Am. J. Hum. Genet. 53:264–278, 1993

*PRB*1 Gene Variants Coding for Length and Null Polymorphisms among Human Salivary Ps, PmF, PmS, and Pe Proline-rich Proteins (PRPs)

E. A. Azen,* P. Latreille,* and R. L. Niece†

*Departments of Medicine and Medical Genetics and †Biotechnology Center, University of Wisconsin, Madison

Summary

Six closely linked *PRP* (proline-rich protein) genes code for salivary PRPs that show frequent length and null polymorphisms. We report assignment of Ps proteins to the *PRB*1 gene, the derived primary structures of Ps 1 and Ps 2 proteins, and the molecular basis for some null alleles among *PRB*1-coded PRPs (Ps, PmF, PmS, and Pe). The derived primary structures of Ps 1 and Ps 2 proteins were determined by sequencing exon 3 of the different-length *PRB*1M (medium) and *PRB*1L (large) copies from subject C.J. with the Ps 1-2 phenotype. The *PRB*1L copy (coding for Ps 2) contained three additional tandem repeats within the Ps coding region, and the different-length Ps 1 and Ps 2 proteins can be explained on this basis. The molecular basis for the Ps 0 and the Pe− phenotypes was determined in another individual (M.V.O., a *PRB*2/1 fusion-gene heterozygote) with a single *PRB*1L copy. A premature stop mutation (CGA [Arg] →TGA [stop]) occurred at residue 61 in the Ps-coding region. The identical mutation was found in the *PRB*1L and *PRB*1/2S (small) copies of a second individual (E.A.) with reduced Pe protein and the Ps 0 phenotype. This individual is a *PRB*1/2 fusion-gene heterozygote (Azen et al. 1992) with probably three mutated *PRB*1 copies (*PRB*1L-*PRB*1L-*PRB*1/2S). DNA sequences of the postulated crossover region of the *PRB*1/2S fusion-gene copy supported the postulated crossover. The PmF− and PmS− phenotypes in the three subjects were due to both the stop mutation and the lack of suitable proteolytic cleavage sites in the *PRB*1-coded precursor proteins.

Introduction

Several important functions for the abundant salivary PRPs (proline-rich proteins) have been described, including binding to hydroxyapatite, calcium, and certain intraoral bacteria; mediating adherence of the microorganisms to the coated tooth surface; forming part of the enamel pellicle; inhibiting hydroxyapatite formation; and modifying the lubricative properties of saliva (Minaguchi and Bennick 1989). There are extensive genetic polymorphisms of salivary PRPs (Azen and Maeda 1988; Minaguchi and Bennick 1989), and it seems possible that these electrophoretic variations, especially the frequent null types, may explain, in part, individual susceptibilities to common intraoral and/or dental diseases. A molecular understanding of these genetic variations will constitute a firm scientific basis for future attempts to correlate genetically determined salivary phenotypes with clinical disease susceptibility.

Six closely linked human *PRP* genes (on chromosome 12p13.2) code for a large number of salivary PRPs. Four *PRP* genes—termed "*PRB*1," "*PRB*2," "*PRB*3," and "*PRB*4"—code for basic and glycosylated PRPs, and two *PRP* genes—termed "*PRH*1" and "*PRH*2"—code for acidic PRPs (Maeda 1985). The *PRP* genes have been physically linked (Kim et al. 1990), and their complete sequences and postulated evolution have been presented (Kim et al. 1993). Multiple PRPs can be produced from a single *PRP* gene by allelic variations, posttranslational cleavages, and differential RNA processing (Maeda et al. 1985). Frequent intragenic homologous and unequal crossovers within tandemly repeated sequences of the third exon of *PRP* genes result in frequent DNA length polymorphisms

Received January 4, 1993; revision received February 12, 1993.

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

0002-9297/93/5301-0029$02.00

(Azen et al. 1984; Lyons et al. 1988*b*). Two mouse *Prp* genes, termed "MP2" and "M14," have also been sequenced (Ann and Carlson 1985; Ann et al. 1988), and they are closely linked on distal chromosome 6 (Azen et al. 1989).

In this paper, we report studies that elucidate the molecular basis of polymorphisms among PRPs (Ps, PmF, PmS, and Pe) that are coded by the *PRB*1 gene. Earlier protein studies (Azen and Denniston 1980) identified a frequent length and null genetic polymorphism among the Ps proteins, which are basic PRPs. The different-size Ps proteins were termed "Ps 1" and "Ps 2," with Ps 2 being larger, and the null type was termed "Ps 0." The Ps proteins are not among the PRPs whose primary structures were determined elsewhere (Azen and Maeda 1988; Minaguchi and Bennick 1989; Kauffman et al. 1991). Two other basic PRPs, termed "PmF" (Ikemoto et al. 1977) and "PmS" (Azen and Denniston 1980), also show polymorphism, and the proteins are usually either both present (PmF+ and PmS+) or absent (PmF− and PmS−) from saliva. The polymorphic PmF protein is electrophoretically the same (Anderson et al. 1982; Azen 1989) as the basic PRPs IB-9 (Kauffman et al. 1982) and P-E (Isemura et al. 1982), and the polymorphic PmS protein is electrophoretically the same (Anderson et al. 1982; Azen 1989) as the basic PRP IB-6 (Kauffman et al. 1986). The primary structure of the IB-9 protein (Kauffman et al. 1982) is the same as that for the P-E protein (Isemura et al. 1982). The Pe protein also shows polymorphism and is either present (Pe+) or absent (Pe−) from saliva (Azen and Yu 1984*a*). The Pe protein is electrophoretically the same (Azen 1989) as the IIB-2 protein (Kauffman and Keller 1979), whose primary structure has been determined (Kauffman et al. 1991).

In a later comparison of protein and DNA sequences, the polymorphic PRPs PmF, PmS, and Pe were found to be coded by the *PRB*1 gene (Lyons et al. 1988*a;* Minaguchi and Bennick 1989). However, because of a lack of protein sequence data, the Ps proteins could not be securely assigned to one of the four *PRB* genes that code for basic PRPs. On the basis of nucleotide and derived amino acid sequence comparisons of two *PRB*1 copies, Lyons et al. (1988*a*) determined that different Pm phenotypes may be due to differences in the proteolytic cleavage sites of the *PRB*1 precursor protein. The PmS and PmF proteins are electrophoretically seen when cleavage occurs, and, in the absence of cleavage, the Pm proteins are not seen.

To determine the molecular basis for the Ps protein polymorphism, we first assigned the Ps proteins to their cognate *PRP* gene and determined their location within the coding region, since it was known that several PRPs could be proteolytically cleaved from a single precursor protein (Maeda et al. 1985). We then determined the molecular basis for the Ps protein polymorphism by sequencing the major coding exon 3 portions of the *PRB*1 copies of three subjects (one with the Ps 1-2 phenotype and the other two with the Ps 0 phenotype). Since the three subjects also showed the PmF−/S− phenotypes, we could test the hypothesis of Lyons et al. (1988*a*), described above, to explain these phenotypes. We also determined the molecular basis for the Pe− phenotype, since one of the subjects with the Ps 0 and PmF−/S− phenotypes also showed the Pe− phenotype.

Material and Methods

Purification of Ps Proteins for Asp-N Digests

The Ps 1 and Ps 2 proteins from salivas of single donors were gel-purified by cutting the Ps-containing bands from acid polyacrylamide gels, eluting the proteins by boiling in SDS solution, and then precipitating the proteins in acetone, as described elsewhere (Azen et al. 1979). Single bands were seen by SDS PAGE (fig. 1, lanes A6 and A9).

N-Terminal Amino Acid Sequencing of the Ps I Protein and Its Asp-N Fragment N1

For N-terminal amino acid sequencing of the intact Ps 1 protein, approximately 3 μg were sequenced on a pulsed-liquid phase sequencer (model 477A; Applied Biosystems) by using protocols and reagents supplied by the manufacturer (Wadsworth et al. 1990). Extended reaction time for proline residues was programmed for all cycles; initial yield was 20 pmol, or 12% of calculated load.

