Am. J. Hum. Genet. 47:686–697, 1990

Alleles at the *PRB3* Locus Coding for a Disulfide-bonded Human Salivary Proline-rich Glycoprotein (Gl 8) and a Null in an Ashkenazi Jew

E. A. Azen,*† K. Minaguchi,‡ P. Latreille,* and H.-S. Kim§

*Laboratory of Genetics and †Department of Medicine, University of Wisconsin, Madison; ‡Department of Forensic Odontology, Tokyo Dental College, Chiba City, Japan; and §Department of Pathology, University of North Carolina, Chapel Hill

Summary

From electrophoretic analysis, we identified in the saliva of an Ashkenazi Jew a disulfide-bonded major glycoprotein variant (Gl 8) that is a product of the proline-rich protein (PRP) locus *PRB3*. A previous study of this variant protein misidentified it as Pa 2 and as a product of a different *PRP* locus. The other *PRB3* allele in this individual is an apparent null. To identify the mutations, we sequenced the tandemly repetitious exon 3 (the major protein-coding portions) of both alleles. A CGT→TGT (Arg→Cys) mutation was found in one allele (*PRB3*S^{cys}), which accounts for the disulfide-bonded and peroxidase-modifying properties of Gl 8. A single nucleotide insertion was found in the other allele (*PRB3*M^{null}) that leads to a frameshift with a premature termination codon that causes an apparent lack of gene expression. Null alleles are frequent at *PRP* loci coding for basic and glycosylated PRPs, and the mechanism described might explain other null phenotypes among PRPs. From nucleotide comparisons, a model of intragenic unequal crossing-over is proposed to explain, in part, the generation of the *PRB3*M^{null} allele. The Gl 8 protein variant is found in Ashkenazi Jews (gene frequency around .008) but not in the general white, black, or Japanese populations. It is interesting that products of different *PRP* genes, Gl 8 from *PRB3* and Pa 1 from *PRH1*, are both disulfide bonded and probably modify salivary peroxidase (part of an important intraoral antibacterial system) through formation of disulfide-bonded heterodimers.

Introduction

The human *PRP* gene family consists of six closely linked genes on human chromosome 12p13.2 that code for acidic, basic, and glycosylated proline-rich proteins (PRPs) (reviewed by Azen and Maeda 1988; Minaguchi and Bennick 1989). Two genes code for acidic PRPs (*PRH1* for Db, Pa, and PIF, and *PRH2* for Pr), one gene (*PRB3*) codes for the major salivary glycoprotein (Gl), and two genes (*PRB2* and *PRB4*) code for other basic and glycosylated PRPs. We previously described a disulfide-bonded salivary protein variant, termed Pa 2, in an Ashkenazi Jewish family (Azen 1977). This uncommon protein variant showed some properties characteristic for the acidic PRP termed Pa 1 (now known to be coded by the *PRH1* locus). These properties included similar electrophoretic mobilities in acid-urea starch gel, disulfide bonding of subunits, and probable modification of salivary peroxidase through the formation of disulfide-bonded heterodimers. The modified forms of salivary peroxidase were converted to the unmodified form after treatment of the saliva with disulfide reduction (Azen 1977). Thus, in these early studies, we assumed that Pa 1 and Pa 2 proteins were coded by alleles at the same *PRP* locus.

However, recent molecular genetic studies indicated that this interpretation was probably incorrect. Maeda (1985) postulated, on the basis of reanalysis of family and population genetic data, that the Pa, Db, and PIF proteins are coded by alleles at the same locus rather than by separate loci, and this hypothesis was confirmed by molecular analysis (Azen et al. 1987). This led us

Received March 5, 1990; revision received April 18, 1990.

Address for correspondence and reprints: Edwin A. Azen, M.D., Department of Medical Genetics, University of Wisconsin, 214 Genetics Building, Madison, WI 53706.

0002-9297/90/4704-0014$02.00

to reexamine the acidic PRP phenotypes in the family with the Pa 2 protein. We noted that salivas of some individuals showing the Pa 2 protein also showed the Db and Pa 1 proteins (the latter two proteins coded by alleles at the *PRH*1 locus). Thus, it seemed unlikely that the Pa 2 protein is also coded by the *PRH*1 locus. The simplest interpretation of the data (although there are other possibilities) is that the Pa 2 protein is coded by another locus, perhaps a *PRP* locus. Therefore, we restudied the saliva of an individual with the Pa 2 protein from this family and used a variety of more recently devised electrophoretic techniques (summarized by Azen [1989]) in order to correctly identify and assign the Pa 2 protein to a specific locus.

In this paper, we report that, in the individual investigated (A.F.), one allele at the *PRB*3 locus codes for a disulfide bonded Gl protein variant (termed Gl 8 for reasons to be described later) and the other allele is unexpressed (null). The mutations are found on two different polymorphic length *PRB*3 alleles, and to identify the mutations we sequenced the tandemly repetitious exon 3, which is the major coding portion of the alleles. We identified a CGT→TGT (Arg→Cys) mutation in one allele (termed *PRB*3S^{cys}) that accounts for the disulfide-bonded and peroxidase-modifying properties of Gl 8. We also found a single nucleotide insertion in the other allele (termed *PRB*3M^{null}) that leads to a frameshift with a premature termination codon that might cause apparent lack of gene expression. We found the Gl 8 protein variant at a low frequency in Ashkenazi Jews but not in the general white, black, or Japanese populations. From nucleotide comparisons of *PRB*3 alleles, a model of intragenic homologous and unequal crossing-over is proposed to explain, in part, the generation of *PRB*P3M^{null}.

