *PRB*1/2 gene from a Madison white individual. After Southern blot analysis of *Eco*RI-digested DNAs, the *PRB*1/2 heterozygote genotype would be expected to show three gene signals (two *PRB*1 and one *PRB*1/2) at the *PRB*1 position (see fig. 1*a* and *b*), which may be distributed over one, two, or three polymorphic bands. Three families showed individuals with three gene signals at the *PRB*1 position. As an example, in one family, the mother's polymorphic *PRB*1 type is 1,2-3 and the father is type 0,2. Southern blot analysis and densitometry show the *PRB*1 pattern of a child to be a type 2,2-3 (three gene copies) with a signal 2.6 times the intensity of the father's *PRB*1-type signal (a *PRB*2/1 heterozygote with one gene copy). Further evidence for the occurrence of the *PRB*1/2 gene was obtained from study of a Madison white. After Southern blot analysis of the *Eco*RI-digested DNA, this individual was found to possess a *PRB*1 length polymorphism with three *PRB*1-type gene copies distributed over two bands (two copies at one band and one copy at the other band), compared with a "normal" *PRB*1 heterozygote with two different-length *PRB*1-type copies (one copy at each band). Densitometric analysis of bands at the *PRB*1 position gave a signal intensity ratio, of the presumed *PRB*1/2 heterozygote (three gene copies)/"normal" *PRB*1 heterozygote (two gene copies), of 1.6. Restriction mapping, with *Eco*RI and *Hin*dIII, of a cloned *PRB*1-type gene from this individual showed a pattern characteristic of the *PRB*1/2 gene (representing one of the three *PRB*1-type gene copies seen on the Southern analysis).

### *Pe, Pm, and Ps Polymorphic PRPs Are Absent from Salivas of Two Utah White* PRB2/1 *Homozygotes and a Madison White (MVO) Who Is Probably* PRB2/1 *Heterozygotic*

Figure 6 shows a protein-stained Western blot (from an SDS gel) of the salivas of four individuals. The Pe, PmS, and Ps polymorphic PRPs (coded by *PRB*1) are missing from the salivas of two Utah white *PRB*2/1 homozygotes, 8136 and 8483 (fig. 6, lanes 2 and 3, respectively). The PmF protein (coded by *PRB*1) is also missing (not shown). Other salivary PRPs that are present in all the samples include basic PRPs (Po, IB-4, IB-1, and Pc), the glycosylated PRP (Gl), and some acidic PRPs (Pr, PIF, and Db). The basic PRP IB-7 is present in all the samples (not shown). Amylase is also present. Salivas from the above two *PRB*2/1 homozygotes were typed for other salivary polymorphic PRPs (not shown) and were both found to be Pr 1-1 (coded