For sequencing the largest Asp-N fragment N1, the Ps 1 protein was completely digested with sequencing-grade Asp-N (Boehringer Mannheim Biochemicals), which cleaves on the N-terminal side of aspartic and cysteic acid residues when used in phosphate, acetate, or Tris buffer, pH 6.0–8.5. The Ps 1 protein was digested according to the manufacturer's instructions at a ratio of 1/125 enzyme to protein by weight. One hundred micrograms of Asp-N–digested Ps 1 protein were analyzed by SDS PAGE (Azen and Yu 1984*b*) and were electrophoretically transferred to polyvinylidene difluoride membrane, and the membrane was stained with Coomassie blue. Three Asp-N bands (N1, N2, and N3) were seen, as illustrated in other digests (fig. 1,

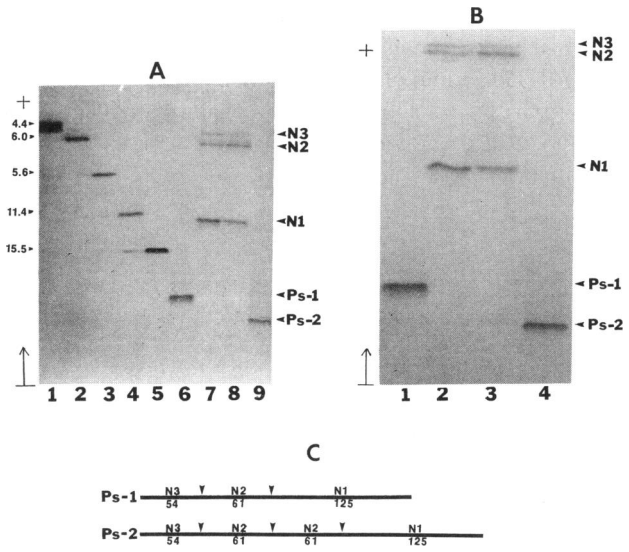

Figure 1 Asp-N digests of Ps 1 and Ps 2 proteins. Amido black–stained western blots from SDS PAGE. *A*, Sizes of Asp-N fragments of Ps 1 and Ps 2 proteins, estimated by comparison with PRP markers TZ (4,371 daltons), 1B-9 (6,024 daltons), IB-4 (5,590 daltons), IB-6 (11,446 daltons), and protein C (15,512 daltons). Molecular weights (in kD) of the markers are shown on the left margin. For an undetermined reason, the IB-9 and IB-4 proteins are not migrating exactly according to molecular size. Asp-N fragments are labeled "N1," "N2," and "N3." Lane 1, TZ peptide. Lane 2, IB-9. Lane 3, IB-4. Lane 4, IB-6. Lane 5, Protein C. Lane 6, Ps 1, 7.5 μg. Lane 7, Ps 1, Asp-N cut, 7.5 μg. Lane 8, Ps 2, Asp-N cut, 7.6 μg. Lane 9, Ps 2, 7.6 μg. *B*, Western blot for densitometry of Ps 1 and Ps 2 proteins and their Asp-N fragments. Lane 1, Ps 1, 15 μg. Lane 2, Ps 1, Asp-N cut, 15 μg. Lane 3, Ps 2, Asp-N cut, 15 μg. Lane 4, Ps 2, 15 μg. *C*, Asp-N fragment map of Ps 1 and Ps 2 proteins. The number of amino acids in the fragments N1, N2, and N3 is given below the line.

lanes A7 and B2). The stained Asp-N band N1 (representing the largest fragment) was cut from the membrane, and N-terminal sequencing was done by using the protocols and a gas phase sequencer (model 470A; Applied Biosystems) as described by Wadsworth et al. (1990). Initial yield was 10 pmol.

Laser Desorption Spectroscopy of the Ps 1 Protein

Mass spectra of the Ps 1 protein purified as described above were obtained by laser desorption–time-of-flight mass spectrometry (model VT2000; Vestec) by using sinapinic acid as the matrix. Myoglobin was used as the internal mass standard.

Cloning and Sequencing of PRB*1 Copies from Subjects with Ps, Pm, and Pe Protein Variants*

The term *PRB*1 "copy," rather than "allele," is often used in this text, since frequent *PRB*1 coding-region deletions and duplications occur in the population. For example, in a *PRB*1/2-"normal" heterozygote with two *PRB*1 coding regions on one chromosome and one *PRB*1 coding region on the other chromosome, the *PRB*1 coding regions that are allelic would not be obvious. Thus, the three types of chromosomes include *PRB*2/1, *PRB*1/2, and "normal," with zero, two, and one *PRB*1 coding regions, respectively (Azen et al. 1992). Terminology based on genomic Southern analysis for the three different-length copies of *PRB*1 (representing a length polymorphism due to different numbers of tandem repeats in exon 3) is that of Lyons et al. (1988*b*). These copies include *PRB*1L (large), *PRB*1M (medium), and *PRB*1S (small).

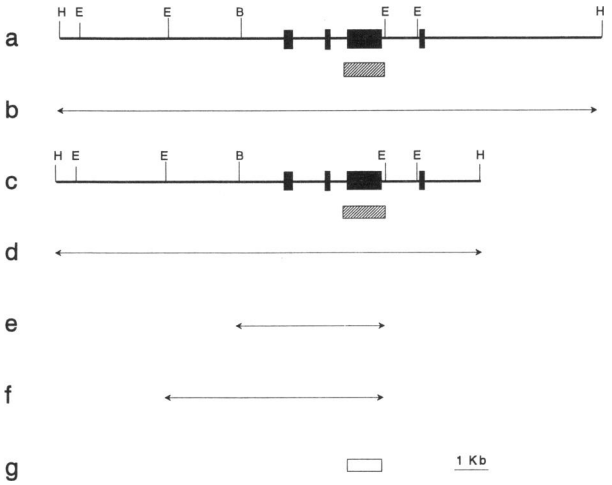

Figure 2 Restriction map of *PRB*1 and *PRB*1/2 genes and cloning strategy. *a*, Prototype *PRB*1 gene restriction map (Kim et al. 1993). *PRB*1L and *PRB*1M copies from subjects C.J., E.A., and M.V.O. show different lengths of the tandemly repetitious third exons. The region sequenced is shown below the line, as a hatched box covering exon 3. *b*, Approximately 15-kbp *Hin*dIII (H) *PRB*1 gene fragments obtained from the genomic libraries of subjects C.J. (*PRB*1M and *PRB*1L), E.A. (*PRB*1L), and M.V.O. (*PRB*1L). *c*, Prototype *PRB*1/2 fusion-gene restriction map. The *PRB*1/2S copy is from subject E.A. The intergenic crossover between *PRB*1 and *PRB*2 that generates the *PRB*1/2 fusion gene occurs in intron 3 downstream from the *PRB*1 coding region (see Azen et al. 1992 and text). The region sequenced is shown below the line, as a hatched box covering exon 3. *d*, Approximately 12-kbp *Hin*dIII *PRB*1/2 gene fragment obtained from the genomic library of subject E.A. (*PRB*1/2S). *e*, Approximately 4-kbp *Bam*HI (B)/*Eco*RI (E) fragments subcloned from the *PRB*1 and *PRB*1/2 copies of subjects C.J., E.A., and M.V.O. and used for sequencing. These cloned fragments were deleted with exonuclease III from the *Eco*RI site to prepare overlapping series of subclones for dideoxy-chain termination sequencing. *f*, Approximately 6-kbp *Eco*RI *PRB*1 gene fragment as shown on the genomic Southern blot (fig. 6). *g*, PRP *Hin*fI 980 probe used for cloning, subcloning, and genomic Southern analysis.

The *PRB*1M and *PRB*1L copies were cloned as approximately 15-kbp *Hin*dIII fragments (fig. 2*a* and *b*), and the *PRB*1/2S fusion-gene copy of subject E.A. was cloned as an approximately 12-kbp *Hin*dIII fragment (fig. 2*c* and *d*) from size-fractionated (subjects E.A. and C.J.) or alkaline phosphatase–treated (subject M.V.O.) genomic DNA, in Charon 40 (Dunn and Blattner 1987).

Recombinant bacteriophage clones were detected by hybridization to the *Hin*fI 980 PRP probe (fig. 2*g*) (from the tandemly repeated sequences of exon 3 of *PRB*1 [Azen et al. 1984]), which hybridizes to exon 3 of all six *PRP* genes (Maeda 1985). The different *PRB*1 and *PRB*1/2 copies in phage clones could be identified on the basis of characteristic polymorphic length differences, as well as by typical restriction maps (fig. 2*a* and *c;* Azen et al. 1992; Kim et al. 1993). The approximately 4-kbp *Bam*HI/*Eco*RI fragments from the phage clones (fig. 2*e*) were subcloned into plasmid Bluescript (Stratagene; La Jolla, CA). The cloned fragments were then deleted from the *Eco*RI site with exonuclease III and mung-bean nuclease (Stratagene methods) to generate an overlapping series of subclones that entirely spanned exon 3, the major coding portion of the *PRB*1 gene. The complete exon 3 portions of the *PRB*1 copies, including their splice junctions, were sequenced twice, as double-stranded DNA, by using Sequenase (United States Biochemical, Cleveland) and the reverse sequencing primer, by dideoxy-chain termination (Sanger et al. 1977).

To sequence the region between the *PRB*1 and *PRB*2 genes wherein the postulated crossover occurred to produce the *PRB*1/2S (subject E.A.) fusion-gene copy, we analyzed a portion of the third intron and adjacent downstream exon 4 region. A previous study of the *PRB*2/1 fusion gene (reciprocal to the *PRB*1/2 fusion gene) showed the crossover to occur in a 743-bp region of virtual complete identity of the *PRB*1 and *PRB*2 genes in intron 3 (Azen et al. 1992). Thus, two adjacent fragments, the 900-bp *Eco*RI and the 2.0-kbp *Eco*RI/*Hin*dIII fragments covering the intron 3 and exon 4 regions of the fusion gene (fig. 2*c*), were subcloned into Bluescript, and the ends of the fragments were sequenced as double-stranded DNA, by using Sequenase and the reverse and M13-20 primers, by dideoxy-chain termination. For the 2.0-kbp *Eco*RI/*Hin*dIII fragment, the coding strand was first partially sequenced from the *Eco*RI end by using the reverse primer. To confirm the nucleotides in the coding strand of this fragment at the six anomalous positions of mismatch with *PRB*1 and *PRB*2 genes on the 3′ side of the crossover (to be discussed), the noncoding strand of the fragment was partially sequenced by using an internal primer (5′-GGC-AGCTGGCTAAATTAGG-3′) complementary to the underlined sequence in figure 7. All sequence data were analyzed by using software provided by the 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, and Southern blots were hybridized to the *Hin*fI 980 PRP probe according to methods described elsewhere (Vanin et al. 1983; Azen et al. 1984).