Methods

1. *Studies of Native and Disulfide-reduced Gl Proteins in Acid/Lactate Polyacrylamide Gels*

Parotid salivas were collected as described elsewhere (Azen and Denniston 1974). After electrophoresis of parotid saliva in the acid/lactate gel system followed by staining with the Schiff reagent, only the major salivary glycoprotein (Gl) is detectable as a heavily stained band(s) of characteristic mobility (Azen et al. 1979). To test for disulfide reduction of variant Gl proteins, parotid saliva samples that were dried and resuspended in 1 × vol of 0.2 M Tris/HCl, pH 8.8, were first reduced with 1 mM dithioerythritol at 37°C for 1 h, then alkylated with 2.2 mM iodacetamide at 23°C for 1 h (Azen 1978). Incubated control samples were treated with 2.2 mM iodacetamide but not with dithioerythritol. The native and disulfide-reduced samples were then dialyzed, neuraminidase treated, concentrated 10×, and electrophoresed in 0.03 M Tris-lactate/lactic acid, pH 2.4 polyacrylamide gels and stained for carbohydrate with the Schiff reagent as described elsewhere (Azen et al. 1979).

2. *Studies of Native and Disulfide reduced Gl Proteins in SDS Polyacrylamide Gels*

To prepare native Gl proteins for electrophoresis, parotid saliva samples were dried, reconstituted in 1 × vol of SDS sample buffer without 2-mercaptoethanol and then boiled for 3–5 min (modified from Laemmli [1970]). To test for disulfide reduction of variant Gl proteins, the salivas were prepared as described above except that 2-mercaptoethanol was included in the sample buffer. To better separate the higher-molecular-weight Gl proteins which migrate close to the origin of the gel (Azen et al. 1979), the concentration of polyacrylamide was reduced in the running gel from the usual 15% to 8.25% and the gels were prepared according to the method of Laemmli (1970). The samples were electrophoresed and electrophoretically transferred to nitrocellulose as described elsewhere (Azen and Yu 1984). The Gl proteins were stained with amido black (Towbin et al. 1979) or with concanavalin A (Con A), which is much more sensitive for detecting the heavily glycosylated Gl proteins (Azen and Yu 1984). This SDS gel technique combined with Con A staining was used for screening saliva samples of populations for the slower-mobility Gl protein variants (including the disulfide-bonded Gl 8 variant herein described). This technique was more convenient for screening than the standard acid/lactate gel (Azen 1979) for several reasons: there was better separation of the slower-mobility Gl proteins; smaller amounts (about one-tenth) of saliva were required; and the Con A method was more sensitive than staining with the Schiff reagent. With our standard technique it was usually difficult to separate the faster-mobility Gl variants (as Gl 1, Gl 2, and Gl 3) from overlapping proteins that stain with Con A. However, in some experiments, the stain could be made less generally sensitive but more specific for Gl proteins by reducing the amount of Con A in the staining reaction approximately 100-fold from standard conditions. Under these modified conditions, the glycoproteins that overlap the faster-mobility Gl variant proteins are only faintly stained.

3. *Studies of Salivary Peroxidase Phenotypes*

Parotid saliva samples were electrophoresed in acid/lactate polyacrylamide gels and stained for peroxidase with *p*-phenylenediamine and hydrogen peroxide as described elsewhere (Azen 1977).

4. *Studies of Acidic PRP Phenotypes among Pr, Db, Pa, and PIF Proteins*

Parotid saliva samples were electrophoresed in pH 3.5–5.2 isoelectric focusing polyacrylamide gels and the bands visualized by precipitation with 20% trichloroacetic acid as described elsewhere (Azen and Denniston 1981; Azen 1989).

5. *Strategy for Cloning PRB3 Alleles from Subject A.F. and the GI 8 Disulfide-bonded Variant*

From the restriction map (H.-S. Kim, unpublished data), the entire *PRB3* gene is contained on a 5.4-kbp *Bam*HI fragment (fig. 1*a*, *b*), and among the six *PRP* genes, only the *PRH2* gene is of similar size, thus allowing for size fractionation and considerable purification of the *PRB3* alleles from the genomic DNA prior

