Am. J. Hum. Genet. 58:143–153, 1996

PRB1, *PRB2*, and *PRB4* Coded Polymorphisms among Human Salivary Concanavalin-A Binding, II-1, and Po Proline-Rich Proteins

E. A. Azen,[1] E. Amberger,[1] S. Fisher,[3,4,5,6] A. Prakobphol,[3] and R. L. Niece[2]

[1]Departments of Medicine and Medical Genetics and [2]Biotechnology Center, University of Wisconsin, Madison; and Departments of [3]Stomatology, [4]Anatomy, [5]Pharmaceutical Chemistry, and [6]Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Francisco

Summary

Six closely linked *PRP* (proline-rich protein) genes code for many salivary PRPs that show frequent length and null variants. From determined protein sequences and DNA sequence analysis of variant alleles, we here report the coding and molecular basis for Con (concanavalin A-binding) and Po (parotid "o") protein polymorphisms. The Con1 glycoprotein is encoded in exon 3 of a *PRB2* allele (*PRB2*L CON1+) with a potential N-linked glycosylation site. Because of a probable gene conversion encompassing ≥684 bp of DNA, the "*PRB2*-like" Con2 glycoprotein is encoded in exon 3 of a *PRB1* allele (*PRB1*M CON2+) with a potential glycosylation site. The PmF protein is also encoded in the *PRB1*M CON2+ allele, thus explaining the previously reported association between Con2 and PmF proteins. A *PRB2*L CON1− allele contains a single nt missense change [TCT(Ser)→CCT(Pro)] that abolishes the potential N-linked glycosylation site (NKS→NKP) in the Con1 protein, and this explains the Con− type. The Po protein and a glycoprotein (II-1) are encoded in the *PRB4* gene, and both proteins are absent in the presence of a mutation in the *PRB4*M PO− allele that contains a single nt change (G→C) at the +1 invariant position of the intron 3 5′ donor splice site. The genetically determined absence of the II-1 glycoprotein leads to altered in vitro binding of *Streptococcus sanguis* 10556 to salivary proteins, which suggests a biological consequence for null mutations of the *PRB4* gene.

Introduction

The specialized salivary glands (submandibular, parotid, and sublingual) and small buccal glands produce the complex whole human saliva that is an easily obtained source of many proteins that are useful for genetic analyses (Azen and Maeda 1988). The stimulated parotid saliva component (collected with the Curby cup) has advantages over whole saliva, since it is more uniform and less subject to enzymatic degradation or contamination with food and bacteria. Many parotid protein components have been genetically studied, such as cystatins (cysteine protease inhibitors), amylases, statherin (a protein important for modulating calcium homeostasis), histatins (antifungal and antibacterial proteins), and especially proline-rich proteins (PRPs), the most abundant class of proteins which are the subject of this paper.

The salivary proline-rich proteins constitute about two thirds of parotid salivary proteins and consist of ≥25 electrophoretically discernible species that have been subdivided into glycosylated, basic, and acidic types. The salivary PRPs have several important activities: they bind hydroxyapatite, calcium, and certain intraoral bacteria; they mediate adherence of microorganisms to the coated tooth surface; they inhibit hydroxyapatite formation; they modify lubricative properties of saliva; and they detoxify dietary tannins (Minaguchi and Bennick 1989).

Six closely linked *PRP* genes, located on chromosome 12p13.2, code for these many salivary PRPs that show frequent polymorphisms (Azen and Maeda 1988; Minaguchi and Bennick 1989; Azen 1993). Among the *PRP* genes, *PRB1*, *PRB2*, *PRB3*, and *PRB4* encode basic and glycosylated PRPs, and *PRH1* and *PRH2* encode acidic PRPs (Maeda 1985). The complete sequences, evolutionary relationships, and physical linkage of the six genes were reported by Kim et al. (1990, 1993). A single *PRP* gene may produce multiple PRPs by allelic variations, posttranslational cleavages, and differential RNA processing (Maeda et al. 1985). Frequent intragenic and intergenic crossovers occurring within the *PRP* genes lead to DNA length polymorphisms and gene deletions and duplications (Lyons et al. 1988*b*; Azen et al. 1992). The molecular bases for protein polymorphisms coded by the *PRH1* gene (Azen et al. 1987), the *PRB1* gene

Received March 16, 1995; accepted for publication October 17, 1995.

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

0002-9297/96/5801-0017$02.00

(Azen et al. 1992; Azen et al. 1993*a*), and the *PRB*3 gene (Azen et al. 1990; Azen et al. 1993*b*) have been determined. Among at least 15 genetic polymorphisms of PRPs is that of two proteins named "Con1" and "Con2" because of their in vitro binding of concanavalin A (Con A). The Con proteins (of unknown structure and function) are of special biological interest, since they, like other glycosylated PRPs, might bind to specific oral bacteria and modulate intraoral disease susceptibilities (Murray et al. 1992). The Con protein polymorphism includes the different-sized Con1 and Con2 proteins (Con1+ and Con2+ types) and the Con− type and is determined by concanavalin A staining for glycoproteins on Western blots (Azen and Yu 1984*b*). Because of the lack of protein sequence data and gene assignment, the molecular basis for the Con protein polymorphism was unknown, although it was postulated (wrongly, in retrospect) that the Con proteins may be encoded by the *PRB*4 gene (Lyons et al. 1988*a*). Thus, we wished to determine the coding, structure, and molecular basis for the Con protein polymorphism.

We also studied the Po (parotid "o") protein polymorphism (Po+ and Po− types) for several reasons. First, although the Po protein is encoded in the *PRB*4 gene (Lyons et al. 1988*a*), the molecular basis for the Po protein polymorphism (Azen and Yu 1984*a*) was not known. Second, Kauffman et al. (1993) showed that another salivary glycoprotein (II-1 protein) is also encoded in the *PRB*4 gene, and we wished to determine the effect of the Po− mutation on the II-1 glycoprotein. If the Po− type is associated with a postulated II-1− type, we could then test the effect of the mutation on specific bacterial binding to the II-1 glycoprotein.

We here report the protein coding and molecular bases for polymorphisms among the Con, Po, and II-1 PRPs. The Con1 glycoprotein is encoded in the *PRB*2 gene, and the "*PRB*2-like" Con2 glycoprotein is encoded in the *PRB*1 gene because of a probable gene conversion. The Con1− allele contains a single nt change that abolishes the potential N-linked glycosylation site in the Con1 protein and explains the Con− phenotype. The Po− type is associated with the absence of the II-1 glycoprotein. A Po− allele contains a single nt change at the intron 3 5′ donor splice site of the *PRB*4 gene that explains the Po-/II-1− types. In the genetically determined absence of the II-1 protein, there is altered in vitro binding of *Streptococcus sanguis* 10556 to salivary proteins. This result suggests a biological consequence for the null mutations of the *PRB*4 gene.