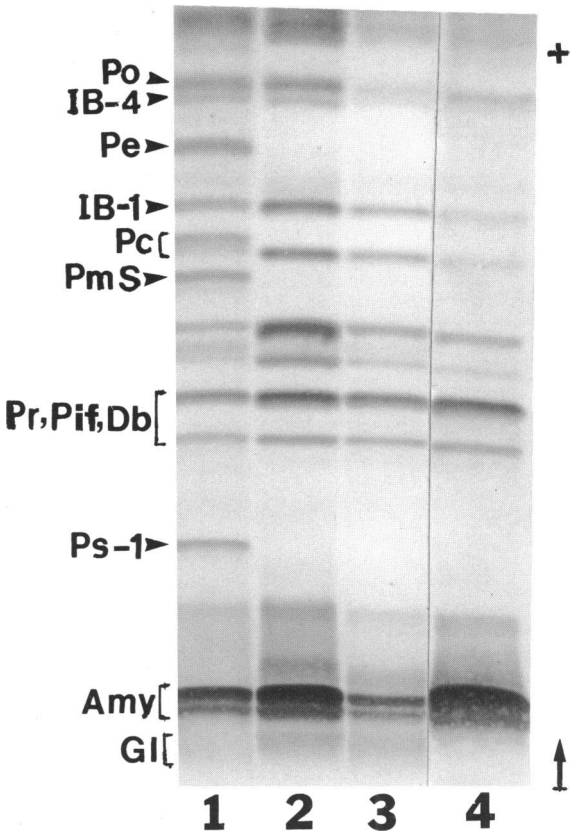


**Figure 6** Salivary PRPs of two Utah white *PRB*2/1 homozygotes and MVO who is a *PRB*2/1 heterozygote. A protein-stained Western blot of an SDS gel is shown. Lane 1, Control, Pe +, PmS +, and Ps 1. The saliva is also PmF + (not shown). Lane 2, White *PRB*2/1 homozygote (8483), Pe-, PmS-, PsO. Lane 3, White *PRB*2/1 homozygote (8136), Pe-, PmS-, PsO. Lane 4, MVO who is probably a *PRB*2/1 heterozygote, Pe-, PmS-, PsO. PmF is also missing from salivas of the two *PRB*2/1 homozygotes and MVO (not shown). Other salivary PRPs shown that are present in all the samples include basic PRPs (Po, IB-4, IB-1, and Pc), the glycosylated PRP (Gl), and some acidic PRPs (Pr, PIF, and Db). Amylase (Amy) is also present in all the samples.

by *PRH*2), Db+ and PIF+ (both coded by *PRH*1), Gl 1-3 (coded by *PRB*3), and CON 1+ (coded by *PRB*4).

The salivary proteins of MVO are shown in figure 6, lane 4. The Pe, PmS, and Ps proteins are missing from the saliva. The PmF protein is also missing (not shown). Saliva was also typed for other salivary PRP polymorphisms (not shown) and was found to be Pr 1-1, Db+, PIF+, Gl 2-3, and CON 1+.

Figure 5 shows a Southern blot analysis with the *Hin*fI 980 PRP probe of *Eco*RI- and *Bam*HI-digested DNAs of three individuals, including MVO. In the *Eco*RI digest, densitometric analysis of the *PRB*1 hybridization signal of MVO (figure 5, lane 1), compared with that of a "normal" *PRB*1 heterozygote with two different-length *PRB*1 alleles (fig. 5, lane 3 in left panel), shows that MVO is a *PRB*2/1 heterozygote with one *PRB*1 allele, and supportive family segregation data are available. Thus, in the *Eco*RI digest, when hybridization signals at the *PRB*1 position of both individuals are normalized to those of *PRB*4 for each individual and are compared, the *PRB*1 signal of the "normal" *PRB*1 heterozygote (two *PRB*1 alleles) is 2.4 times the *PRB*1 signal of MVO (presumed to have one *PRB*1 allele). Since MVO is not a *PRB*2/1 homozygote and has one *PRB*1 allele, there must be some other explanation for the absence of Pe, PmS, PmF, and Ps proteins from the saliva of this individual.

## Discussion

Previously, Lyons et al. (1988*b*) described homologous and unequal intragenic crossing-over (facilitated by tandemly repeated sequences in exon 3 of *PRP* genes) that leads to frequent length polymorphisms. We now report homologous and unequal crossing-over between *PRB*1 and *PRB*2 genes that is probably facilitated by a 743-bp region of identity with 1-bp mismatch in the third intron of the two genes. When the DNA sequences of all six *PRP* genes are compared (Kim et al., unpublished data), this region of virtual complete identity is the largest found and may be the main factor that accounts for the occurrence of the *PRB*2/1 fusion gene. Indeed, among the three subgroups of *PRP* genes, the nucleotide sequences of the *PRB*1 and *PRB*2 genes are the most closely related and have the most recent evolutionary divergence (3.3% overall sequence differences), compared with *PRB*3 and *PRB*4 genes (9.8% overall sequence differences) and *PRH*1 and *PRH*2 genes (8.7% overall sequence differences) (Kim and Maeda 1986; H.-S. Kim, unpublished data). An additional factor that might facilitate unequal and homologous crossing-over between *PRB*1 and *PRB*2 genes is the physical ordering between *PRP* genes within the six-member gene complex. *PRB*1 and *PRB*2 genes are physically adjacent members of the *PRP* gene complex and are separated by about 60 kbp (Kim et al. 1990).