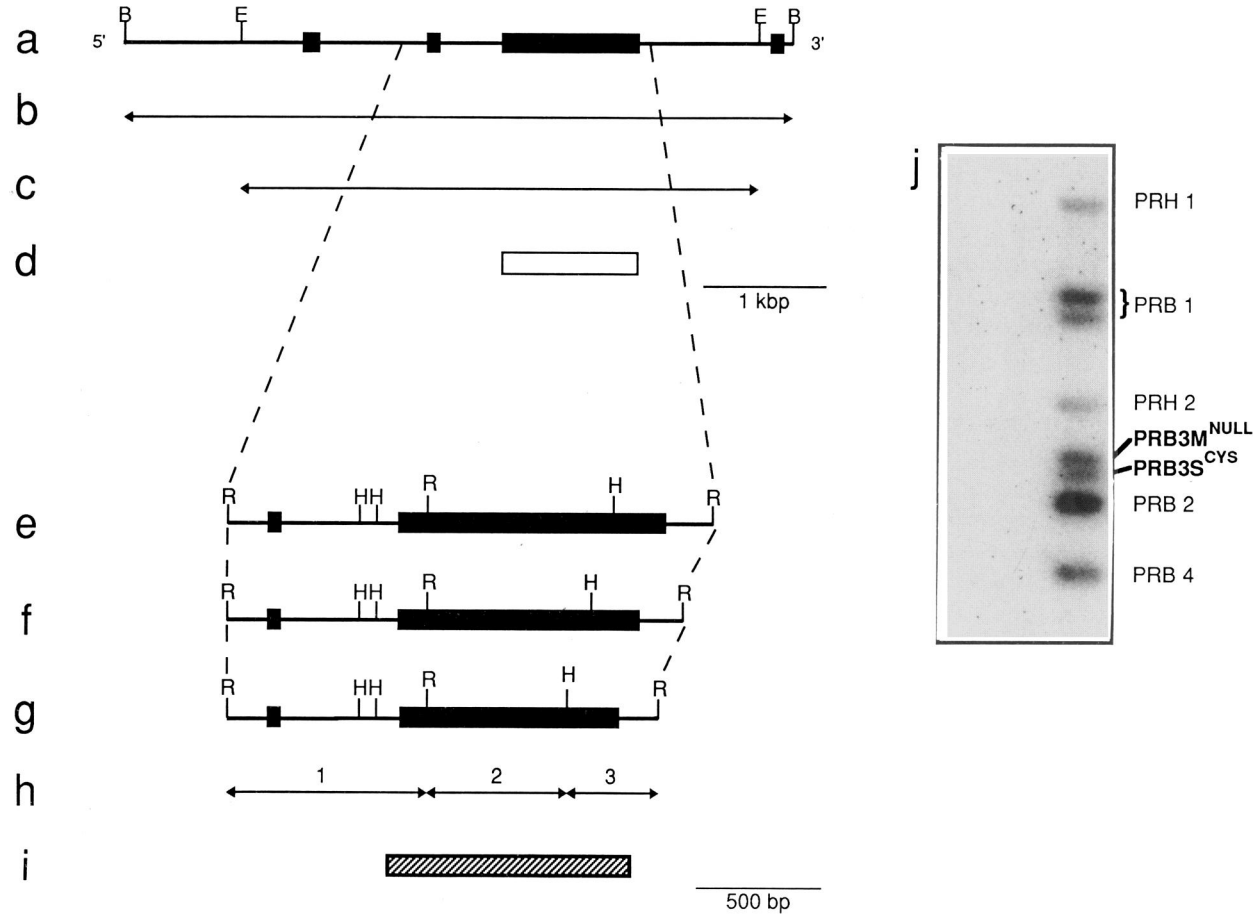

Figure 1 Restriction map of *PRB3* alleles and cloning strategy. *a*, Restriction map of the *PRB3*L allele. The four exons are shown as solid black boxes (H.-S. Kim, unpublished data). *b*, Approximately 5.4-kbp cloned *PRB3* fragments, obtained from the *Bam*HI (B) genomic library of subject A.F. (see text). *c*, An approximately 4.2-kbp *Eco*RI (E) genomic fragment for *PRB3*, shown here and in the genomic Southern result (*j*). *d*, The PRP probe, *Hin*fI 980, that was used for cloning, subcloning, and genomic Southern analysis. *e*, Map of the exon 3 region from the *PRB3*L allele (Lyons et al. 1988*a*, 1988*b*). To obtain the entire exon 3 which encodes the tandem proline-rich protein repeats, *Rsa*I (R) fragments were subcloned into Bluescript plasmid. *f*, Map of the exon 3 region from the *PRB3*M^{null} allele of subject A.F. (see text). *g*, Map of the exon 3 region from the *PRB3*S^{cys} allele of subject A.F. (see text). *h*, Subcloned fragments for sequencing using *Rsa*I (R) and *Hae*III (H) sites (illustrated for *PRB3*S^{cys}). *i*, The regions sequenced for *PRB3*M^{null} and *PRB3*S^{cys}, which cover the entire exon 3 (illustrated for *PRB3*S^{cys}). *j*, Southern genomic blot of *Eco*RI-digested DNA from subject A.F. hybridized to the probe, *Hin*fI 980. Bands representing the six *PRP* genes are indicated on the right. *PRB1* and *PRB3* show closely spaced doublets due to length polymorphisms.

to cloning into lambda phage. Lambda phage clones of the *PRB3* type are easily distinguished from those of the *PRH2* type by fragments of characteristic size that hybridize to the *Hin*fI 980 PRP probe (fig. 1*d*). The *Hin*fI 980 PRP probe (fig. 1*d*) hybridizes to exon 3 of all six PRP genes. Since the *PRB3* gene in subject A.F. with the disulfide-bonded Gl 8 variant showed a deletion/insertion-type length polymorphism on genomic Southern blots probed with the *Hin*fI 980 PRP probe (figs. 1*c* and 1*j*), we could easily distinguish the fragments from the two alleles (termed *PRB3*S^{cys} and *PRB3*M^{null} for reasons to be discussed later) based on a polymorphic size difference of about 200 bp. To search for mutations in the *PRB3* alleles of subject A.F., we subcloned and sequenced only the exon 3 portions of the two alleles, since the bulk of the more than 200 amino acids in the various polymorphic Gl proteins is contained in exon 3 and there are only five amino acids in exon 1 and 12 amino acids in exon 2. The last (fourth) exon is noncoding. From the restriction map of the exon 3 region of *PRB3* (figs. 1*e*, 1*f*, and 1*g*), the entire exon 3 is contained on two *Rsa*I fragments, a 5′ and a 3′ fragment: the bulk of exon 3 (including the allelic length difference) is present on the 864-bp and 1,054-bp 3′ *Rsa*I fragments (figs. 1*f* and 1*g*). Therefore, we cloned the *Bam*HI fragments containing both *PRB3* alleles in lambda DL10 and then subcloned the 5′ *Rsa*I and 3′ *Rsa*I fragments into Bluescript plasmid. The larger 3′ *Rsa*I fragment was further subdivided into 459-bp and 649-bp *Rsa*I/*Hae*III and 405-bp *Hae*III/*Rsa*I fragments (figs. 1*f* and 1*g*) which were subcloned into Bluescript plasmid (as illustrated for *PRB3*S^{cys} in fig. 1*h*). To completely sequence the larger 649-bp *Rsa*I/*Hae*III fragment from allele *PRB3*M^{null} (fig. 1*f*), it was necessary to prepare a nested set of deletions with exonuclease III and mung-bean nuclease (Stratagene methods). The entire exon 3 portions of both alleles including the 5′ and 3′ splice junctions and adjacent regions of intron DNA were then sequenced at least two times (as illustrated for *PRB3*S^{cys} in fig. 1*i*).