Electrophoretic Analysis of Salivary PRP Polymorphisms and Asp-N Fragments of Ps 1 and Ps 2 Proteins

Parotid salivary PRPs were typed for Ps, Pe, and PmS polymorphisms by 10% SDS PAGE (Azen and Yu 1984*a;* Azen 1989) and for the PmF polymorphism by acid/lactate PAGE (Azen and Denniston 1980; Shintani et al. 1990). The identities of polymorphic PRPs in SDS PAGE were established elsewhere, by electrophoretic comparisons with biochemically purified PRPs (summarized by Azen 1989). In SDS PAGE, most of the PRPs designated in figure *5B* migrate as single bands and do not overlap other proteins, with the following exceptions: the acidic PRPs, Pr and PIF, comigrate, as do faster-migrating G1 variants with amylase (Amy). Acid/lactate PAGE (fig. *5A*) is primarily used for typing PmF, as other PRPs are not as well resolved.

For size determination of the Asp-N fragments, Ps 1 and Ps 2 proteins were completely digested with Asp-N (see previous section), were analyzed by 10% PAGE, and were western blotted on nitrocellulose, and the transfers were stained for protein with amido black (Azen and Yu 1984*b*). Since comparisons by SDS PAGE by using standard marker proteins may give spurious size estimates for PRPs (Anderson et al. 1982), PRP size markers were used. The PRP marker proteins were provided by Drs. A. Bennick (TZ peptide and protein C) and D. Kauffman (IB-9, IB-4, and IB-6).

Results

The Ps Proteins Are Coded by the PRB1 *Gene*

The N-terminal amino acid sequence was determined for the largest (N1) of the three Ps 1 Asp-N fragments (fig. 1, lanes A7 and B2). This amino acid sequence (D-K-S-R-S-P-Q-S-P-P) was then compared with the previously deduced amino acid sequences from the coding regions of all six human *PRP* genes (Kim et al. 1993). The sequence was unique to exon 3 of

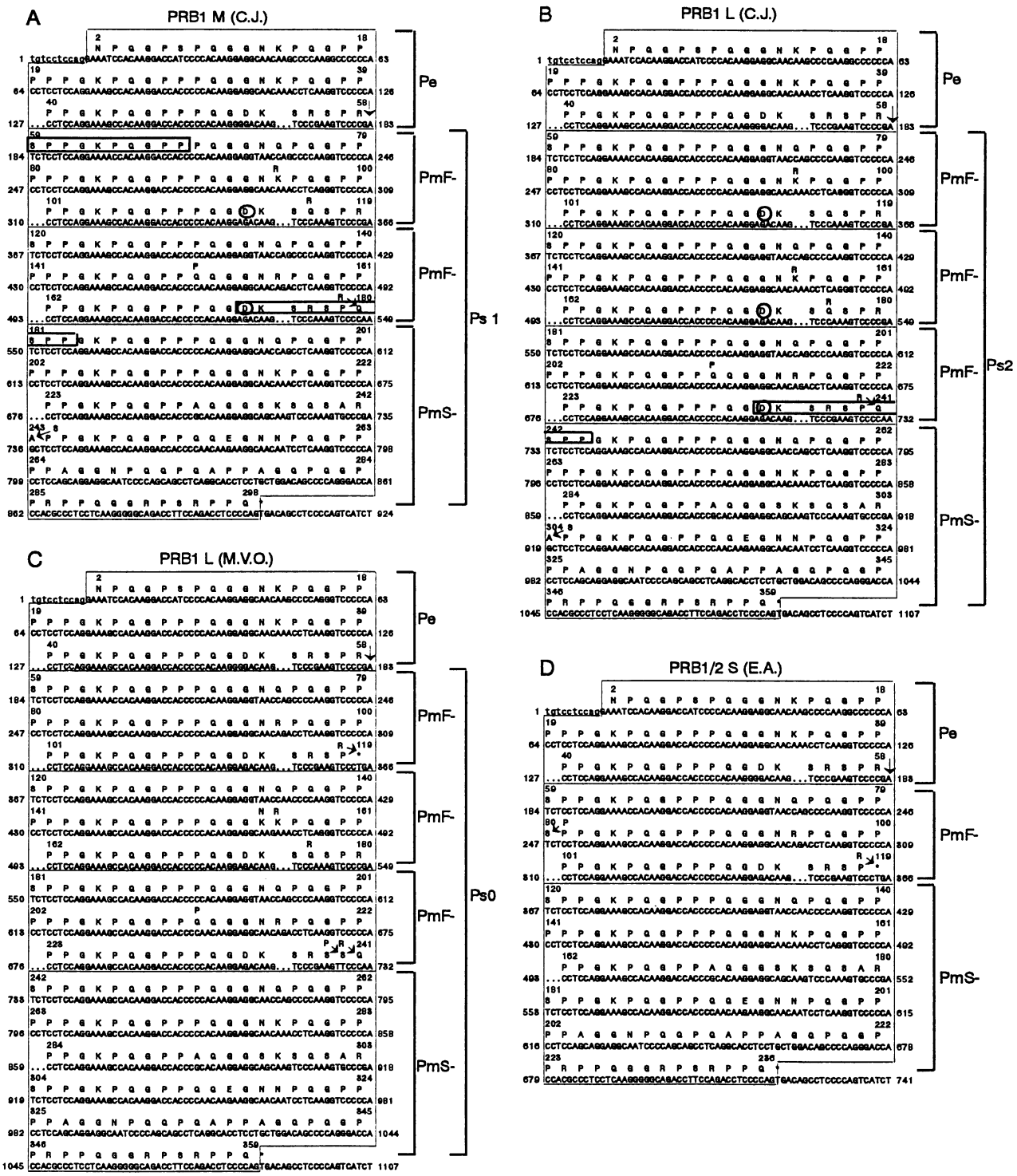

Figure 3 Nucleotide and deduced amino acids of exon 3 regions of *PRB1* copies. Relevant Asp residues (cleaved by Asp-N on the N-terminal side) are circled in *A* and *B*. The translated portion of exon 3 is boxed (thinner lines). The sequences of the amino termini of the Ps 1 protein and the largest Asp-N fragment N1 are boxed (thicker lines in *A* and *B*). The downward-pointing arrow represents the proteolytic cleavage site at Arg 58. An asterisk (*) indicates a stop codon. The amino acids and nucleotides are numbered according to their position in exon 3. Intron sequence is indicated by lowercase letters. The major portion of the Pe protein is coded by exon 3, and 17 amino acids are coded by

the *PRB*1 gene (an M copy). The same sequence described above was found in *PRB*1M and *PRB*1L copies of subject C.J. with the Ps 1-2 phenotype (boxed in fig. 3*A* and *B*).

Localization of the Ps Proteins within the Exon 3 Coding Region of the PRB*1M and* PRB*1L Copies (Alleles) of Subject C.J. with the Ps 1-2 Phenotype*

1. *Asp-N fragment sizes of the Ps 1 and Ps 2 proteins compared with those deduced from DNA sequences of* PRB*1M (subject C.J.) and* PRB*1L (subject C.J.) copies (alleles).*—The sizes of the Asp-N fragments N1, N2, and N3 were determined from a protein-stained western blot after SDS PAGE, and the mobilities of the fragments compared with PRP standards are shown in figure 1*A*. Three fragments (N1, N2, and N3) were seen in both Ps 1 and Ps 2 digests, and Asp-N bands N1, N2, and N3 in the Ps 1 digest were compared, by densitometry, with their corresponding bands in the Ps 2 digest. The protein-stained western blot used for this comparison is shown in figure 1*B*. For the Ps 1 and Ps 2 digests, the densities of the Asp-N bands (fig. 1, lanes B2 and B3) were normalized to the densities of their cognate intact Ps protein bands (fig. 1, lanes B1 and B4) before they were compared. The ratios of the normalized Asp-N bands of the Ps 2/Ps 1 digests were N1/N1 = 1.0, N2/N2 = 2.1, and N3/N3 = 1.3. There is at least a twofold greater density of the N2 band in the Ps 2 digest compared with that in the Ps 1 digest. The simplest hypothesis to explain these data is that the Ps 1 digest contains single N1, N2, and N3 fragments, whereas the Ps 2 digest contains single N1 and N3 fragments and perhaps two or more N2 fragments. From these results, it is inferred that the smaller Ps 1 protein is coded by the smaller *PRB*1M (subject C.J.) copy and that the larger Ps 2 protein is coded by the larger *PRB*1L (subject C.J.) copy.