A similar mechanism of intergenic recombination occurs between other related genes, such as members of the β-like globin gene family, with the Lepore hemoglobins as an example. Metzenberg et al. (1991) presented data indicating that unequal Lepore crossovers are more likely to occur in relatively large stretches of sequence identity between δ- and β-globin genes. The consensus of many studies in mammalian cells (reviewed by Bollag et al. 1989; Metzenberg et al. 1991) is that approximately 200 bp of identity are required for efficient recombination. This condition is met for the postulated crossing-over region in the third introns between *PRB*1 and *PRB*2 genes, with identical sequences extending 443 bp 5′ and 299 bp 3′ to the 1-bp mismatch. The crossing-over most likely occurred in the 443-bp 5′ region of identity, although it could have occurred in the 3′ region if the single nucleotide difference arose as a mutation after the crossing-over event.

Kim et al. (1990) have proposed an evolutionary model for the *PRP* genes. This model includes four recombinational events, and the generation of *PRB*1 and *PRB*2 genes from their precursor may be the most recent event. To explain the 743-bp region of virtual complete identity in the third introns of *PRB*1 and *PRB*2 genes, we postulate that a gene conversion event (as proposed for the third exons of the *PRH*1 and *PRH*2 genes [Kim and Maeda 1986]) may have occurred sometime after this most recent gene duplication. Unlike the Lepore-type fusion genes, where selection due to malaria may account in part for their population frequencies (reviewed by Weatherall and Clegg [1981]), there is no apparent selection for the *PRB*2/1 fusion gene. Thus, the relatively high *PRB*2/1 gene frequencies in several populations may be mainly due to recurring crossing-over mutations facilitated by the relatively large region of identity between *PRB*1 and *PRB*2 genes.

The *PRB*2/1 homozygote is easy to recognize on genomic Southern blots, because of a loss of a *PRB*1 or *PRB*2 hybridizing band (see figs. 4 and 5). By contrast, the *PRB*1/2 homozygote would be more difficult to recognize, since it would be predicted not to show

an absent or unique band on Southern blot analysis (see fig. 1*a* and *b*). However, there is no a priori reason why the alternative crossover product (*PRB*1/2) would not exist. Although we did not find the *PRB*1/2 homozygote type, we provided evidence for the occurrence of the *PRB*1/2 heterozygote type from Southern blot analysis and densitometry in families from the Utah population and from a restriction analysis of a cloned candidate *PRB*1/2 gene from a Madison white.

The segregation of "normal" (noncrossover), *PRB*1/2 and *PRB*2/1 chromosomes in the population results in different *PRB*1 copy numbers in different individuals. Some individuals with *PRB*1 copy numbers varying from 0 to 4 might show a corresponding copy-dependent variation in production of *PRB*1-coded salivary proteins, which varies from 0 to 2 (with 1 representing the production of two expressed *PRB*1 alleles). However, this postulated correlation between *PRB*1 copy number and gene expression may not be strict in the population, since, as will be discussed, some *PRB*1 alleles (for undetermined reasons) may not be expressed.

The assignment of the numerous PRPs to their cognate genes is of continued interest. Since the crossing-over leading to the *PRB*2/1 fusion gene occurs in the third intron beyond the coding region of *PRB*1, the *PRB*2/1 gene is nonproductive for *PRB*1 products and thus, in the homozygous state, may lead to null salivary PRP phenotypes for those proteins coded by *PRB*1. This led us to determine which PRPs were missing from the salivas of *PRB*2/1 homozygotes. Lyons et al. (1988*a*) and Minaguchi and Bennick (1989) previously (on the basis of PRP and decoded DNA sequence comparisons) assigned the Pe (DEAEII-2), PmF (IB-9, P-E), and PmS (IB-6) polymorphic proteins to *PRB*1. Kauffman et al. (1991) found the IB-4 and IB-7 proteins to be encoded in *PRB*2. The work of Lyons et al. (1988*a*) suggests that the PmS protein may also be encoded in *PRB*2; however, the presence of a potential proteolytic cut site in the decoded amino acid sequence of PmS makes it unlikely that an intact PmS protein would be produced by *PRB*2. Although amino acid sequence data were not available, Lyons et al. (1988*a*) tentatively assigned the Ps protein to the *PRB*2 gene, on the basis of (*a*) an amino acid composition resembling that of either the *PRB*1 or *PRB*2 gene and (*b*) a weak reaction of the protein to the periodic acid-Schiff reagent, which may indicate the presence of carbohydrate. In this regard, the *PRB*2 gene, but not the *PRB*1 gene, contains N-linked glycosylation sites. However, when the partial amino acid sequence data from the Ps protein are compared with the decoded DNA sequences of *PRB*1 and *PRB*2 genes, the indication is that the Ps protein is coded by the *PRB*1 gene (E. A. Azen and R. Niece, unpublished data). Thus, in the homozygous state, the *PRB*2/1 gene (as predicted) may lead to PsO, Pm −, and Pe − null salivary phenotypes as shown in figure 6. Since the *PRB*2 coding region is present in the *PRB*2/1 gene, the IB-4 and IB-7 proteins coded by *PRB*2 are present as expected in *PRB*2/1 homozygotes.