Material and Methods

Cloning and Sequencing PRB*1,* PRB*2, and* PRB*4 Alleles from Subjects with Con and Po Protein Variants*

DNA samples were isolated from peripheral blood leukocytes of three subjects (R.S., L.S., and J.J.) who

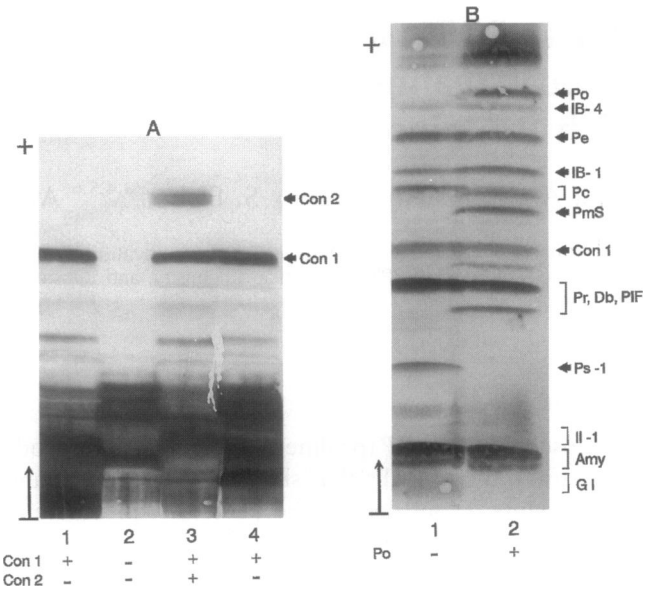

Figure 1 Po and Con protein polymorphisms. Parotid saliva proteins were separated in 15% SDS PAGE, electrophoretically transferred to nitrocellulose, and stained for glycoproteins with Concanavalin A (*A*) or for proteins with Amido Black (*B*). *A*, Samples 1 and 4: Con1+/Con2−, controls. Sample 2: Con1-/Con2−, subject L.S. Sample 3: Con1+/Con2+, subject R.S. *B*, Sample 1: Po−, subject J.J. Sample 2: Po+, control. Electrophoretic positions of different PRPs (see table 1 and Azen 1989) and amylase (Amy) are shown. The II-1 protein is not well resolved in this system.

show characteristic Con and Po variant types (fig. 1). Terminology for *PRP* DNA variants in this paper is based on genomic Southern analysis of *Eco*RI-digested DNAs with the *Hin*fI 980 PRP probe, derived from the sequences in exon 3 of *PRB*1 (Azen et al. 1984) and which cross-hybridizes with exon 3 of all six *PRP* genes (Maeda 1985). The *PRB*1, *PRB*2, and *PRB*4 DNA variants show characteristic length differences designated "small" (S), "medium" (M), and "large" (L) (Lyons et al. 1988*a*, 1988*b*) as well as typical restriction maps (Azen et al. 1992; Kim et al. 1993). For this paper, the allele designations have been modified to indicate the particular mutations discussed for the Con and Po proteins, and these alleles include *PRB*1M CON2+, *PRB*2L CON1+, *PRB*2L CON1−, and *PRB*4M PO−.

Recombinant bacteriophage clones in charon 40 libraries (Dunn and Blattner 1987) were detected by hybridization to the *Hin*fI 980 PRP probe (fig. 2*d*). The *PRB*1, *PRB*2, and *PRB*4 genes (figs. 2*c*, *a*, and *b*, respectively) were cloned as characteristic 17-, 15-, and 18-kb genomic fragments from Charon 40 libraries made from size-selected *Hin*dIII-(for *PRB*1) or *Bam*HI- (for *PRB*2 and *PRB*4) digested genomic DNAs. Subclones in plasmid Bluescript were prepared for sequencing. Subclones include (see restriction maps in fig. 2): for *PRB*1, 4-kb

*Bam*HI/*Eco*RI and 1-kb *Hin*fI fragments; for *PRB2*, 2-kb *Xba*I/*Eco*RI and 1-kb *Hin*fI fragments; and for *PRB4*, a 3-kb *Hin*dIII/*Eco*RI fragment. The subcloned fragments were deleted with exonuclease III and mung-bean nuclease to generate overlapping series of fragments that entirely spanned exons 3, the major coding portions of *PRP* genes. Two deletion series (one for each strand) were prepared for each *PRB1* and *PRB2* allele. One series initiated from the 3′ *Eco*RI site and the other from the 5′ *Hin*fI site. A single deletion series was prepared for the noncoding strand of the *PRB4* allele from the 3′ *Eco*RI site. Double-stranded DNA sequencing of the exon 3 regions (fig. 2, hatched boxes below genes) was done by using Sequenase (USB) and reverse or M13-20 primers by dideoxy-chain termination (Sanger et al. 1977). For *PRB1* and *PRB2* alleles, both DNA strands were sequenced once, and the noncoding strand of the *PRB4* allele was sequenced twice. The intron 3 5′-donor splice site mutation of the *PRB4* allele was further confirmed by sequencing a portion of the coding strand from a primer 5′ and adjacent to the mutation (fig. 3*C*, primer shown by interrupted underline). Sequence data were analyzed using software provided by the Genetics Computer Group, Inc.

For gene conversion analysis of the *PRB*1M Con2+ allele, the exon 3 regions of several *PRP* genes were sequenced by double-stranded sequenase methods, and the regions downstream of exon 3 were sequenced by cycle sequencing (cyclist methods, Stratagene; sequitherm methods, Epicentre) of overlapping PCR fragments amplified from the DNAs of single-recombinant bacteriophage plaques as schematically shown in figure 4*A*. The PCR amplifications were done for 40 cycles of 1 min at 94°C, 2 min at 55°C, and 2 min at 72°C, and the PCR fragments include P1/P6 (1187 bp), P3/P6 (765 bp), P3/P7 (1255 bp), and P5/P7 (511 bp).

PCR and sequencing primers were based on published *PRB1* and *PRB2* gene sequences (Kim et al. 1993) and are numbered accordingly. The PCR and sequencing primers include (P1) 5′-CCAGTAATCTAGGATTC-3′ (3609–3625); (P2) 5′-GTGTTGGCAACAAGGAAG-3′ (3876–3893); (P3) 5′-CAGCAGCTAAATGGTGT-CTC-3′(4060–4079); (P4) 5′-GTGCAAGGGCAT-CCAAAGAG-3′ (4646–4665); (P5) 5′-CAGGAAGT-GAATAAGAAGATG-3′ (4828–4848); (P6) 5′-CATC-CTCTTATTCACTTCCTG-3′ (4848–4828); and (P7) 5′-GACTAGCTTGGCCAACATGG-3′ (5356–5337).

Gene Conversion Analysis of PRB*1M CON2+ Allele*

To determine the extent of the probable gene conversion between *PRB1* and *PRB2* genes to produce the *PRB*1M CON2+ allele, three DNA sequences were compared: *PRB*1L (Azen et al. 1993*a*), *PRB*2L CON1+ (this paper), and *PRB*1M CON2+ (this paper). The sequences compared extend ~2.5 Kb from the *Hin*fI (H) site just 5′ to exon 3 to ~200 bp 3′ to exon 4 (see fig. 2*a* and *c* and fig. 4*A*).