Estimates of Asp-N fragment sizes of the Ps 1 and Ps 2 proteins, from SDS PAGE, were compared with sizes expected from the analysis of deduced exon 3 sequences of *PRB*1M (subject C.J.; fig. 3*A*) and *PRB*1L (subject C.J.; fig. 3*B*) copies (relevant Asp residues are circled). It is assumed that the putative amino-terminus of the Ps proteins immediately follows the proteolytic cleavage site at Arg 58 (shown by an arrow) and that the protein terminates at the normal stop codon (shown by an asterisk) in exon 3. This postulated amino terminus seemed reasonable, since the Pe protein had previously been localized to the coding region immediately upstream from the Arg 58 cleavage site (Lyons et al. 1988*a*; Minaguchi and Bennick 1989; Kauffman et al. 1991). The observed Asp-N fragment sizes (in kD) of the Ps 1 protein from the gel study (fig. 1*A*), versus expected sizes from the deduced exon 3 sequence of *PRB*1M (subject C.J.; fig. 3*A*), are N1 = 11.7/12.3, N2 = 5.2/6.0, and N3 = 4.7/5.2. The observed and expected sizes are in good agreement and are within the error of the SDS PAGE method for estimating sizes of proteins. The Asp-N fragment sizes of the Ps 2 protein from SDS PAGE also fit those expected from the deduced exon 3 sequence of *PRB*1L (C.J.; fig. 3*B*), which contains single N1- and N3-sized fragments and a duplicated N2-sized fragment. An Asp-N fragment map model for the Ps 1 and Ps 2 proteins is shown in figure 1*C*.

2. *Size and N-terminal sequence of the Ps 1 protein compared with the deduced exon 3 sequence of the* PRB*1M (subject C.J.) copy (allele).*—To further confirm the localization of the Ps 1 protein in the exon 3 region of the *PRB*1M (subject C.J.) copy, its size and N-terminal sequence were determined. The gel-purified Ps 1 protein was analyzed by laser desorption spectroscopy. In figure 4, the peaks labeled "1+" and "2+" represent the mass/charge (m/Z) ratios of single- and double-charged Ps 1 protein ions. The average molecular mass of the Ps 1 protein, on the basis of m/Z ratios of $(M + H)^+$ and $(M + 2H)^{2+}$ ions, was calculated to be 23,469.7 daltons, by using software provided by the manufacturer. This value closely corresponds (within 0.04%) to the mass of the Ps 1 protein (23,459.9 daltons) calculated from the amino acid sequence of its putative coding region in the third exon of the *PRB*1M copy (subject C.J.; fig. 3*A*). The putative amino-terminus (discussed previously) immediately follows the proteolytic cleavage site at Arg 58, and the protein terminates at the normal stop codon. The small size difference between the observed and expected values is within the error of the method and excludes posttranslational modifications such as glycosylation and phosphorylation. The size of the Ps 1 protein, by spectros-

exons 1 and 2. Dots are introduced into the nucleotide sequence to line up repeats according to the method of Lyons et al. (1988*a*). Differences of the published amino acid sequences of PmF and PmS proteins from the deduced sequences are shown above the deduced sequences. *A*, *PRB*1M (subject C.J.) copy. *B*, *PRB*1L (subject C.J.) copy. *C*, *PRB*1L (subject M.V.O.) copy. The sequence of the *PRB*1L (subject E.A.) copy is the same as that for the *PRB*1L (subject M.V.O.) copy that is shown in *C*, except for a single silent nucleotide change in the Pe coding region. *D*, *PRB*1/2S (subject E.A.) copy.

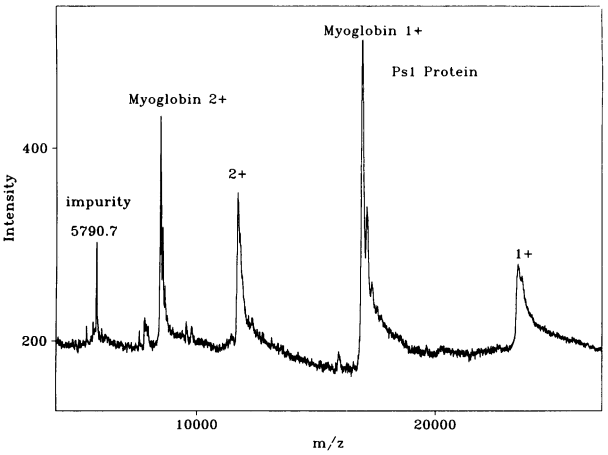

Figure 4 Size of Ps 1 protein by laser desorption spectroscopy. Gel-purified Ps 1 protein was analyzed. Myoglobin was loaded as an internal standard. 1+ and 2+ peaks represent single- and double-charged Ps 1 protein ions. The m/Z estimates for the Ps 1 peaks shown on the horizontal axis were used to calculate a mass of 23,469.7 daltons for the Ps 1 protein.

copy (23.5 kD), is also consistent with the sum of sizes of its three Asp-N fragments (21.6 kD) as determined by SDS PAGE (discussed previously).

The N-terminal sequence of the Ps 1 protein was determined to be X-P-P-G-K-P-Q-G-X-P ("X" indicates indeterminate assignment). Although this sequence is found in several locations downstream from the Arg 58 cleavage site, its location as an N-terminal sequence in the position immediately following Arg 58 (boxed in fig. 3*A*) is consistent with the estimated sizes of the Ps 1 protein and its Asp-N fragments.

Salivary PRPs, and PRB1 Length Polymorphisms of Subjects and Controls

Salivary PRPs of two subjects (C.J. and E.A.), together with those of four controls, are shown on a protein-stained western blot after SDS PAGE (fig. 5), and the Pe, Pm, and Ps phenotypes are given in the figure legend. The genomic Southern blot of *Eco*RI-digested DNAs from the same individuals was hybridized to the *Hin*fI 980 PRP probe, and the autoradiogram is shown in figure 6. Closely spaced doublet bands represent length polymorphisms at *PRB*1 (B1), *PRB*3 (B3), and *PRB*4 (B4) gene positions.

To determine the *PRB*1 genotypes for the samples, densitometry was performed on the autoradiogram shown in figure 6. For each sample, after normalization of the *PRB*1 gene signal to the *PRB*2 gene signal (a convenient two-copy control as an internal loading standard), the ratios of the signal at the *PRB*1 position to that of the two-copy *PRB*1 control 4 (lane 5) were lane 1 (control 1), 1.0; lane 2 (subject C.J.), 1.1; lane 3 (control 2), 1.3; lane 4 (subject E.A.), 1.7; and lane 6 (control 3), 1.1. The ratio of the 1L band signal of subject E.A. (lane 4) to the 1L band signal of the two-copy control 4 (lane 5) was 2.0, and the ratio of the 1S band signal of subject E.A. (lane 4) to the 1S band signal of the two-copy control 4 (lane 5) was 1.2.

From the densitometric data above, it is clear that all the samples except that of subject E.A. show two *PRB*1 copies and that subject E.A. shows three *PRB*1-type copies. The genotypes are control 1, *PRB*1M-*PRB*1M; subject C.J., *PRB*1M-*PRB*1L; control 2, *PRB*1L-*PRB*1L; subject E.A., *PRB*1L-*PRB*1L-*PRB*1/2S (the *PRB*1/2S copy will be discussed later); control 4, *PRB*1L-*PRB*1S; and control 3, *PRB*1L-*PRB*1L.

Exon 3 Nucleotide and Deduced Amino Acid Sequences of PRB1M and PRB1L Copies (Alleles) from Subject C.J. with Ps 1-2, PmF−/S−, and Pe+ Phenotypes

1. *PRB1M (subject C.J.) (fig. 3A).*—The Ps 1 protein, coded by the *PRB*1M (subject C.J.) copy, is 240 amino acids in length and (in addition to the Pe protein)

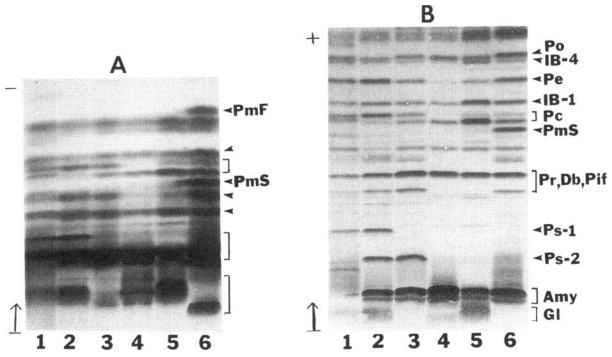

Figure 5 Salivary PRP polymorphisms of subjects C.J. and E.A. and controls (same identity as those on the genomic Southern blot; fig. 6). *A*, Protein-stained acid polyacrylamide gel for typing the PmF polymorphic protein. Other PRPs, not as well resolved as in the SDS PAGE gel (*B*), are indicated as "PmS" and with brackets and arrowheads, along the right margin of gel *A*. In order of migration from anode to cathode, these PRPs include G1, Ps and acidic PRPs (overlapping); IB-1, Pe, PmS, Pc and IB-4 (overlapping), and Po. *B*, Protein-stained western blot after SDS PAGE for typing polymorphic Ps, PmS, and Pe proteins. Other PRPs (Pr, Db, PIF, Po, IB-1, P-c, IB-4, and G1) are also shown, as is amylase. The order and identity of the samples are the same in *A* and *B*. Lane 1, Control 1, Ps 1, PmF−/S−, Pe+. Lane 2, Subject C.J., Ps 1-2, PmF−/S−, Pe+. Lane 3, Control 2, Ps 2, PmF−/S−, Pe+. Lane 4, Control 3, Ps 0, PmF−/S−, Pe+. Lane 5, Subject E.A., Ps 0, PmF−/S−, Pe+. Lane 6, Control 4, Ps 0, PmF+/S+, Pe+.