However, the occurrence of the *PRB*2/1 gene may not be the only explanation for the PsO, Pm-, and Pe- salivary protein phenotypes, for the following reasons: First (assuming the *PRP* gene frequencies are similar in the two populations), we compared by $\chi^2$ analysis the proportions of homozygotes for Pe- (19/317), PsO (14/150), and PmF − (101/140) protein phenotypes in a previously studied Madison white population (Azen and Denniston 1980; Azen and Yu 1984*a*) with the smaller proportion of *PRB*2/1 homozygotes (2/76) herein reported in the Utah white population. Borderline significant or highly significant differences were found for PsO/*PRB*2/1 and PmF-/*PRB*2/1 comparisons ($P = .06$ and $< .0001$, respectively). Second, we show in this paper an example of a Madison white subject (MVO) who possesses the PsO, Pm-, and Pe- salivary PRP phenotypes (fig. 6) yet who, on the basis of Southern blot and densitometric analysis of the DNA, is a *PRB*2/1 heterozygote with one *PRB*1 allele (fig. 5). We do not know the reason for the PsO, Pm-, and Pe- phenotypes in subject MVO but speculate that there may be one or more as yet undetermined mutations in the *PRB*1 gene of this individual.

Lyons et al. (1988*a*) have suggested one mechanism to account for the frequent null alleles among basic PRP polymorphisms. Nucleotide analysis of a *PRB*1 allele coding for the PmF- and PmS- phenotypes showed an alteration in the proteolytic cleavage site of the precursor protein. The PmF and PmS proteins are normally seen when cleavage occurs, but, in the absence of cleavage, the larger precursor protein may not be resolved electrophoretically. Another mechanism for producing null alleles among basic PRPs is the recently described single-nucleotide insertion in exon 3 of the *PRB*3 gene (coding for the Gl protein), which leads to a frameshift with a premature termination codon that causes an apparent lack of gene expression (Azen et al. 1990). Thus, it seems likely that other interesting mutations will be found to explain frequent null phenotypes among salivary PRPs, especially Po − and CON − (from the *PRB*4 gene) and Pm-, Pe-, and PsO (from the *PRB*1 gene).

## Acknowledgments

We greatly appreciate the technical assistance of P. Latreille and the help of M. Leppert, R. White, B. Ogden, H. Kazazian, Jr., A. Metzenberg, L. Chih-Chuan, Q. Li, and Y.W. Kan in obtaining samples. We thank A. Metzenberg, P. Powers, and B. Storer for helpful comments. This investigation was supported by National Institutes of Health grant DEO3658-26.

## References

Azen EA (1989) Genetic protein polymorphisms of human saliva: In: Tenovuo JO (ed) Human saliva: clinical chemistry and microbiology, vol 1. CRC, Boca Raton, FL, pp 161–195

Azen EA, Denniston C (1980) Polymorphism of Ps (parotid size variant) and detection of a protein (PmS) related to the Pm (parotid middle band) system with genetic linkage of Ps and Pm to Gl, Db and Pr genetic determinants. Biochem Genet 18:483–501

——— (1981) Genetic polymorphisms of PIF (parotid isoelectric focussing variant) with linkage to the PPP (parotid proline-rich protein) gene complex. Biochem Genet 19: 475–485