6. *Cloning Procedures*

Human genomic DNA from subject A.F. with the disulfide-bonded variant of Gl (termed Gl 8) was prepared from white blood cells (Poncz et al. 1982). The complete *Bam*HI DNA digest was size fractionated by electrophoresis in 0.8% agarose gel, and the 5–6-kbp fraction was used for cloning into lambda DL10 (Windle 1986). The 980-bp *Hin*fI fragment (fig. 1*d*) from exon 3 of the *PRP* gene, *PRB*1 (Azen et al. 1984), was used to screen the lambda phage library and subclones. The phage library was packaged in vitro using Gigapack Gold packaging extracts (Stratagene) and screened without amplification. All subcloning was done in Bluescript plasmids (Stratagene) for double-stranded sequencing by dideoxy-chain termination (Sanger et al. 1977). Sequence data were analyzed using software provided by the University of Wisconsin Computer Group (Devereux et al. 1984).

Results

1. *A Disulfide-bonded Gl Variant Protein (Gl 8) of Slow Electrophoretic Mobility Is Found in Saliva of A.F. from the Ashkenazi Jewish Family Previously studied (Azen 1977)*

We studied the salivary proteins of subject A.F. by several electrophoretic techniques in order to identify the Pa 2 protein with a product of a specific *PRP* locus. We first studied the acidic PRPs in an isoelectric focusing gel system that is used for typing Pr, Db, Pa, and PIF (Azen and Denniston 1981) and found that subject A.F. is Pr1-1, Db+, Pa−, and PIF+ and has no "Pa-like" acidic protein that is dissociated by treatment of saliva by disulfide reduction. A positive control saliva with the Pa 1 protein (Pa+) showed dissociation of the Pa 1 protein by disulfide reduction as expected (not shown). Since we previously could assign most of the visibly stained salivary proteins to specific *PRP* loci (Azen 1989), the entire range of salivary proteins of A.F. was studied (with and without disulfide reduction) in the SDS gel system. Among the proteins visibly stained with amido black, a slowly migrating Gl protein variant was found that was also dissociated by 2-mercaptoethanol (not shown). This seemed like a good candidate for the Pa 2 protein, since the Pa 1 protein is the only other PRP that was previously known to be disulfide bonded (Azen and Maeda 1988).

We then more specifically analyzed the Gl proteins of A.F. and controls in the acid/lactate gel system stained for carbohydrate with the Schiff reagent. As shown in figure 2*A*, lane 6, saliva of A.F. shows a slow-mobility Gl protein that comigrates with Gl 4 (a slow-mobility Gl protein variant previously described [Azen et al. 1979]). No other distinct Gl band is seen in the A.F. sample, although there is a smear migrating faster than the Gl band. This smear is often seen in other saliva samples, and its significance is uncertain (Azen et al. 1979; Minaguchi et al. 1981). Disulfide reduction shows dissociation of the slow-mobility Gl band (presumed disulfide-bonded dimer) in the A.F. sample and appearance of a faster band (presumed thiol monomer) at or