Electrophoretic Analysis of Salivary PRP Polymorphisms

The terminology for PRPs and their gene assignments is consolidated in table 1. For the Po (parotid "o" protein) and Con (concanavalin A binding) protein polymorphisms, Western blots from 15% SDS PAGE gels were stained with Amido Black or Con A, respectively (Azen and Yu 1984*a*, 1984*b*; Azen 1989). We modified the Con A stain in accordance with recommendations of the Vector Company and used Immobilon PSQ (Millipore) rather than the standard nitrocellulose to enhance sensitivity. For optimal identification of the II-1 glycoprotein (a PRP described by Kauffman et al. 1993), the polyacrylamide concentration in the separation gel was reduced to 10%, and the separated proteins were electrophoretically transferred to Immobilon PSQ for Con A staining.

N-Terminal Amino Acid Sequencing of Glycosylated Con Proteins

Glycosylated parotid salivary PRPs from two subjects were partially purified by Con A affinity chromatography according to published procedures (Ausubel et al. 1994). The Con2 protein was isolated from saliva of subject R.S. (Con1+/Con2+ type) and the Con1 protein from saliva of another subject, A.F., (Con1+ type).

The eluted salivary glycoproteins from affinity chromatography were separated by SDS-PAGE and transferred to polyvinylidene fluoride (PVDF) membrane for sequencing. Con1 and Con2 proteins were identified on the PVDF membrane by comparison to salivary proteins in adjacent lanes whose identities were established and verified by characteristic mobilities and lectin binding assays in separate experiments. The PVDF membranes prepared for sequencing of Con1 and Con2 proteins were not subjected directly to staining. Proteins were sequenced on a Perkin Elmer—Applied Biosystems Division model 477A automated protein sequencer with on-line model 120A phenylthiohydantoin (PTH)–amino acid analyzer equipped with chart recorder and 610A data analysis system. The PTH-amino acid standard was 1 pmol (0.002 absorbance units full scale). Special proline cycles were utilized for all cycles in sequencing of both Con1 and Con2 proteins.

Overlay Assay for Assessing Bacterial Binding to Glycoproteins

Bacterial receptors were detected using a modification of a whole cell overlay assay described elsewhere (Murray et al. 1992; Azen et al. 1993*b*). In brief, duplicate

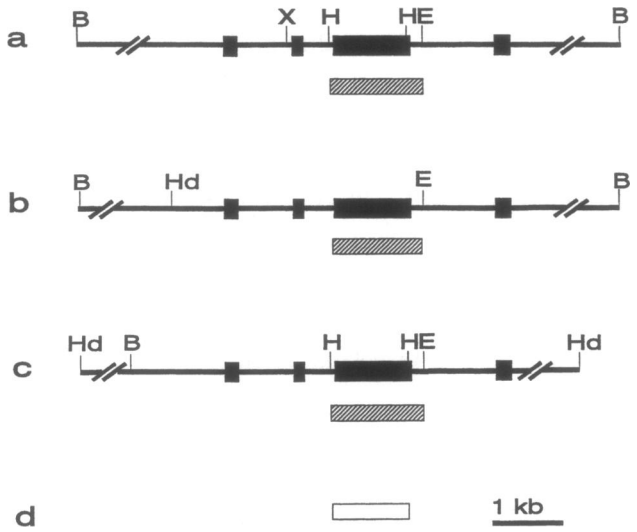

Figure 2 Restriction map and cloning strategy for *PRB*1, *PRB*2, and *PRB*4 genes. Exonuclease III deleted and sequenced regions covering exon 3 are shown below the lines as hatched boxes. Restriction sites used in cloning are shown. To facilitate presentation of the map, the interrupted regions between restriction sites are not drawn to scale, but other regions are drawn to scale. Restriction sites shown are *Bam*HI (B), *Xba*I (X), *Hin*fI (H), *Eco*RI (E), and *Hin*dIII (Hd). Panels *a*, *b*, and *c* depict prototype *PRB*2, *PRB*4, and *PRB*1 genes, respectively. Panel *d* depicts PRP *Hin*fI 980 probe covering exon 3 and used for cloning, subcloning, and genomic analysis.

samples of parotid saliva were electrophoretically separated (10% SDS-PAGE), and the proteins were transferred to immobilon membranes. One replica was stained with biotinylated lens culinaris agglutinin (or concanavalin A) to visualize the II-1 protein, and the other was overlaid with [^{35}S]methionine-labeled *S. sanguis* 10556. Salivary components that supported bacterial attachment were detected by autoradiography.

Results

Assignment of Con1 and Con2 Proteins to PRP Genes and Deduced Glycosylation

The N-terminal amino acid sequences of Con1 and Con2 were similar to those derived previously from analysis of the *PRB*2 gene (Kim et al. 1993). To determine the basis for the Con protein polymorphism, *PRB*2 alleles were cloned and sequenced from subjects R.S. (Con1+/Con2+ type) and L.S. (Con− type). The determined N-terminal 61 amino acids of the glycosylated Con1 protein match a coding region within the *PRB*2L CON1+ allele of subject R.S. and includes the region surrounding the first potential N-linked glycosylation site (NKS) (figs. 3*A* and 5*A*). Since asparagine (N) (underlined in fig. 5*A*) is not blocked (i.e., can be determined in the N-terminal amino acid sequence), this site is not glycosylated, indicating that the other NKS site downstream in the Con1 coding region is probably glycosylated (fig. 3*A*).

The determined N-terminal amino acid sequence of the Con2 protein differed from the Con1 protein at position 55 in both *PRB*2 alleles of subject R.S., so the Con2 protein must be encoded in another *PRP* gene. The Con2 protein is always associated with the PmF protein in saliva (Azen and Yu 1984*b*), known to be encoded by the *PRB*1 gene (Lyons et al. 1988*a*). Thus, *PRB*1 clones of subject R.S. were also studied. We have determined that subject R.S. (Con1+/Con2+ type) is a *PRB*1/2 fusion gene heterozygote (Azen et al. 1992; 1993*a*) with three *PRB*1-type copies that include *PRB*1M, *PRB*1/

Table 1

PRP Genes and Products[a]

Locus and Polymorphism	Biochemically Purified PRPs: Same as Polymorphic Proteins	Biochemically Purified PRPs: Other Proteins
*PRH*1:		
Db, Pa, and PIF[b]	Db, Pa, and PIF	IB-8b (P-C)
*PRH*2:		
Pr	PrI-IV	IB-8b (P-C)
*PRB*1:		
Pe	II-2	
PmS	IB-6 (P-I)	
PmF	IB-9 (P-E)	
Ps		
Con2[c]		
*PRB*2:		
Con1		IB-1, IB-4 (P-H), IB-8C (P-F), and IB-7 (P-G)
*PRB*3:		
Gl	I-A	CD-IIf
*PRB*4:		
Po	IB-5 (P-D)	CD-IIg
II-1	II-1	

[a] The provisional gene assignments of polymorphic PRPs and other PRPs are shown. In most cases, but not all (as Ps, Con1, and Con2 proteins), the polymorphic PRPs were shown to be electrophoretically the same as biochemically purified PRPs, and these alternative biochemical names are also given. The gene assignments of other PRP polymorphisms as Pmo1, Pmo2, P3, and Pc are unknown and are not shown. These data are based on previous studies (summarized by Azen and Maeda 1988; Azen 1989, 1993; Minaguchi and Bennick 1989; Kauffman et al. 1991, 1993; Azen et al. 1993*a*) and the present work.