Azen EA, Hurley CK, Denniston C (1979) Genetic polymorphism of the major parotid salivary glycoprotein (Gl) with linkage to the genes for Pr, Db and Pa. Biochem Genet 17: 257–279

Azen EA, Kim H-S, Goodman P, Flynn S, Maeda N (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 EA, Lyons KM, McGonigal T, Barrett NL, Clements LS, Maeda N, Vanin EF, et al (1984) Clones from the human gene complex coding for salivary proline-rich proteins. Proc Natl Acad Sci USA 81:5561–5565

Azen EA, Maeda N (1988) Molecular genetics of human salivary proteins and their polymorphisms. In: Harris H, Hirschhorn K (eds) Advances in human genetics. Plenum, New York, pp 141–199

Azen EA, Minaguchi K, Latreille P, Kim H-S (1990) Alleles at the *PRB3* locus coding for a disulfide-bonded human salivary proline-rich glycoprotein (Gl 8) and a null in an Ashkenazi Jew. Am J Hum Genet 47:686–697

Azen EA, Yu PL (1984*a*) Genetic determinants of Pe and Po salivary proteins with probable linkage of their genes to the salivary protein gene complex (SPC). Biochem Genet 22:1065–1080

——— (1984*b*) Genetic polymorphisms of CON1 and CON2 salivary proteins detected by immunologic and concanavallin A reactions on nitrocellulose with linkage of CON1 and CON2 genes to the salivary protein gene complex (SPC). Biochem Genet 22:1–19

Bollag RJ, Waldman AS, Liskay RM (1989) Homologous recombination in mammalian cells. Annu Rev Genet 23: 199–225

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Dunn IS, Blattner FR (1987) Charons 36 to 40: multi-enzyme, high capacity, recombination deficient replacement vectors with polylinkers and polystuffers. Nucleic Acids Res 15:2677–2698

Kauffman DL, Bennick A, Blum M, Keller PJ (1991) Basic proline-rich proteins from human parotid saliva: relationships of the covalent structures of ten proteins from a single individual. Biochemistry 30:3351–3356

Kim H-S, Maeda N (1986) Structure of two *Hae*III-type genes in the human salivary proline-rich protein multigene family. J Biol Chem 261:6712–6718

Kim H-S, Smithies O, Maeda N (1990) A physical map of the human salivary proline-rich protein gene cluster covers over 700 kbp of DNA. Genomics 6:260–267

Loenen WAM, Blattner FR (1983) Lambda Charon vectors (Ch 32, 33, 34 and 35) adapted for cloning in recombination-deficient hosts. Gene 26:171–179

Lyons KM, Azen EA, Goodman PA, Smithies O (1988*a*) Many protein products from a few loci: assignment of human salivary proline-rich proteins to specific loci. Genetics 120:255–265

Lyons KM, Stein JH, Smithies O (1988*b*) Length polymorphisms in human proline-rich protein genes generated by intragenic unequal crossing over. Genetics 120:267–278

Maeda N (1985) Inheritance of human salivary proline-rich proteins: a reinterpretation in terms of six loci forming two subfamilies. Biochem Genet 23:455–464

Maeda N, Kim H-S, Azen EA, Smithies O (1985) Differential RNA splicing and post translational cleavages in the proline-rich protein gene system. J Biol Chem 260: 11123–11130

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci USA 74:560–564

Messing J (1983) New M13 vectors for cloning. Methods Enzymol 101:20–78

Metzenberg AB, Wurzer G, Huisman THJ, Smithies O (1991) Homology requirements for unequal crossing over in humans. Genetics 128:143–161

Minaguchi K, Bennick A (1989) Genetics of human salivary proteins. J Dent Res 68:2–15

O'Connell P, Lathrop M, Law M, Leppert M, Nakamura Y, Hoff M, Kumlin E, et al (1987) A primary genetic linkage map for human chromosome 12. Genomics 1:93–102

Vanin EF, Henthorn PS, Kioussis D, Grossveld F, Smithies O (1983) Unexpected relationships between four large deletions in the human β-globin gene cluster. Cell 35:701–709

Weatherall DJ, Clegg JB (1981) The thalassemia syndromes, 3d ed. Blackwell Scientific, Oxford