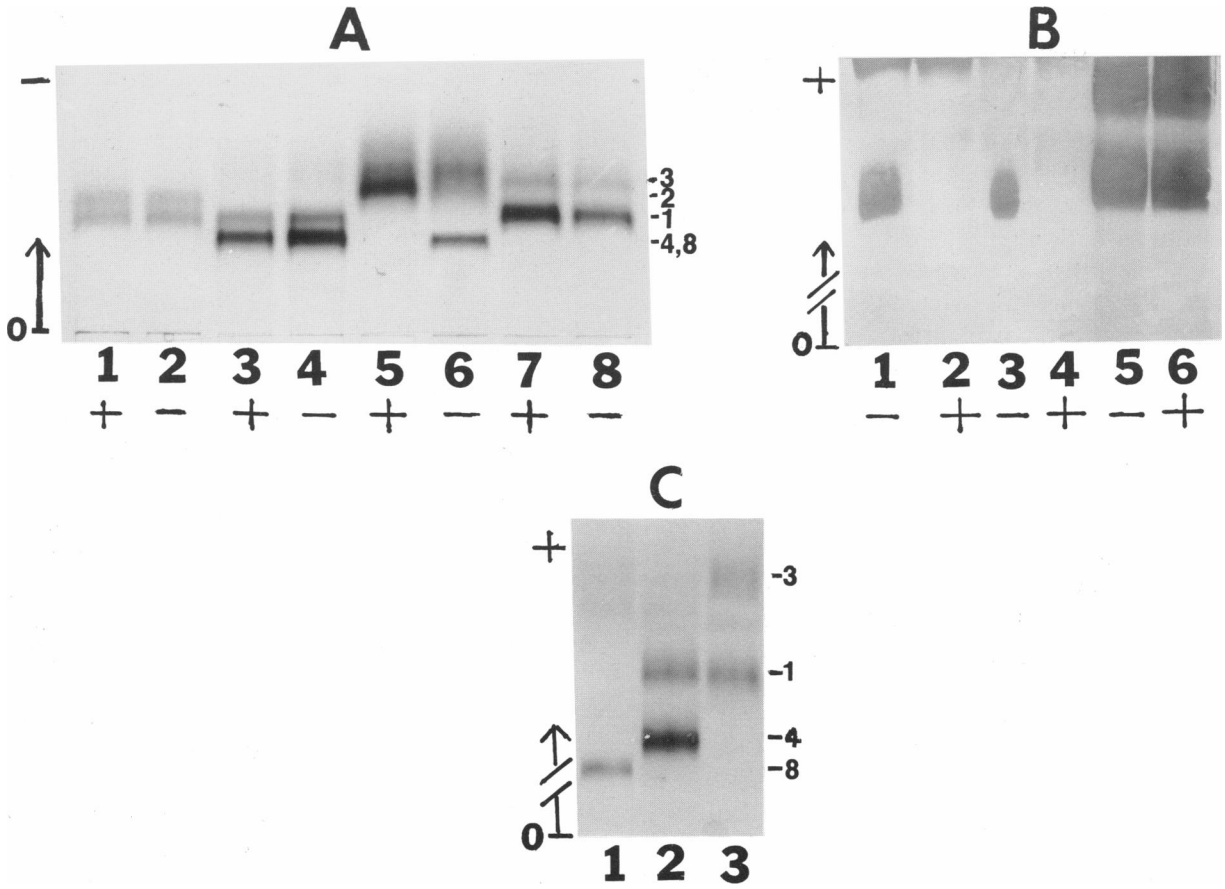

Figure 2 Electrophoretic characterization of Gl 8 in acid/lactate and SDS polyacrylamide gels. *A,* acid/lactate gel stained with the Schiff reagent for carbohydrate. Saliva samples with different Gl type were treated either by disulfide reduction with dithioerythritol followed by iodoacetamide (+) or by iodoacetamide alone (−). Lane 1, Gl 1-2; lane 2, Gl 1-2; lane 3, Gl 1-4; lane 4, Gl 1-4; lane 5 (A.F.), Gl 8; lane 6 (A.F.), Gl 8; lane 7, Gl 1-3; lane 8, Gl 1-3. Gl 8 alone is affected by disulfide reduction and is the only distinct Gl protein present in the A.F. sample (lane 6). *B,* SDS gel with Con A stain of transferred blot for glycoproteins. Saliva samples were (+) or were not (−) treated by disulfide reduction with 2-mercaptoethanol. Lane 1 (W.K.), Gl 8; lane 2 (W.K.), Gl 8; lane 3 (A.F.), Gl 8; lane 4 (A.F.), Gl 8; lane 5, undefined control Gl variant; lane 6, undefined control Gl variant. Both Ashkenazi Jews A.F. and W.K. possess a slow-mobility Gl variant (Gl 8) that is dissociated by disulfide reduction. Faster-mobility Gl variants overlap other unrelated glycoproteins and are not easily distinguished by the standard method of staining. *C,* SDS gel with Con A stain of transferred blot. The stain was modified to more specifically detect Gl proteins (see Methods). The saliva samples were not treated by disulfide reduction. Lane 1 (A.F.), Gl 8; lane 2, Gl 1-4; lane 3, Gl 1-3. Note that only Gl 8 is present in the A.F. sample (lane 1).

close to the position of Gl 2 (fig. 2*A*, lane 5). As will be discussed later, we have named the disulfide-bonded Gl variant Gl 8. Note that, in other samples of figure 2*A*, none of the other Gl proteins, such as Gl 1, Gl 2, Gl 3, or Gl 4, are affected by disulfide reduction.

To screen for Gl 8 in population studies, we modified the SDS gel system to better separate and sensitively stain the slow-mobility Gl protein variants that migrate close to the gel origin. All slow-mobility variants found were then tested with and without disulfide reduction of the saliva samples by 2-mercaptoethanol. Figure 2*B* (Con A stained for glycoproteins) shows dissociation of the Gl 8 protein in sample A.F. (see lanes 3 and 4) and another unrelated Ashkenazi Jewish sample W.K. to be discussed later (see lanes 1 and 2). As a negative control, a different and undefined slow-mobility Gl variant from another Ashkenazi Jewish sample is not affected (cf. lanes 5 and 6).

Under the standard conditions of staining with Con A, the rapid-mobility monomer products of disulfide reduction as well as rapid-mobility native Gl variant proteins are not easily detected in this system because