[b] Historically, Db, Pa, and PIF proteins were designated as separate polymorphisms and the products of three different PRP genes, but later they were determined to be three alleles at the *PRH*1 locus.

[c] Con2, a "*PRB*2-like" protein, is encoded along with other "*PRB*2-like" proteins (IB-4 and IB-8c) in some *PRB*1 alleles by a probable gene conversion.

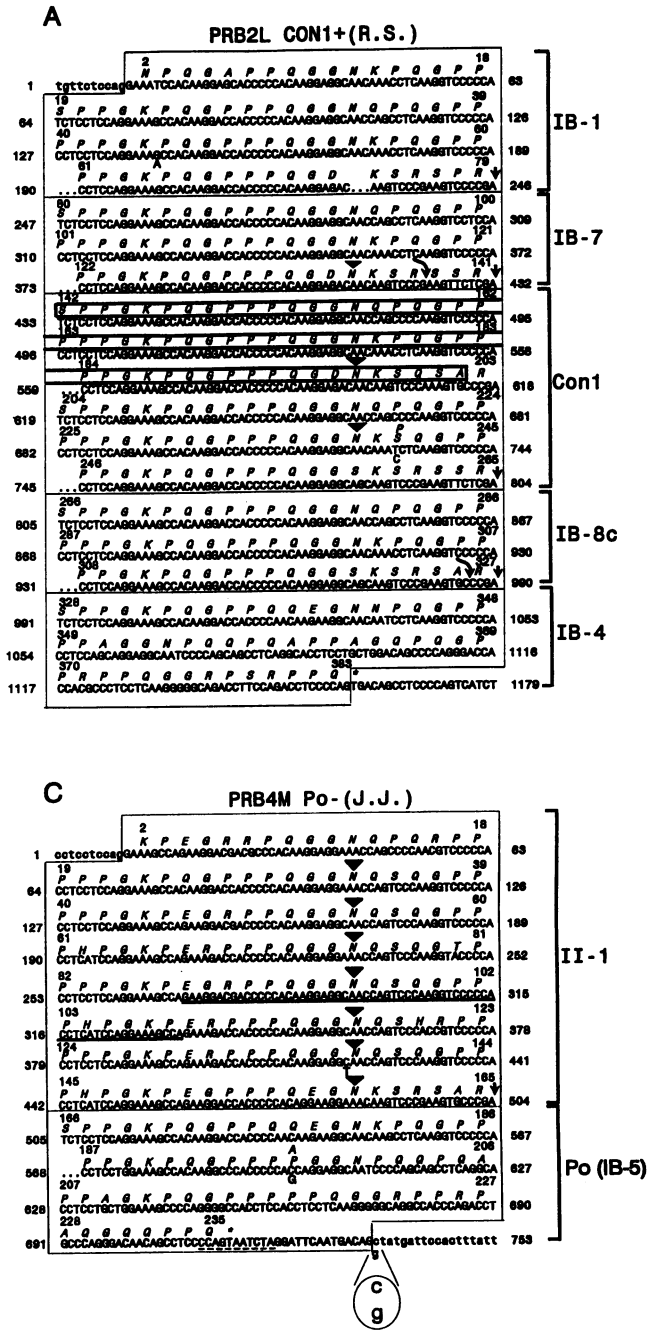

Figure 3 Nucleotide and deduced amino acids of exon 3 regions of *PRB*2L CON1+, *PRB*1M CON2+, and *PRB*4M PO−. Dots are introduced into nucleotide sequences to align repeats according to the method of Lyons et al. (1988*a*). Translated portions of exon 3 for *A* and *B* and the entire exon 3 for *C* are enclosed in the thin-lined boxes. The amino acids and nucleotides are numbered according to their positions in exon 3. Intron sequence is indicated by lowercase letters. The determined N-terminal amino acids of Con1 and Con2 proteins are enclosed in thick lines in *A* and *B*. Arrowheads represent potential N-linked glycosylation sites. For IB-1, Pe (II-2), and II-1 proteins in *A*, *B*, and *C*, respectively, the 17 N-terminal amino acids are encoded in exons 1 and 2. The localizations of PRPs within exon 3 are shown at the right margins in *A*, *B*, and *C* (for terminology, see table 1). Regions identical with the CD-IIg peptide sequences of Shimomura et al. (1983) within the provisional II-1 coding region in C are not shown. Straight arrows represent putative cleavage sites at basic-X-X-Arg sequences (Schwartz 1986). Curved arrows represent termination of some PRPs (IB-7 and IB-8C). *A*, *PRB*2L CON1+ allele (subject R.S.). The change at nt 730 that abolishes the potential N-linked glycosylation site and the silent change at nt 141 in the *PRB*2L CON1− allele are also shown. *B*, *PRB*1M CON2+ copy (subject R.S.). The potential N-linked glycosylation site is at amino acid 175. *C*, *PRB*4M PO− allele (subject J.J.). The exon 3 sequence is compared to that of a *PRB*4S allele from an individual with the Po+ phenotype (Kim et al. 1993). The mutation at the +1 position of the 5′ donor splice site in intron 3 is circled. The sequence of the primer used to verify the mutation on the coding strand is shown (interrupted underline). The continuous underline represents one extra repeat in the *PRB*4M PO− sequence compared to the *PRB*4S sequence (Kim et al. 1993).

2M, and *PRB*1S. The determined N-terminal 60 amino acids of the Con2 protein match a coding region within the *PRB*1M copy (here termed "*PRB*1M CON2+") and include the region surrounding the potential N-linked (NKS) glycosylation site (figs. 3*B* and 5*B*). Since the amino acid at the expected asparagine (N) position is undetermined (X underlined in fig. 5*B*), this site is presumably blocked and probably indicates glycosylation. The Con2 sequence is immediately downstream from the PmF sequence in the PRB1*M allele (fig.* 3B).