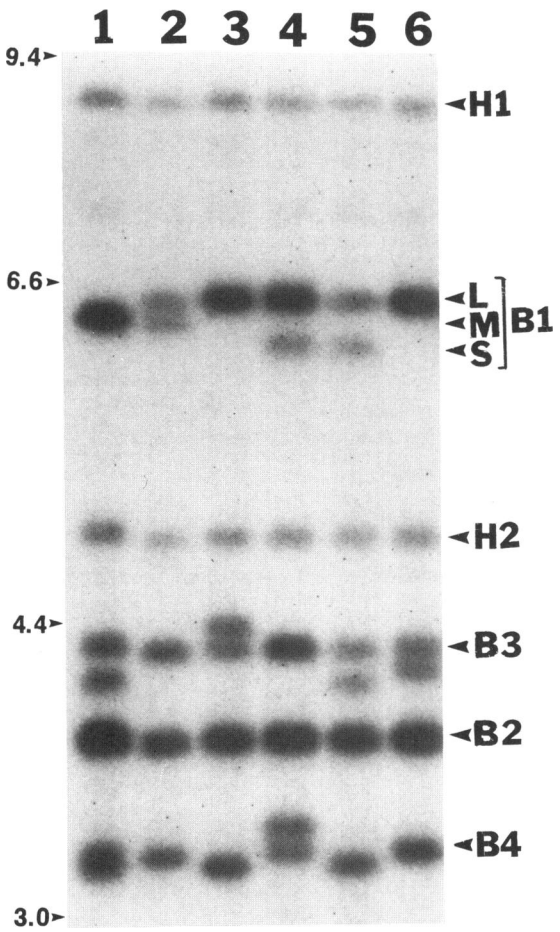

Figure 6 Genomic Southern blot of *PRB*1 length polymorphisms and genotypes of subjects C.J. and E.A. and controls (same identity as those on the protein gel; fig. 5). Closely spaced doublet bands represent length polymorphisms seen here at *PRB*1, *PRB*3, and *PRB*4 gene positions. Positions of size markers (in kbp) are shown on the left margin. The Southern blot of *Eco*RI digests was hybridized to the *Hin*fI 980 PRP probe that covers exon 3 of *PRB*1 (fig. 2*g*) and that cross-hybridizes to the exon 3 regions of the other *PRP* genes. B1, B2, B3, and B4 represent *PRB*1, *PRB*2, *PRB*3, and *PRB*4 genes, respectively; and H1 and H2 represent *PRH*1 and *PRH*2 genes, respectively. Genotypes from densitometry are as follows (see text): Lane 1, Control 1, *PRB*1M-*PRB*1M. Lane 2, Subject C.J., *PRB*1M-*PRB*1L. Lane 3, Control 2, *PRB*1L-*PRB*1L. Lane 4, Subject E.A., *PRB*1L-*PRB*1L-*PRB*1/2S. Lane 5, Control 4, *PRB*1L-*PRB*1S. Lane 6, Control 3, *PRB*1L-*PRB*1L.

is proteolytically cleaved from a precursor protein at Arg 58. For this and subsequent *PRB*1 copies to be discussed, we have included in figure 3, above the deduced amino acid sequence, the few amino acid differences from the IB-9 (PmF) (Kauffman et al. 1982) and IB-6 (PmS) (Kauffman et al. 1986) primary structures. Note that, in this *PRB*1 copy and subsequent ones to be discussed, the proteolytic cleavage site at Arg 58 is intact. However, at all other sites where cleavage could occur to generate PmF and PmS proteins, as at amino acids 119 or 180 in this figure, there is a substitution at one of the two critical Args that are separated by two amino acids at the cleavage site (Schwartz 1986; Lyons et al. 1988*a*). The substitution in this and the other *PRB*1 copies to be discussed is either Arg→Gln or Arg→stop(*).

2. PRB*1L (subject C.J.) (fig.* 3B).—The Ps 2 protein, coded by the *PRB*1L (subject C.J.) copy, is 301 amino acids in length and differs from the Ps 1 protein in that it contains three additional tandem repeats with duplication of an N2-sized Asp-N fragment, as was previously discussed. Otherwise the two proteins and nucleotide sequences are identical. Although cleavage at Arg 58 can generate a Pe protein, there are no other cleavage sites present at amino acids 119, 180, or 241.

Exon 3 Nucleotide and Deduced Amino Acid Sequences from the Single PRB*1L Copy of Subject M.V.O. with the Ps 0, PmF−/S−, and Pe− Phenotypes (Fig.* 3C)

Subject M.V.O. is a *PRB*2/1 heterozygote with one *PRB*2/1 fusion-gene copy (lacking a *PRB*1 coding region) and one *PRB*1L copy. This interpretation is supported by the densitometric analysis of subject M.V.O.'s genomic Southern blot reported elsewhere (Azen et al. 1992, fig. 5) and by family studies, as well as by isolation of both *PRB*1L and *PRB*2/1 copies (see below). The Ps 0, PmS−, and Pe− salivary phenotypes of subject M.V.O. were shown elsewhere (Azen et al. 1992, fig. 6). The *PRB*2/1 and *PRB*1L copies in subject M.V.O. are both contained on approximately 15-kbp *Hin*dIII fragments (Azen et al. 1992), and we isolated both copies from the genomic library of subject M.V.O. The two copies were distinguished on the basis of characteristic *Eco*RI restriction maps (Azen et al. 1992), and the *PRB*2/1 copy was not further studied.

DNA analysis of the single *PRB*1L (subject M.V.O.) copy offered an excellent opportunity to observe the effect of the postulated null mutation on the protein products. The *PRB*1L (subject M.V.O.) copy contains a premature stop mutation (CGA [Arg]→TGA [stop]) at amino acid 61 in the Ps protein (amino acid 119 in exon 3).

Exon 3 Nucleotide and Deduced Amino Acid Sequences of PRB*1L and* PRB*1/2S Copies from Subject E.A. with the Ps 0, PmF−/S−, and Pe+ Phenotypes (Fig.* 3D)

Subject E.A. is a *PRB*1/2 heterozygote with three *PRB*1 coding regions represented as a single *PRB*1L and *PRB*1/2S copies on one chromosome (both with