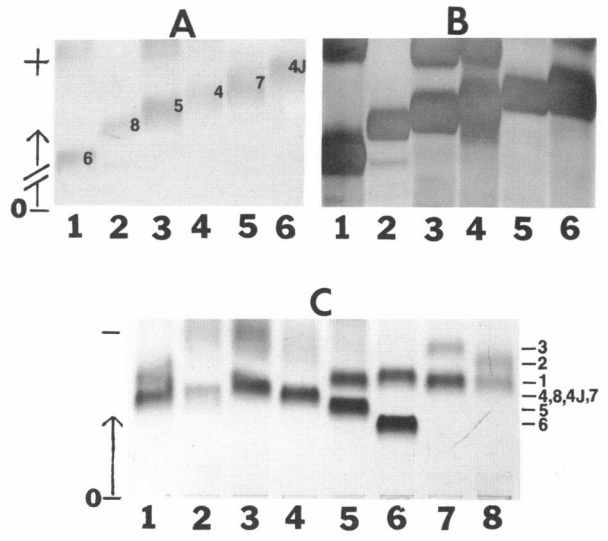

Figure 3 Electrophoretic comparison of slow-mobility salivary Gl variants. *A* and *B*, SDS gel. The blot transferred from the SDS gel was subdivided and the duplicated samples (not treated by disulfide reduction) were stained for protein (*A*) or Con A (*B*). Gl 4 and Gl 8 are from whites, and Gl 4J, Gl 5, Gl 6, and Gl 7 are from Japanese. Lane 1, Gl 6; lane 2 (A.F.), Gl 8; lane 3, Gl 5; lane 4, Gl 4; lane 5, Gl 7; lane 6, Gl 4J. Gl 8 is electrophoretically distinct from the other slow-mobility Gl variants. *C*, Acid/lactate gel stained with the Schiff reagent. The saliva samples were not treated by disulfide reduction. Lane 1, Gl 1-4; lane 2 (A.F.), Gl 8; lane 3 Gl 4J; lane 4, Gl 7; lane 5, Gl 1-5; lane 6, Gl X-6; lane 7, Gl 1-3; lane 8, Gl 1-2. In this system, the slow mobility variants Gl 4, Gl 8, Gl 4J, and Gl 7 cannot be easily distinguished electrophoretically. The faster-mobility Gl protein in lane 6 shows a slightly faster mobility than Gl 1 proteins in adjacent lanes and has been designated X (sample phenotype, Gl X-6).

of overlapping bands. However, by modifying the Con A stain (see Methods) it can be made more specific for Gl proteins in the SDS gel system, and the overlapping bands that obscure typing of faster-mobility Gl proteins as Gl 1, Gl 2, and Gl 3 are much fainter (fig. 2*C*). Figure 2*C* (no disulfide reduction) shows that sample 1 from A.F. (typed in the acid/lactate gel system as Gl 8) shows only a single Gl band, whereas control samples 2 and 3 (typed in the acid/lactate gel system as heterozygous Gl 1-4 and Gl 1-3 respectively) show two bands as expected.

2. *Comparison of Different Slow-ELectrophoretic-Mobility Gl Variants in SDS and Acid/Lactate Polyacrylamide Gels*

Slow-mobility Gl variants from salivas (not disulfide reduced) of whites (Gl 4 and Gl 8) and Japanese (Gl 4J, Gl 5, Gl 6, and Gl 7) show different electrophoretic mobilities when transfers from an SDS gel are stained for protein (fig. 3*A*) or Con A (fig. 3*B*). Gl proteins of more rapid mobility (such as Gl 1, Gl 2, and Gl 3) are not easily distinguished from overlapping and unrelated proteins with standard conditions of staining, and thus are not shown. The Con A stain (fig. 3*B*) is much more sensitive than the protein stain (fig. 3*A*) in detecting the slow-mobility Gl proteins. Among these slow-mobility variants, only Gl 8 (A.F.) is dissociated by disulfide reduction when 2-mercaptoethanol is added to the sample solution (data shown for Gl 8 in fig. 2*B*, lanes 3 and 4, but not for the other slow-mobility variants).

The same slow-mobility Gl variants are shown for comparison in the acid/lactate polyacrylamide gel stained with the Schiff reagent (fig. 3*C*). In this system, Gl 4, Gl 8, Gl 4J, and Gl 7 (fig. 3*C*, lanes 1–4) are not easily distinguished electrophoretically. Note that Gl 8 shows a different relative mobility in the acid/lactate versus the SDS gel. In the SDS gel (fig. 3*A*, *B*, lanes 2) its mobility is between that of Gl 5 and Gl 6 (lanes 3 and 1), but in the acid/lactate gel (fig. 3*C*, lane 2) its mobility is faster than that of either Gl 5 or Gl 6 (lanes 5 and 6). Note that faster-mobility Gl proteins are clearly seen only in lanes 1 and 5–8, which contain different Gl protein heterozygotes. However, distinct faster-mobility Gl proteins are not seen in lanes 2–4, where the samples are typed as Gl 8, Gl 4J, and Gl 7 respectively.

To summarize, Gl 8 is electrophoretically distinct from other slowly migrating Gl variants and is the only one to be dissociated by disulfide reduction. The Gl 8 protein in sample A.F. is not accompanied by an additional allelic Gl protein, as assessed in two different gel systems. This is important, since, as will be shown later from DNA studies, two different molecular-size allelic Gl proteins would be expected if both *PRB3* alleles were expressed in subject A.F. Thus, A.F. possesses the Gl 8 heterozygous phenotype (Gl 8-0) with one expressed and one null allele.

3. *Association of the Disulfide-bonded Gl 8 Protein with Modified Salivary Peroxidase*

The unmodified fast-migrating (F) form of peroxidase is shown in figure 4, lane 1, from saliva of a control who is Pa− and Gl 1 type. In contrast, Gl 8 and Pa 1 proteins are associated with different larger modified forms of salivary peroxidase, presumably by heterodimer disulfide-bond formation. Thus, in saliva of subject A.F., who is Pa− and Gl 8 type, only the very-slow-migrating (V) peroxidase is seen (fig. 4, lane 3). The same modified V peroxidase is also seen in the saliva

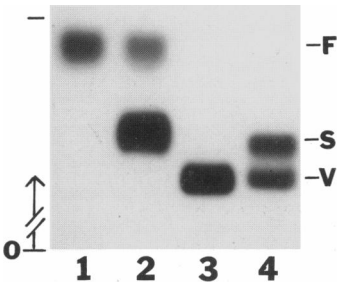

Figure 4 Salivary peroxidase types in acid/lactate gel stained with *p*-phenylenediamine. The saliva samples were not treated by disulfide reduction. Lane 1, Pa–, Gl 1; lane 2, Pa+, Gl 1-2; lane 3 (A.F.), Pa–, Gl 8; lane 4 (W.K.), Pa+, Gl 8. F is unmodified peroxidase and S and V are modified peroxidases associated with the Pa 1 and Gl 8 proteins respectively.

of an unrelated Ashkenazi Jew (W.K.) with the Gl 8 protein (fig. 4, lane 4). This individual is Pa+ and thus also shows the slow-migrating (S) peroxidase characteristic for Pa+. For comparison, lane 2 shows the S peroxidase of a control who is Pa+ and Gl 1-2 type.