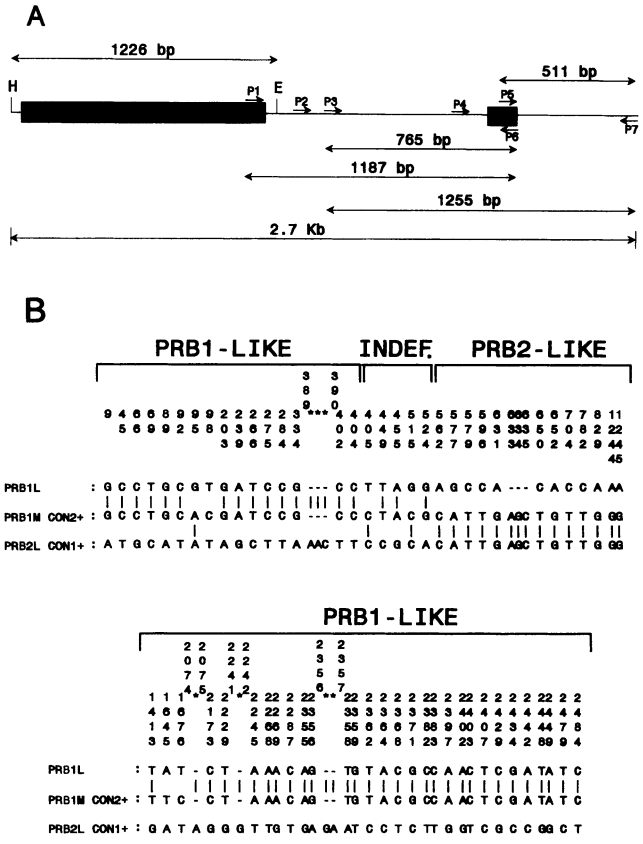

Figure 4 DNA sequence analysis of a probable conversion between *PRB*1 and *PRB*2 genes. *A*, Region of *PRP* genes compared. PCR and sequencing primers (P1–P7) and the PCR fragments analyzed are shown. H = *Hin*fI; E = *Eco*RI. Exons 3 and 4 are shown as black boxes. *B*, DNA sequence comparison of *PRB*1M CON2+ with *PRB1L* and *PRB*2L CON1+. The nucleotides of *PRB*1M CON2+ are numbered (shown vertically above sequence) beginning at the *Hin*fI (H) site. Only those nucleotides of *PRB*1M CON2+ that differ from one or both of the other two genes are shown. Each asterisk (*) indicates a nucleotide that is absent in the *PRB*1M CON2+ and *PRB1L* sequences but is present in the *PRB*2L CON1+ sequence. The positions of these gaps are shown by numbers that are offset and indicate the nucleotides immediately adjacent to the gaps. From the DNA comparison, the *PRB*1M CON2+ sequence is divided into four regions ("*PRB*1-like," "indefinite" "*PRB*2-like," and "*PRB*1-like"). The exon 3 sequence extends from nt 52–1043, and the exon 4 sequence from nt 2188–2290.

The Deduced Structure of the Con1 Protein and a PRB*2 Mutation That Eliminates the Proposed Potential N-linked Glycosylation Site*

When the exon 3 DNA sequence of the *PRB*2L allele of subject L.S. (Con− type) was compared with that of the *PRB*2L allele of subject R.S. (Con1+/Con2+ type) two nt changes were seen (fig. 3*A*). The first is a silent G→A transition upstream of the Con1 coding region at nt 141. The second is a single T→C transition at nt 730 causing a missense change [TCT(Ser)→CCT(Pro)] that abolishes the potential N-linked glycosylation site (NKS→NKP) (figs. 3*A* and 6*A* and *B*). The same two substitutions are also seen in the second *PRB*2L allele of subject R.S., who is thus a heterozygote for *PRB*2L CON1− and *PRB*2L CON1+ alleles. In figure 3*A* are also shown other *PRB*2 encoded PRPs [IB-1, IB-7, IB-8c, and IB-4] that have been biochemically purified and sequenced by others (summarized by Kauffman et al. 1991; see table 1). The N-terminal 17 amino acids of IB-1 are encoded by exons 1 and 2.

The Con2 Protein, a Product of a Probable Conversion between PRB*1 and* PRB*2 Genes*

From restriction analysis of the Lambda phage isolate, the *PRB*1M CON2+ copy is located on "*PRB*1-sized" 17-kb *Hin*dIII and 6.5-kb *Eco*RI fragments (Kim et al. 1993). However, exon 3 of the PRB1M CON2+ copy (fig. 3*B*) is a hybrid of *PRB*1 and *PRB*2 genes encoding "*PRB*1-like" proteins (Pe and PmF) and "*PRB*2-like" proteins (Con2, IB-8c, and IB-4). This conclusion is supported by a comparison of the *PRB*1M CON2+ DNA sequence with that of *PRB2L CON1+* and that of a previously published *PRB*1L gene (Azen et al. 1993*a*) (fig. 4).

To determine the extent of the putative gene conversion, all positions where the *PRB*1M CON2+ sequence differ from either one or both of the other two sequences were compared as shown in figure 4*B*. From this comparison, the sequence can be divided into several regions beginning just 5′ of exon 3 (nt 52–1043). The first two regions (labeled "*PRB*1-like" and "indefinite") are contiguous, and the next region ("*PRB*2-like") is bounded by 37-bp and 167-bp regions (not shown) in which the three DNA sequences are identical and are therefore not informative. The first region is "*PRB*1-like" (nt 9–404), and it includes a small portion of intron 2 and one-third of exon 3. *PRB*1M CON2+ resembles *PRB*1L in 17/18 and *PRB*2L CON1+ in 1/18 informative comparisons. In a second smaller region (nt 405–524) in exon 3 (labeled "indefinite"), *PRB*1M CON2+ resembles *PRB*1L in 3/5 and *PRB*2L CON1+ in 2/5 comparisons. The third region is "*PRB*2-like" (nt 562–1245), and it includes about one-half of exon 3 and a small portion (~200 bp) of intron 3. In all 15 comparisons, *PRB*1M CON2+ resembles *PRB*2L CON1+. The fourth region is "*PRB*1-like" (nt 1413–2484), and it includes most of intron 3, exon 4 (nt 2188–2290), and ~200 bp 3′ to exon 4. *PRB*1M CON2+ resembles *PRB*1L in all 32 informative comparisons. From this analysis, there is a probable gene conversion that encompasses at least a 684-bp "*PRB*2-like" region inserted into a "*PRB*1-type" gene.

A PRB*4 Intron 3 5′ Donor Splice Site Mutation that May Account for the Po− and Il-1− Protein Phenotypes*

The DNA sequence of a *PRB*4M PO− allele from subject J.J. (Po− type) was compared with that of a previously