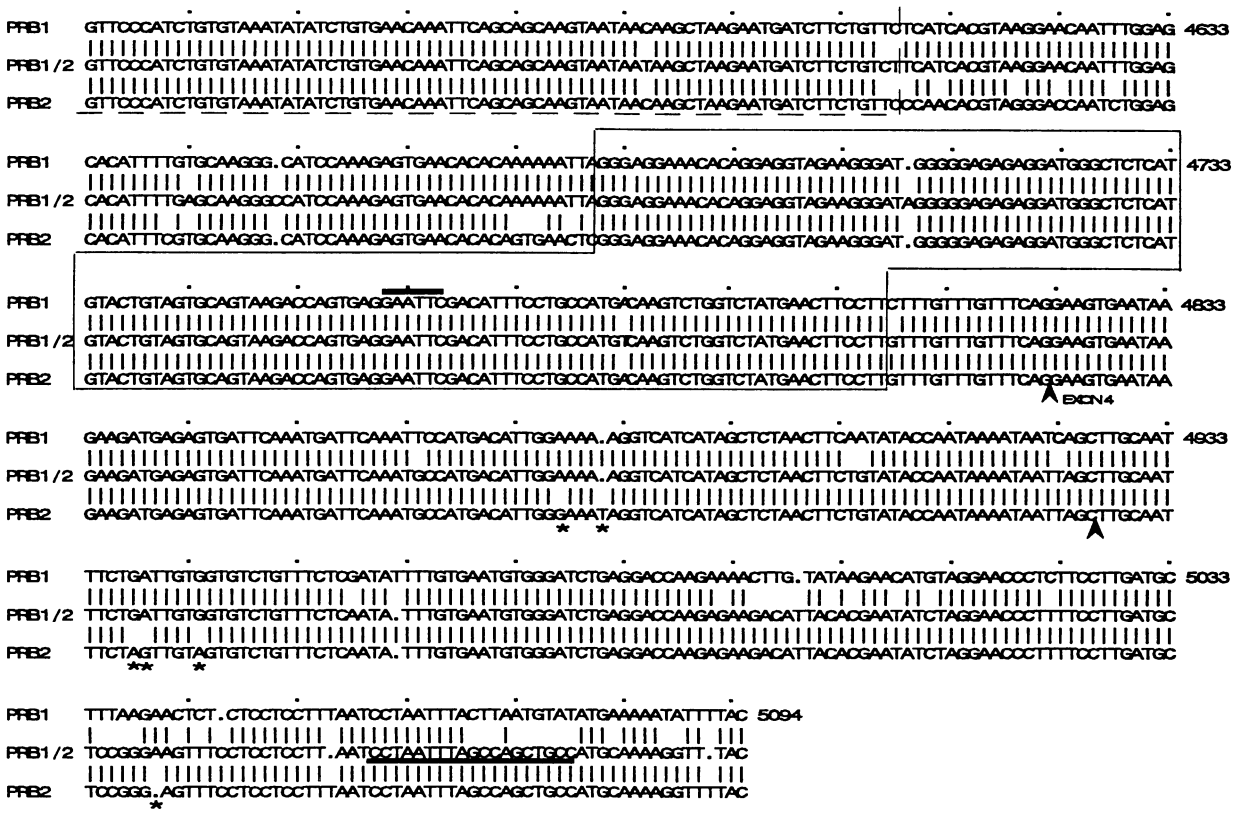

Figure 7 Nucleotide sequence comparison of the intron 3 regions of *PRB*1, *PRB*2, and *PRB*1/2 genes. The *PRB*1 and *PRB*2 sequences are from Kim et al. (1993) and begin at nucleotide 4534. The *PRB*1/2 fusion gene sequence is from the S copy (subject E.A.). The intron 3 regions for the three genes surround the *Eco*RI site (overlined) that is located 55 bp 5′ to the beginning of exon 4 (delineated by upward-pointing arrows; for orientation, see fig. 2). The 127-bp region of identity between the three genes (with the exception of two nucleotides unique to the *PRB*1/2S fusion-gene copy) is boxed by a solid line, and the postulated crossover that generated the *PRB*1/2S copy may have occurred in this region. The 3′ end of the 743-bp region of almost complete identity wherein crossover may have occurred to generate the *PRB*2/1 fusion gene (reciprocal to the *PRB*1/2 fusion gene) is partially enclosed by a dashed line. Asterisks indicate anomalous mismatches 3′ to the postulated 127-bp crossover region (see text). A primer complementary to the underlined sequence of the *PRB*1/2 gene was used to partially sequence the noncoding strand, in order to confirm in the coding strand the six anomalous mismatches indicated by asterisks (see text).

intact *PRB*1 coding regions) and as a single *PRB*1L copy on the other chromosome (Azen et al. 1992). This interpretation is supported by the previously discussed densitometric analysis of the genomic blot samples (fig. 6).

The *Hin*dIII and *Eco*RI restriction map of the *PRB*1S-sized (subject E.A.) copy is that of a *PRB*1/2 fusion gene with an intact *PRB*1 coding region (fig. 2*c;* Azen et al. 1992). To further confirm and detail the nature of the postulated crossover event between *PRB*1 and *PRB*2 genes that generated the *PRB*1/2S copy, the DNA sequences of the *PRB*1/2S (subject E.A.) copy and the sequences of the *PRB*1 and *PRB*2 genes (Kim et al. 1993) from intron 3 and exon 4 were compared (fig. 7). From the sequence results, the crossover is postulated to occur in a 127-bp region of identity in intron 3 (with the exception of two nucleotides that are unique to the *PRB*1/2S copy), and this crossover region is boxed by a solid line. The 3′ end of this crossover region is 15 bp 5′ to exon 4. The 5′ end of this crossover region begins 72 bp 3′ to the end of the 743-bp region of almost complete identity (partially enclosed by a dashed line) in which the crossover that generated the *PRB*2/1 fusion gene is postulated to have occurred (Azen et al. 1992). Immediately 5′ to the postulated *PRB*1/2 crossover region, there are 11 nucleotide positions where the *PRB*1/2S copy is the same as the *PRB*1 gene but differs from the *PRB*2 gene. There are no examples of the reverse situation in this region. In the region immediately 3′ to the 127-bp crossover region,

there are 37 nucleotide positions where the *PRB*1/2S copy is the same as the *PRB*2 gene but differs from the *PRB*1 gene. However, in the same region, there are six nucleotide positions where the *PRB*1/2S copy is the same as the *PRB*1 gene but differs from the *PRB*2 gene (indicated by asterisks). The nucleotides of the *PRB*1/2S copy at these six anomalous mismatches were confirmed by sequencing both DNA strands.

1. PRB*1L (subject E.A.) (not shown)*.—This copy is identical (except for a single silent nucleotide difference in the Pe coding region) to the previously discussed *PRB*1L (subject M.V.O.) copy (fig. 3*C*) and contains the same premature stop mutation. Although the *PRB*1L (subject E.A.) copy with the stop mutation is identical in length to the Ps 2 productive *PRB*1L (subject C.J.) copy, there are nine single-nucleotide differences (including the stop mutation) between them within the Ps coding region, but these changes would be irrelevant to the Ps 0 phenotype and are not shown.

2. PRB*1/2S (subject E.A.) (fig. 3D)*.—The premature stop mutation is identical to that seen in the larger *PRB*1L copies of subjects E.A. and M.V.O. and also abolishes the proteolytic cleavage site that, in part, could generate the PmF and PmS proteins, as was seen from a Pm productive *PRB*1S copy sequenced elsewhere (Lyons et al. 1988*a*). The second *PRB*1L copy that we have not sequenced in subject E.A. probably also contains the same premature stop mutation, since the subject shows the Ps 0 phenotype.

DNA Sequencing-Gel Results

DNA sequencing-gel autoradiograms show nucleotides surrounding and including the Arg 61 codon in the Ps protein (amino acid 119 in exon 3) that is present in Ps productive copies *PRB*1M and *PRB*1L from subject C.J. (fig. 8*A* and *B*) and that is mutated to a stop codon in the *PRB*1L and *PRB*1/2S copies from subject E.A. (fig. 8*C* and *D*) and in the *PRB*1L copy of subject M.V.O. (fig. 8*E*).

Discussion

We report here the derived structures of the Ps 1 and Ps 2 proteins and the molecular basis for polymorphisms among *PRB*1-coded *PRP*s (Ps, PmF, PmS, and Pe). The identity and structure of the Ps 1 protein was determined by jointly considering its estimated size, the sizes of Asp-N fragments, experimentally determined amino acid sequences of the intact protein and its largest Asp-N fragment N1, and the deduced amino acid sequence of the *PRB*1M (subject C.J.) allele. Although amino acid sequence and size were not determined for the larger Ps 2 protein, its amino acid sequence was deduced from the nucleotide sequence of the *PRB*1L (subject C.J.) allele. This amino acid sequence is identical to that of the Ps 1 protein (with the exception of three additional tandem repeats in the Ps 2 protein). The sizes and number of deduced Asp-N fragments are consistent with the observed fragment pattern.

The Ps proteins were elsewhere tentatively and incorrectly assigned to the *PRB*2 gene (Lyons et al. 1988*a*), in part on the basis of comparison of the amino acid composition of the Ps 1 protein (Goodman et al. 1985) with that derived from a decoded peptide of a *PRB*2L copy. In the present paper we have assigned the Ps proteins to the *PRB*1 gene and have localized the coding region entirely within exon 3. This assignment is further supported by a study elsewhere of two *PRB*2/1 fusion-gene homozygotes who lack *PRB*1 coding regions in their *PRB*2/1 fusion genes and who show the electrophoretic absence of Ps proteins in their saliva (Azen et al. 1992).

We initially described a length (Ps 1 and Ps 2) and null (Ps 0) polymorphism on the basis of SDS PAGE studies (Azen and Denniston 1980). Subsequent analysis of proteolytic digests of the Ps proteins by SDS PAGE (Goodman and Karn 1983) supported the proposed length difference of the Ps 1 and Ps 2 proteins. In a Japanese study, five Ps length variants were found: Ps 1, Ps 1^F, Ps 2^F, Ps 2^S, and Ps 3 (Minaguchi et al. 1988). However, in these early studies, the molecular basis for the length variants was not determined. Among the mechanisms that could account for these length differences are unequal crossing-over, translational modifications (such as differential proteolysis or glycosylation), and mutation to a premature termination codon.

In the present study, the decoded Ps 1 and Ps 2 proteins are, respectively, 240 and 301 amino acids in length and differ by three additional tandem repeats in the Ps 2 protein. It is likely that other length variants among Ps proteins (Minaguchi et al. 1988) will differ in the number of tandem repeats. Length differences among the glycosylated G1 proteins coded by the *PRB*3 gene (Lyons et al. 1988*a*) and among the acidic PRPs (Pa, Db, and PIF) coded by the *PRH*1 gene (Maeda et al. 1985; Azen et al. 1987) show similar patterns. Lyons et al. (1988*b*) studied the extensive insertion/deletion polymorphisms within the *PRB*1, *PRB*2, *PRB*3, and *PRB*4 loci and explained the generation of DNA length variants by homologous and unequal intragenic exchanges in exon 3. It is likely that a similar mechanism at the *PRB*1 locus explains the length differences