4. *Population Survey for the Gl 8 Protein*

We searched for the Gl 8 protein in saliva samples (not disulfide reduced) of several populations of unrelated individuals by electrophoresis of the samples in SDS gels and use of the Con A staining technique. All slow-mobility variants were then tested by disulfide reduction with 2-mercaptoethanol. Since the Gl 8 protein was initially found in an Ashkenazi Jewish family, we studied salivas from American Ashkenazi Jews whose maternal and/or paternal progenitors were of Ashkenazi origin. Two samples among 130 showed a slow-mobility Gl variant that comigrated in the SDS gel with the Gl 8 from A.F. and was dissociated with 2-mercaptoethanol, as is shown for one of these samples (W.K.) in figure 2*B*, lanes 1 and 2. The two samples were further tested in the acid/lactate gel system stained with the Schiff reagent (not shown). The slow-mobility Gl variant in these samples comigrated with Gl 8 from A.F. and was dissociated by disulfide reduction to a monomer that comigrated with that of Gl 8 from A.F. Also, the Gl variant in these samples, like Gl 8 from A.F., was associated with the same modified (V) form of salivary peroxidase, as is shown for one of the samples (W.K.) in figure 4, lane 4. Since the variant Gl protein in the two samples is disulfide bonded, comigrates with the Gl 8 protein from A.F. in two different gel systems, and is associated with the same modified peroxidase (V) as Gl 8 from A.F., we conclude that the samples also possess the Gl 8 protein. The gene frequency of Gl 8 among Ashkenazi Jews is approximately .008. Gl 8 was not found among a general population of 128 whites, 38 blacks, and 104 Japanese. Although the general white population was not ascertained as to Jewish background, the frequency of Jews in this population is believed to be low.

5. *DNA Analysis of Exon 3 Portions of Alleles* PRB3S^{cys} *and* PRB3M^{null} *in Subject A.F.*

The nucleotide and derived amino acid sequence comparisons of exon 3 in *PRB*3S^{cys}, *PRB*3M^{null}, and two *PRB*3 alleles (*PRB*3L and *PRB*3S) previously reported by Lyons et al. (1988*a*, 1988*b*) are shown in figure 5. The sequences are arranged to better compare the tandem repeats. The major finding in allele *PRB*3S^{cys} is a C→T mutation at nucleotide 45 which causes a CGT→TGT (Arg→Cys) change in amino acid 15 (fig. 5 and fig. 6, top). *PRB*3S^{cys} codes for a 229-amino-acid protein (212 amino acids in exon 3 and 17 amino acids in exons 1 and 2).

The major finding in allele *PRB*3M^{null} is an insertion of a single nucleotide (C) at position 621 (fig. 5, fig. 6, bottom) This occurs precisely between the two 63-bp tandem repeats 10 and 11. This nucleotide insertion leads to a frameshift with a premature termination codon that occurs 42 nucleotides downstream from the insertion. Although there is no evidence for expression, *PRB*3M^{null} would code for a 237-amino acid protein (220 amino acids in exon 3 and 17 amino acids in exons 1 and 2). This protein is 19% smaller than the normally coded protein (without the C insertion), which would be 292 amino acids in length. There are no other small nucleotide differences between alleles *PRB*3S^{cys} and *PRB*3M^{null}, although *PRB*3M^{null} is three 63-bp tandem repeats longer than *PRB*3S^{cys}.

When the lengths of the exon 3 portions of the four *PRB*3 alleles shown in figure 5 are compared, they differ from each other by integral numbers of tandem repeats. In order of size from largest to smallest, *PRB*3L (15 tandem repeats) > allele *PRB*3M^{null} (14 tandem repeats) > *PRB*3S and *PRB*3S^{cys} (11 repeats each).

Aside from the length differences, when the nucleotide sequences of exon 3 portions of alleles *PRB*3S^{cys} and *PRB*3M^{null} are compared to those previously determined for alleles *PBR*3L and *PRB*3S, several single nucleotide differences are seen. The CGT→TGT mutation leading to a cysteine substitution is unique to allele *PRB*3S^{cys}, and the nucleotide C insertion with a premature termination codon is unique to allele *PRB*3M^{null}. Eight other single nucleotide differences, which in seven cases lead to amino acid differences be-