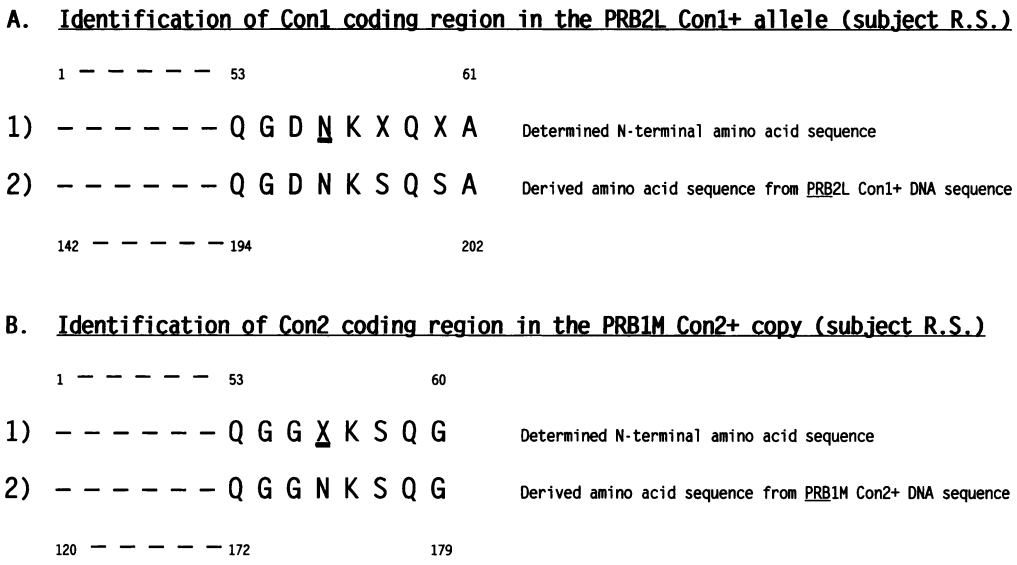

Figure 5 Comparison of determined N-terminal with derived amino acid sequences. The determined N-terminal amino acid sequences of Con1 and Con2 proteins match the derived amino acid sequences of the *PRB*2L CON1+ allele and *PRB*1M CON2+ copy of subject R.S. in exon 3. Only the amino acids surrounding the putative nonfunctional glycosylation site in the Con1 protein (*A*) and the potential glycosylation site in the Con2 protein (*B*) are shown. The single letter amino acid code is used (X means "undetermined").

reported *PRB*4S allele from an individual with the Po+ type (Kim et al. 1993) (fig. 3*C*). A G→C transversion was seen at the +1 invariant position (nt 736) of the intron 3 5′ donor splice site (figs. 3*C* and 6*C*). Two other single-nucleotide substitutions and a length difference were found. The first substitution is a T→C transition at nt 420 (a silent change), and the other is a G→C transversion at nt 598, causing a missense change [GCA(Ala)→CCA(Pro)]. In addition to the Po protein, the II-1 protein is encoded in exon 3 of *PRB*4 (the N-terminal 17 amino acids are encoded in exons 1 and 2) (Kauffman et al. 1993). From comparison of the determined II-1 amino acid sequence

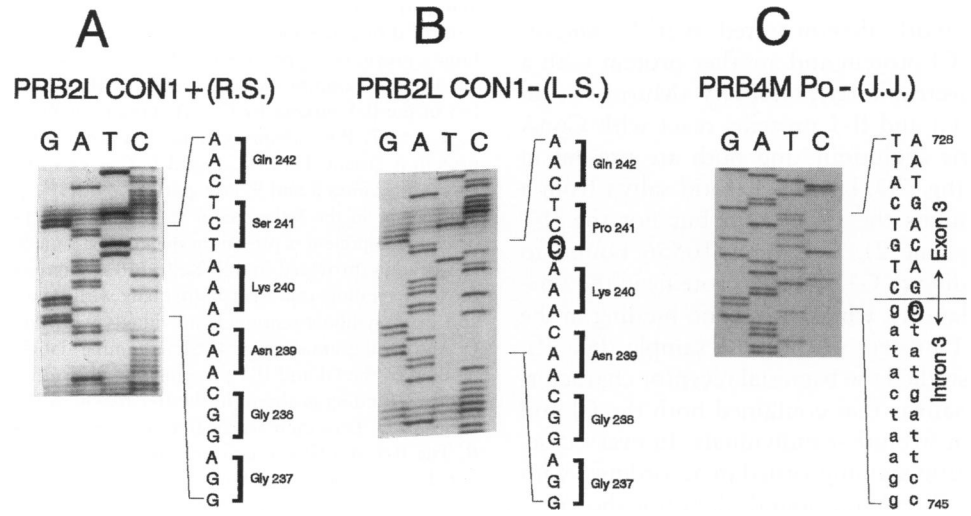

Figure 6 DNA sequence gel results and deduced amino acids. The deduced amino acids (*A* and *B*) and nucleotides (*C*) are numbered according to the sequences in fig. 3. *A*, Potential N-linked glycosylation site (NKS) of the Con1 protein encoded in the *PRB*2L CON1+ allele (subject R.S.). *B*, Same region as in *A*, showing the mutation (circled) that abolishes the potential N-linked glycosylation site of the *PRB*2L CON1− allele (subject L.S.). *C*, Intron 3 5′ donor splice site mutation of the *PRB*4M PO− allele (subject J.J.). The sequence shown is from the noncoding strand.

of Kauffman et al. (1993) with that derived from exon 3 of the *PRB*4M PO− allele (fig. 3*C*), there are four unambiguous single amino acid differences and an internal length difference of eight amino acids (comparison not shown). The eight additional amino acids in the *PRB*4M Po− allele include residues 97–104 in figure 3*C*. It is probable that these represent mainly polymorphic differences, since our studies and those of Kauffman et al. (1993) were done on different individuals.

The II-1 Protein is Absent From Saliva of Subject J.J. and the Subject's Family Members with the Po− Type

Since the *PRB*4 precursor protein encodes both the Po protein and the larger II-1 glycoprotein (fig. 3*C*), the II-1 protein may also be missing in individuals with the Po− type. The saliva samples from this family were previously typed for the Po protein polymorphism (Azen and Yu 1984*a*) and are here typed for the presence (+) or absence (−) of the II-1 protein on a Con A–stained Western blot (figs. 7*A* and *B*). In Po+ individuals, the II-1 glycoprotein appears as two major bands that comigrate with purified II-1 protein provided by P. J. Keller (fig. 7*A*, lane 10). The parents (fig. 7*A*, lanes 8 and 9) are both Po+ and II-1+ and are presumed Po+/Po− heterozygotes. For an unknown reason, the mother's sample (fig. 7*A*, lane 9) shows the faster migrating II-1 component but not the slow one. Four offspring (samples in lanes 1, 3, 5, and 7) are both Po− and II-1− types, and three offspring (samples in lanes 2, 4 and 6) are both Po+ and II-1+ types.

[^{35}S]*Methionine-labeled* S. sanguis *10556 Binding to Salivary Proteins of Different Po and II-1 Types*

Our previous work demonstrated that *S. sanguis* 10556 bound to Gl protein and another protein with a slightly higher electrophoretic mobility (Murray et al. 1992). Both the Gl and II-1 proteins react with ConA and Lens culinaris agglutinin, and both are present in the Po+ sample (fig. 7*D*, lane 1). Parotid saliva from a Po− sample contains the Gl protein but not the II-1 protein (fig. 7*D*, lane 2). *S. sanguis* 10556 bound to bands corresponding to G1 and II-1 proteins in the Po+ sample (fig. 7*C*, lane 1), while there is no binding in the region of the II-1 protein in the Po− sample (fig. 7*C*, lane 2). We also studied the bacterial receptor characteristics of parotid salivas that contained both the Gl and II-1 proteins from five other individuals. In every case, the intensity of lectin staining varied in accordance with the number of [^{35}S]-bacteria bound (data not shown).