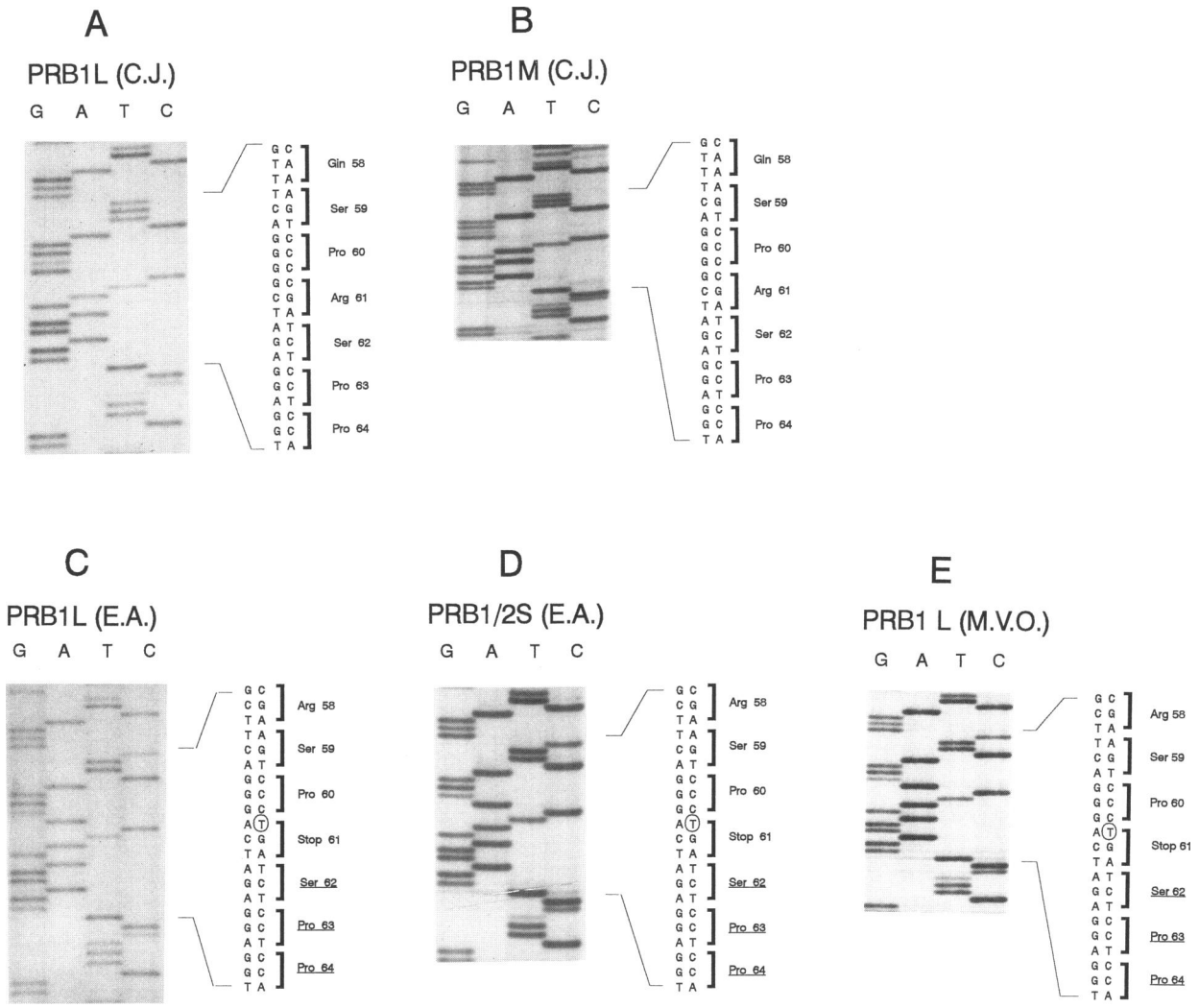

Figure 8 DNA sequence gel results and deduced amino acids from *PRB*1 copies of subjects C.J., M.V.O., and E.A. The sequences are from the noncoding strand. The coding-strand sequences and deduced amino acids, which are numbered according to their position in the Ps-coding region of exon 3 of *PRB*1L and *PRB*1M copies (*A*, *B*, *C*, and *E*), are shown, as is the same region from the *PRB*1/2S copy (*D*). *A* and *B*, *PRB*1L and *PRB*1M copies of subject C.J. Note especially the same Arg 61 in the Ps proteins (amino acid 119 in exon 3). *C* and *D*, *PRB*1L and *PRB*1/2S copies of subject E.A. *E*, *PRB*1L copy of subject M.V.O. Note the identical stop mutation (CGA [Arg]→TGA [stop]) at residue 61 in *C*, *D*, and *E*. Underlined amino acids are untranslated. The mutation is circled.

among the Ps proteins. The Ps 1 coding sequence from the *PRB*1M (subject C.J.) copy that we have studied is identical to that of both a *PRB*1M copy sequenced by Lyons et al. (1988*a*) and another *PRB*1M copy from a different individual, sequenced by Kim et al. (1993). However, the exact Ps genotypes of these two individuals were not known.

We have found that the Ps 0, PmF−/S−, and Pe− phenotypes in subject M.V.O. are probably caused by the premature stop mutation (CGA [Arg]→TGA [stop]) at amino acid 61 in the Ps coding region of the single *PRB*1L copy. This C→T mutation occurs at a CpG dinucleotide on the coding strand and adequately accounts for the Ps 0 and PmS−/F− phenotypes. The CpG dinucleotide is a mutation hot spot that is responsible for a sizable proportion of human genetic disease (reviewed by Cooper and Krawczak 1990). The premature stop mutation could produce a truncated protein 135 amino acids in length, which would be cleaved at Arg 58 in exon 3 to generate a "PmF-like" protein and a

Pe protein, but neither protein was electrophoretically seen in subject M.V.O. However, this truncated protein may be produced because of the presence of a reduced amount of Pe protein in other individuals with the mutation, such as subject E.A. with three mutated *PRB*1 copies and no nonmutated copies. Also, control 3 with the *PRB*1L-*PRB*1L genotype and Ps 0 phenotype is presumed to have two mutated *PRB*1L copies and no nonmutated copies and shows a markedly reduced level of Pe protein (fig. *5B*). Densitometric studies of figure *5B* (not shown) confirm these initial observations, but more genotype/salivary phenotype comparisons will be needed to determine the quantitative effect of the mutation on the Pe protein level. If produced, the PmF-like protein is 60, rather than 61, amino acids in length and would be more acidic, because of the loss of the terminal Arg, with a consequent reduction in basic residues from seven to six. A PmF-like protein might migrate differently than the "normal" PmF protein in the acid polyacrylamide typing gel (Azen and Denniston 1980) and thus not be recognized. Also, it might not be normally exported from the cell or might be rapidly degraded.

Premature stop mutations (as described here) may lead to variably reduced or absent steady-state mRNA levels, as has been seen in a number of other genes (reviewed by Urlaub et al. [1989] and Cheng et al. [1990]), and this mechanism could contribute to the electrophoretically reduced level of Pe protein produced from a mutated *PRB*1 copy. The occurrence of the same stop mutation on two different-length *PRB*1 copies (*PRB*1L and *PRB*1/2S) in subject E.A. is also interesting. It seems likely that intra-allelic spread of the mutation has occurred by homologous and unequal intragenic exchange (Lyons et al. 1988*b*). In the *PRB*1 copies with the same stop mutation, the sequences of exon 3 splice junctions are normal. They conform overall to the consensus sequences (Shapiro and Senapathy 1987) and are the same as those from *PRB*1 copies without the mutation.

As discussed by Schwartz (1986), the amino acid sequence basic-X-X-Arg is common to a variety of monobasic proteolytic cleavage sites, and Lyons et al. (1988*a*) have argued that the sequence Arg-Ser-X-Arg-Ser serves to generate proteolytic cleavage sites in human PRPs, with cleavage occurring on the carboxy side of the second Arg. Lyons et al. (1988*a*) sequenced a *PRB*1M and a *PRB*1S allele and correlated the PmF and PmS phenotypes with the presence or absence of appropriate proteolytic cleavage sites in the derived precursor proteins. Data from our current study strongly support this hypothesis. Thus in the Ps-productive but PmF−/S− *PRB*1M (subject C.J.) and *PRB*1L (subject C.J.) copies, multiple CGA (Arg)→CAA (Gln) substitutions are seen in the coding strand at potential cleavage sites, and these probably represent C→T mutations on the noncoding strand at CpG mutation hot spots. As noted by Lyons et al. (1988*a*), these Arg→Gln substitutions occur at either of the two presumed critical Args, separated by two amino acids, that occur at potential cleavage sites to generate PmF and PmS proteins from the precursor proteins. The same pattern was seen in the *PRB*1L copies of subjects E.A. and M.V.O. with the premature stop mutation. These latter Arg→Gln substitutions are interesting but irrelevant to the phenotype, since they occur downstream from the premature stop mutation.

Lyons et al. (1988*a*) correlated PmF and PmS phenotypes with *PRB*1 length variants and found that all individuals with the PmF+/S+ phenotype carry the *PRB*1S copy, and no individual with the *PRB*1S copy was typed as either PmF− or PmS−. However, in the current study, subject E.A. with the same-length *PRB*1/2S copy shows the PmF−/S− type due to the premature stop mutation. The mutation occurs at a presumed proteolytic cleavage site present in the previously sequenced *PRB*1S copy that was productive for PmF and PmS proteins (Lyons et al. 1988*a*), and this further supports the significance of the cleavage site. Lyons et al. (1988*a*) also showed that most individuals with *PRB*1M copies show the PmF−/S− phenotype, but some show the less common PmF+/S− phenotype. This suggests that some *PRB*1M copies may contain cleavage sites to generate the PmF protein from the precursor protein.