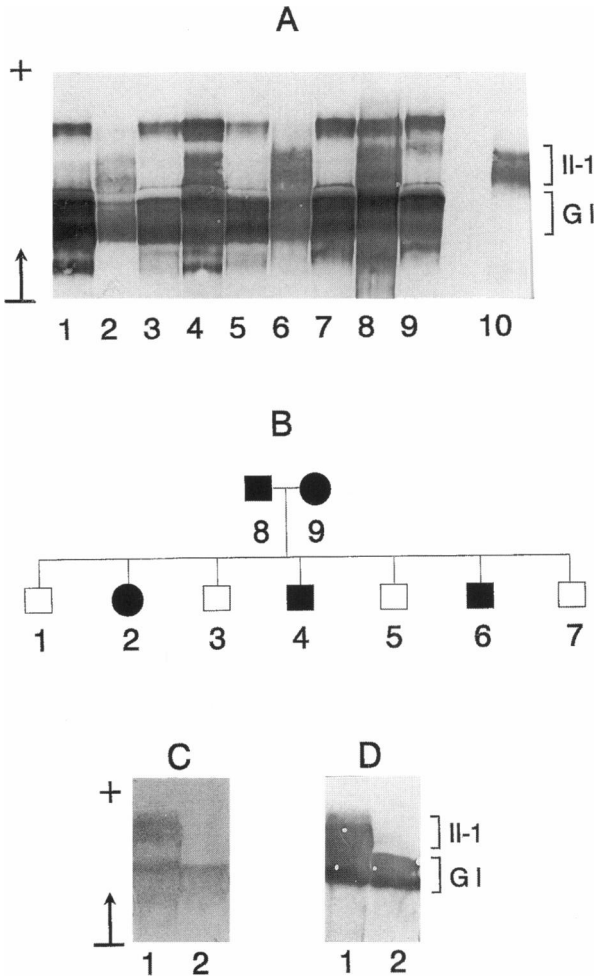

Figure 7 Family study showing association of Po and II-1 types, and [^{35}S]methionine-labeled *S. sanguis* 10556 binding to salivary proteins of different Po and II-1 types. *A*, Parotid saliva samples from a family previously typed for the Po protein polymorphism (Azen and Yu 1984*a*). Samples are here typed for the presence (+) or absence (−) of the II-1 protein by Con A staining of Western blots. Lanes 1, 3, 5, and 7, Po− offspring (lane 1 is from proband J.J.). The II-1 protein is absent. Lanes 2, 4, and 6, Po+ offspring. The II-1 protein is present. Lanes 8 and 9, Po+ parents. Both II-1 protein components are present in the father (lane 8), but only the faster migrating II-1 protein component is present in the mother (lane 9). Lane 10, Purified II-1 protein (provided by P.J. Keller). *B*, Pedigree of family shown in *A*. Solid symbols represent individuals with Po+ and II-1 (+) types, and open symbols represent individuals with Po− and II-1− types. *C*, Autoradiogram showing [^{35}S]methionine–labeled *S. sanguis* 10556 binding to the Gl and II-1 proteins in the Po+/II-1 (+) saliva sample (lane 1). Binding is absent in the II-1 region of the Po-/II-1− sample-(lane 2). *D*, Lens culinaris agglutinin staining of same samples as in *B*. The II-1 and Gl proteins are present in the Po+ sample (lane 1), but the II-1 protein is absent in the Po− sample (lane 2).

Discussion

Among the many salivary PRP polymorphisms, the structure and gene coding of the Con proteins has been enigmatic. The Con proteins were previously determined to be glycosylated PRPs, on the basis of immunological reactivity to anti-PRP sera, Con A reactions on western blots, and linkage of their genetic determinants to the *PRP* gene cluster (Azen and Yu 1984*b*). The Con1 and

Con2 proteins do not correspond electrophoretically, or by derived protein sequence reported here, to other purified PRPs (see table 1).

We have now shown that a 124 aa Con1 glycoprotein is encoded in a *PRB*2L CON1+ allele. The coding region is bounded by basic-X-X-arg sequences common to monobasic proteolytic cleavage sites (Schwartz 1986) such as those used for other PRPs encoded by *PRB*1, *PRB*2, and *PRB*4 genes (see fig. 3; Lyons et al. 1988*a*; Azen et al. 1993*a*). The N-terminus for the Con1 protein was established by the determined amino acid sequence here reported, but the C-terminal sequence has been deduced only from the DNA sequence and is thus less certain, since further processing may occur in the saliva and/or salivary glands. On the basis of the protein and DNA sequence comparisons (figs. 3*A* and 5*A*), we postulate that a single potential N-linked glycosylation site (NKS) is used in the Con1 protein so that the Con1− protein results from a mutation at the potential glycosylation site.

The "*PRB*2-like" Con2 glycoprotein is encoded by the *PRB*1M CON2+ allele. The region encodes 82 amino acids that include a potential N-linked glycosylation site (NKS) that is probably used (figs. 3*B* and 5*B*) and is bounded by basic-X-X-arg potential cleavage sites; the IB–8C coding region occurs immediately adjacent to the postulated downstream cleavage site (fig. 3*B*). The N-terminus of the Con2 protein was established by the amino acid sequence determination, but the deduced C-terminus is uncertain, because further processing may occur.

The protein sequences immediately surrounding the potential N-linked glycosylation site that is probably used in the Con1 and in the Con2 proteins are identical (GGNKSQG) and differ from the apparently nonfunctional glycosylation site in the Con1 protein (GDNKSQS). The presence of the consensus sequence N-X-S/T (where X denotes any amino acid) does not always lead to glycosylation, but, other than proline residues in positions X and Y of the sequence N-X-S/T-Y, there are no obvious local sequence features that correlate with presence or absence of N-linked glycosylation (Gavel and von Heijne 1990). Perhaps the substitution of the charged aspartic (D) for the nonpolar glycine (G) immediately adjacent to the asparagine (N) interferes with glycosylation at the nonfunctional NKS site in the Con1 protein. In a *PRB*2L CON1− allele, a single nt change at the potential N-linked glycosylation site causes a missense change (NKS→NKP) that abolishes glycosylation and appears to be the basis for the Con− phenotype. We do not know if the nonglycosylated Con1 protein is present in salivas of the Con− type.

The "*PRB*2-like" Con2 protein is encoded in a *PRB*1M-type allele (*PRB*1M CON2+), and this hybrid allele probably represents a conversion encompassing ≥684 bp of DNA between *PRB*1 and *PRB*2 genes, although double recombination is also possible. The subject R.S. with the *PRB*1M CON2+ allele is also a *PRB*1/2 fusion gene heterozygote with three *PRB*1-type copies due to intergenic homologous and unequal recombination between *PRB*1 and *PRB*2 genes (Azen et al. 1992, 1993*a*). Gene conversion has been well described in other human genes, such as those coding for globins, histocompatibility proteins, immunoglobulins, and steroid 21-hydroxylase (reviewed by Cooper and Krawczak 1993). In mammalian cell systems, gene conversion can accompany reciprocal recombination (reviewed by Bollag et al. 1989). Perhaps a similar concurrence may be operative in subject R.S., who shows both gene duplication and gene conversion involving the very homologous and contiguous *PRB*1 and *PRB*2 genes (Kim et al. 1990). The probable gene conversion here described should particularly lend itself to further studies, since it is accompanied by an easily identifiable phenotypic marker, e.g., the presence in saliva of the Con2 protein.

Since the PmF protein (but not the PmS protein) is encoded in the *PRB*1M CON2+ allele, two related associations are now explained. First, previous population studies showed a strong association between PmF+ and Con2+ types. Thus, all Con2+ salivas were also PmF+ (Azen and Yu 1984*b*). Second, in an unpublished analysis of the above population data, there was a strong association of the Con2+ type with the PmF+/PmS− minority subtype of PmF+ (~25%) but not with the majority subtype (PmF+/PmS+) that constitutes the rest. The strong association of the Con2 type with the PmF+/PmS− minority subtype is now explained, since the *PRB*1M CON2+ allele encodes both PmF and Con2 proteins but not the PmS protein. In contrast, the *PRB*1S allele encodes both the PmF and PmS proteins and is associated with the PmF+/PmS+ type (Lyons et al. 1988*a*).

The polymorphic Po protein (Po+ and Po− types) was shown to be a PRP by reactivity to anti-PRP sera and by linkage of its genetic determinant to the *PRP* gene cluster (Azen and Yu 1984*a*). The Po protein has the same electrophoretic mobility as the purified and biochemically identical PRPs IB-5 and P-D (Azen 1989; Kauffman et al. 1991). In the studies here reported, a single nt change (G→C) at the +1 invariant position of the intron 3 5′ donor splice site in the *PRB*4M PO− allele appears to be responsible for the Po-/II-1− phenotype. Such mutations generally cause profound and somewhat unpredictable effects, such as absent normal mRNA, exon skipping, and cryptic splice site utilization (Krawczak et al. 1992). From electrophoretic studies, there was no evidence for the presence of protein products from the mutated *PRB*4 allele.

In our previous electrophoretic studies of the Po protein polymorphism, we did not recognize that the absence of the II-1 protein accompanied the Po− type (Azen and Yu 1984*a*). With improved electrophoretic techniques and with recent data that assign the Po protein and the II-1 protein to the *PRB*4 gene (Lyons et al. 1988*a;* Kauffman et al. 1993), we showed that the II-1 protein was absent from saliva samples of the Po− type. The association of the Po− and the II-1− phenotypes with the *PRB*4 splice site mutation also confirms that the II-1 protein is coded by the *PRB*4 gene (Kauffman et al. 1993).

Numerous studies report the bacterial binding properties of PRPs. PRPs are known to selectively bind to hydroxyapatite and to acquired enamel pellicle (reviewed by Hay et al. 1994). Acidic PRPs bound to hydroxyapatite differ in their adherence to a variety of potential intraoral pathogens. The main difference resides in the carboxy-terminal di-peptide (Schluckebier et al. 1992). Thus, different forms of acidic PRPs vary in their promotion of bacterial adherence, and these differences may be reflected in the initial bacterial colonization that is important in dental plaque formation.

Recent in vitro work described bacterial receptor binding properties of human salivary proline-rich glycoproteins (Gillece-Castro et al. 1991; Murray et al. 1992) and the effects of null mutations on binding (Azen et al. 1993*b*). In the absence of the Gl protein due to null mutations of the *PRB*3 gene, *Fusobacterium nucleatum* (a periodontal pathogen) does not interact with saliva in vitro. Previous in vitro work demonstrated that [^{35}S]methionine-labeled *S. sanguis* 10556 binds to the Gl protein and to a second protein with slightly higher electrophoretic mobility (Murray et al. 1992). We here now show that the second protein is II-1, which is absent in a Po-/II-1− individual. The result suggests that in most individuals both Gl and II-1 proteins act as receptors for *S. sanguis* 10556. The adherence of streptococci, particularly *S. sanguis* and *S. mitis,* to the salivary components that coat the tooth surface is believed to be one of the early steps in the formation of dental plaque (reviewed by Murray et al. 1992). Perhaps the genetically determined presence or absence of these salivary glycoproteins is important in establishing differences in the intraoral microbial environment and modulates intraoral disease susceptibilities to dental caries and/or periodontal disease.

Acknowledgments

We greatly appreciate the technical assistance of J. Walent. P. J. Keller kindly provided purified II-1 protein. This work was supported by the National Institute of Health grants 5-R37-DE03658-30 (to E.A.) and DE07244 (to S.F.). We thank L. Sabatini and R. Pauli for critical reading of the manuscript.

References

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K, et al (eds) (1994) Con A-Sepharose affinity chromatography. In: Current protocols in molecular biology, Vol 3, Suppl 21:17.1.5

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

——— (1993) Genetics of salivary protein polymorphisms. Crit Rev Oral Biol Med 4:479–485

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 Pa, Db, and PIF. Am J Hum Genet 41:1035–1047

Azen EA, Latreille P, Niece RL (1993*a*) *PRB*1 gene variants coding for length and null polymorphisms among human salivary Ps, PmF, PmS and Pe proline-rich proteins (PRPs). Am J Hum Genet 53:264–278

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 (Gl8) and a null in an Ashkenazi Jew. Am J Hum Genet 47:686–697

Azen EA, O'Connell PO, 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 (1993*b*) *PRB*3 null mutations result in absence of the proline-rich glycoprotein Gl and abolish *Fusobacterium nucleatum* interactions with saliva in vitro. Infect Immun 61:4434–4439

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

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

Cooper DN, Krawczak M (1993) Single base-pair substitutions. In: Cooper Dn, Krawczak (eds) Human gene mutation. Bios Scientific Publishers, Oxford, pp 150–152

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

Gavel Y, von Heijne G (1990) Sequence differences between glycosylated and nonglycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. Protein Eng 3:433–442

Gillece-Castro BL, Prakobphol A, Burlingame AL, Leffler H, Fisher SJ (1991) Structure and bacterial receptor activity of a human salivary proline-rich glycoprotein. J Biol Chem 266:17358–17368

Hay DI, Ahern JM, Schluckebier SK, Schlesinger DH (1994) Human salivary acidic proline-rich protein polymorphisms and biosynthesis studied by high-performance liquid chromatography. J Dent Res 73:1717–1726

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 DL, Keller PJ, Bennick A, Blum M (1993) Alignment of amino acids and DNA sequences of human proline-rich proteins. Crit Rev Oral Biol Med 4:287–292

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. Mamm 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

Krawczak M, Reiss J, Cooper DN (1992) The mutational spectrum of single base substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet 90:41–54

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

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

Murray PA, Prakobphol A, Lee T, Hoover CI, Fisher SJ (1992) Adherence of oral streptococci to salivary glycoproteins. Infect Immun 60:31–38

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

Schluckebier SK, Hay DI, Gibbons RJ, Schlisinger DH, Ferland MJ (1992) Effects of structural variations in carboxy-terminal residues of human salivary proline-rich proteins (PRPs) on adhesion of *Actinomyces viscosis* LY7 (abstract) J Dent Res 71:762

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

Shimomura H, Kanai Y, Sanada K (1983) Amino acid sequences of glycoprotein of human parotid saliva. J Biochem 93:857–863