Null phenotypes were elsewhere noted to occur at high frequency among basic polymorphic PRPs such as Ps, Pe, and Pm (from *PRB*1); Po, Con1, and Con2 (from *PRB*4); and G1 (from *PRB*3) (Azen and Maeda 1988; Minaguchi and Bennick 1989; Azen et al. 1990, 1992). To summarize, several different mechanisms are now known to explain some of these null phenotypes. In *PRB*2/1 fusion gene homozygotes, Ps, Pm, and Pe proteins are absent from saliva. This is due to deletion of *PRB*1 coding regions from the *PRB*2/1 chromosome (Azen et al. 1992). In the present report we describe a CGA (Arg)→TGA (stop) mutation that also accounts for some Ps 0, Pm−, and Pe− phenotypes. We also confirmed the study by Lyons et al. (1988*a*) that describes the lack of integrity of critical proteolytic cleavage sites to explain some PmF− and PmS− types. A null allele (*PRB*3M^{null}) was elsewhere described in a G1 8-0

heterozygote (Azen et al. 1990). There was a frameshift insertion of a single C nucleotide in exon 3 of this allele, and this produced a premature stop codon. In subsequent studies (Azen et al., submitted), we have found the same mutation in both *PRB*3 alleles of a G1 0-0 individual with no electrophoretically visible G1 protein. As yet, Po− and Con− phenotypes have not been explained.

The combined effects of the premature stop mutation and fusion genes (*PRB*1/2 and *PRB*2/1) on *PRB*1-coded proteins are surprisingly complex. Since both *PRB*2/1 fusion-gene chromosomes (missing *PRB*1 coding regions), *PRB*1/2 fusion-gene chromosomes (with duplicated *PRB*1 coding regions), and "normal" chromosomes (with single *PRB*1 coding regions) are segregating at high frequencies in populations, the *PRB*1 copy number in any individual can vary from zero to four (Azen et al. 1992). However, in any individual the segregation of *PRB*1 copies with the premature stop mutation described herein could modify the level of *PRB*1-coded proteins in a subtle way. Thus the stop mutation eliminates the production of Ps and PmS proteins and reduces the level of Pe protein. It may also result in a PmF-like protein, but this protein is not electrophoretically seen in saliva.

Previous evidence for the existence of the *PRB*1/2 fusion gene was based primarily on densitometric studies of restriction endonuclease-digested DNAs from family members, on genomic Southern blots hybridized to the *Hin*fI 980 PRP probe (Azen et al. 1992). In the present paper we present conclusive evidence for the existence of the *PRB*1/2 fusion gene, with details of the nucleotide sequence surrounding and including the postulated crossover junction between the *PRB*1 and *PRB*2 genes. Two especially interesting features in the crossover region were noted. First, the crossover may have occurred in a 127-bp region of identity between *PRB*1 and *PRB*2 genes in intron 3. This is located only 72 bp 3′ to the 743-bp region of virtual complete identity between *PRB*1 and *PRB*2 genes in intron 3 where crossover may have occurred to produce the reciprocal *PRB*2/1 fusion gene (Azen et al. 1992). Because of their close proximity, it seems likely that this large 743-bp region of virtual complete identity between *PRB*1 and *PRB*2 genes has also influenced the unequal gene alignment and/or strand exchange involved in the smaller postulated crossover region farther downstream that produced the *PRB*1/2S fusion-gene copy (Metzenberg et al. 1991). Second, downstream from the crossover site, there were anomalous mismatches at 6 of 43 nucleotide positions where the *PRB*1/2 gene is the same as the *PRB*1 gene but differs from the *PRB*2 gene. DNA sequence analysis of homologous and unequal crossover exchanges in humans, as with the δ-β (Lepore globin) gene fusions, usually does not show this anomaly; however, such a pattern was seen for the single γ-globin fusion gene (Metzenberg et al. 1991). In the 5′ region, the γ-globin fusion gene was $^{G}\gamma$, but 3′ to the crossover there were features characteristic of both $^{G}\gamma$ and $^{A}\gamma$ before the fusion gene ended up as $^{A}\gamma$. This could be explained by gene conversions (and possibly point mutations) related to the homologous but unequal crossover event that led to the fusion (Metzenberg et al. 1991). A similar patchy repair process has been seen in recombination experiments in yeast (Orr-Weaver et al. 1988). Alternatively, the anomalous mismatches may represent polymorphic differences between *PRB*1 and *PRB*2 genes. Sequence analysis of the potential crossover region of additional *PRB*1 and *PRB*2 genes would help to interpret the mismatches.

Acknowledgments

We greatly appreciate the technical assistance of E. Amberger. We thank M. Lacey and G. Parr for assistance with laser desorption spectroscopy, and we thank L. Waite for help with gel densitometry. We thank A. Metzenberg, L. Zhuo, L. Sabatini, and H.-S. Kim for helpful comments. This investigation was supported by National Institutes of Health grant DE03658-27.

References

Anderson LC, Kauffman DL, Keller PJ (1982) Identification of Pm and PmS human parotid salivary proteins as basic proline-rich proteins. Biochem Genet 20:1131–1137

Ann DK, Carlson DM (1985) The structure and organization of a proline-rich protein gene of a mouse multigene family. J Biol Chem 260:15863–15872

Ann DK, Smith MK, Carlson DM (1988) Molecular evolution of the mouse proline-rich protein multigene family. J Biol Chem 263:10887–10893

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, Davisson MT, Cherry M, Taylor BA (1989) Prp (proline-rich protein) genes linked to markers Es-12 (esterase-12), Ea-10 (erythrocyte alloantigen) and loci on distal mouse chromosome 6. Genomics 5:415–422

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 G1, Db and Pr genetic determinants. Biochem Genet 18:483–501

Azen EA, Hurley CK, Denniston C (1979) Genetic polymorphism of the major parotid salivary glycoprotein (G1) 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 *PRH*1 locus coding for the human salivary-acidic proline-rich proteins 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 *PRB*3 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, O'Connell P, Kim H-S (1992) *PRB*2/1 fusion gene: a product of unequal and homologous crossing-over between proline-rich protein (PRP) genes *PRB*1 and *PRB*2. Am J Hum Genet 50:842–851

Azen EA, Prakobphol A, Fisher SJ. *PRB*3 null mutations result in absence of the proline-rich glycoprotein Gl (PRG) and abolish *Fusobacterium nucleatum* interactions with saliva in vitro (submitted)

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 concanavalin A reactions on nitrocellulose with linkage of CON1 and CON2 genes to the salivary protein gene complex (SPC). Biochem Genet 22:1–19

Cheng J, Fogel-Petrovic M, Maquat LE (1990) Translation to near the distal end of the penultimate exon is required for normal levels of spliced triosephosphate isomerase mRNA. Mol Cell Biol 10:5215–5225

Cooper DN, Krawczak M (1990) The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. Hum Genet 85:55–74

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

Goodman PA, Karn RC (1983) Human parotid size polymorphism (Ps): characterization of the two allelic products, Ps 1 and Ps 2 by limited proteolysis. Biochem Genet 21:405–416

Goodman PA, Yu P-L, Azen EA, Karn RC (1985) The human salivary protein complex (SPC): a large block of related genes. Am J Hum Genet 37:785–797

Ikemoto S, Minaguchi K, Suzuki K, Tomita K (1977) New genetic marker in human parotid saliva (Pm). Science 197:378–379

Isemura S, Saitoh E, Sanada K (1982) Fractionation and characterization of basic proline-rich peptides of human parotid saliva and the amino acid sequence of proline-rich peptide P-E. J Biochem (Tokyo) 91:2067–2075

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

Kauffman D, Hofmann T, Bennick A, Keller PJ (1986) Basic proline-rich proteins from parotid saliva: complete covalent structures of proteins IB-I and IB-6. Biochemistry 25:2387–2392

Kauffman DL, Keller PJ (1979) The basic proline-rich proteins in human parotid saliva from a single subject. Arch Oral Biol 24:249–256

Kauffman D, Wong R, Bennick A, Keller PJ (1982) Basic proline-rich proteins from human parotid saliva: complete covalent structure of protein IB-9 and partial structure of protein IB-6, members of a polymorphic pair. Biochemistry 21:6558–6562

Kim H-S, Lyons KM, Saitoh E, Azen EA, Smithies O, Maeda N (1993) The structure and evolution of the human salivary proline-rich protein gene family. Mammalian Genome 4:3–14

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

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

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

Minaguchi K, Shirotani T, Suzuki K (1988) New variants of Ps salivary polymorphic proteins. Ann Hum Genet 52:11–16

Orr-Weaver TL, Nicolas A, Szostak JW (1988) Gene conversion adjacent to regions of double-strand break repair. Mol Cell Biol 8:5292–5298

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schwartz TW (1986) The processing of peptide precursors: proline-directed arginyl cleavage and other monobasic processing mechanisms. FEBS Lett 200:1–9

Shapiro MB, Senapathy P (1987) RNA splice junctions of different classes of eukaryocytes: sequence statistics and functional implications in gene expression. Nucleic Acids Res 15:7155–7174

Shintani M, Minaguchi K, Lim KA, Hashimoto M, Suzuki K (1990) Salivary proline-rich polymorphisms in Chinese, Malays and Indians in Singapore. Hum Hered 40:89–98

Urlaub G, Mitchell PJ, Ciudad CJ, Chasin LA (1989) Nonsense mutations in the dihydrofolate reductase gene effect RNA processing. Mol Cell Biol 9:2868–2880

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

Wadsworth CL, Niece RL, Gimenez JA, DasGupta BR (1990) Extended N-terminal protein sequencing of *Clostridium botulinum* neurotoxin type A fragments: continuous sequence to seventy residues both from glass-fiber filter using the gas phase and polyvinylidene difluoride using the pulsed liquid phase sequencer. In: Villafranca JJ (ed) Current research in protein chemistry. Academic Press, San Diego, pp 167–178