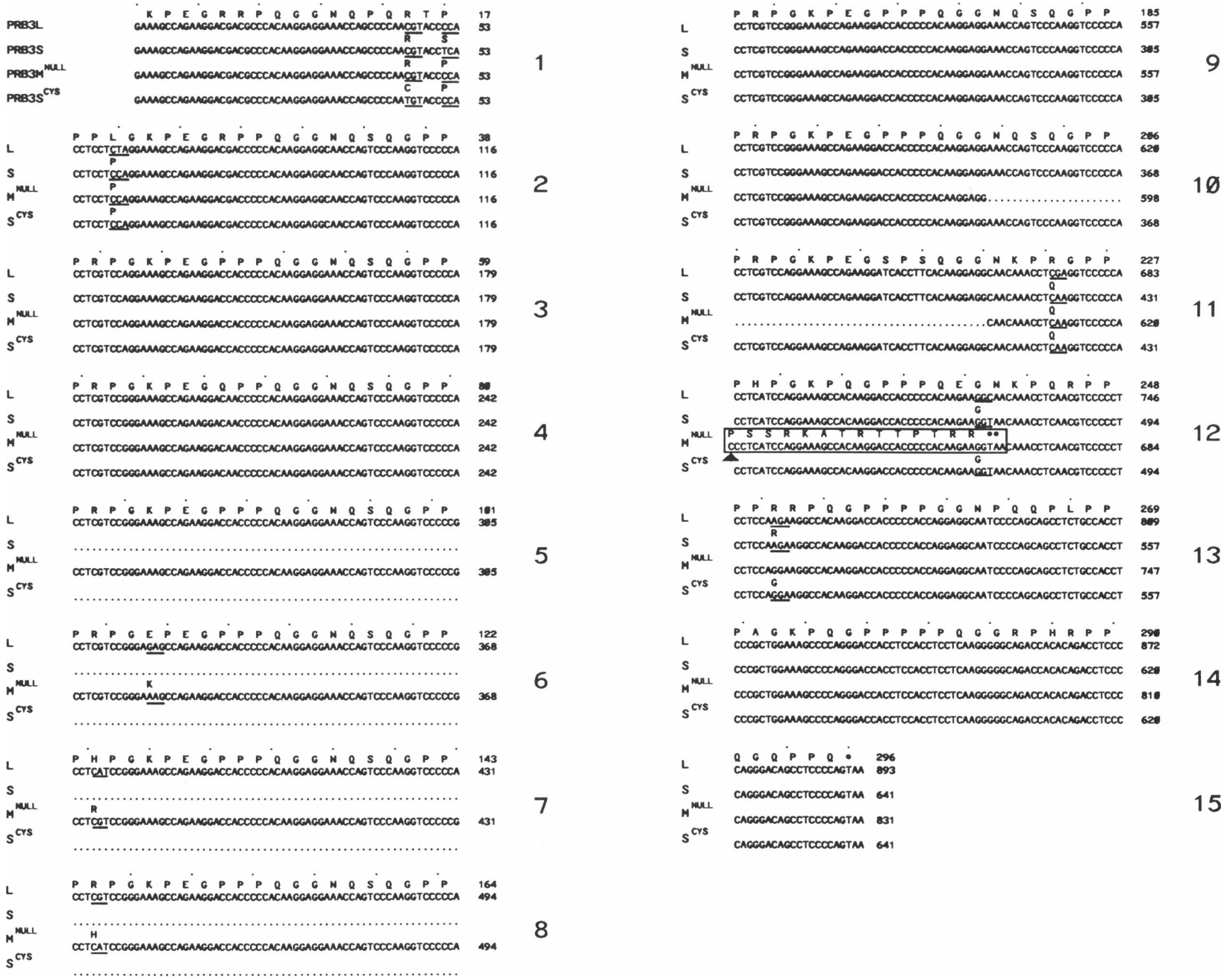

Figure 5 Nucleotide and decoded amino acid sequences of exon 3 coding regions of four *PRB3* alleles. *PRB3*L and *PRB3*S sequences are from Lyons et al. (1988*a*, 1988*b*). *PRB3*M^{null} and *PRB3*S^{cys} sequences are from subject A.F. with the Gl 8-0 phenotype. The sequences are arranged with gaps to permit optimal alignment of tandem repeats that are numbered 1–15 only for the longest allele, *PRB3*L, shown at the top. The nucleotide sequences for all four alleles and the protein sequence for *PRB3*L only are numbered to the right. The normal termination codon is indicated by a single dot and the premature termination codon in *PRB3*M^{null} by two dots. The positions of nucleotide substitutions are indicated by underlining the codons being compared and designating their decoded amino acids. The position of the nucleotide C insertion at position 621 in *PRB3*M^{null} is indicated by an arrowhead, and the region of frameshift downstream from the insertion, including the premature termination codon, is boxed.

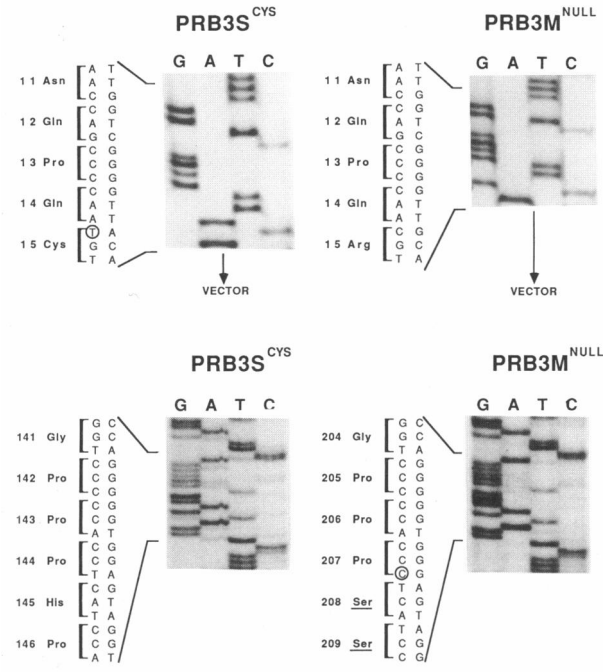

Figure 6 Sequencing gel showing CGT→TGT (Arg→Cys) and nucleotide C insertion mutations in *PRB3* alleles of subject A.F. Antisense sequences are shown in the columns adjacent to the gels. The derived sense sequences and amino acids, which are numbered according to their positions in exon 3 (fig. 5), are shown in the middle and left-hand columns, respectively. The CGT→TGT (Arg→Cys) mutation in the *PRB3*S^{cys} allele is shown above, and the nucleotide C insertion with the frameshift in the *PRB3*M^{null} allele is shown below. The mutated and inserted nucleotides are circled, and the frameshifted amino acids downstream from the C nucleotide insertion in *PRB3*M^{null} are underlined. For both mutations, the normal sequence in the same region of the other allele is shown for comparison.

tween either alleles *PRB3*S^{cys} or *PRB3*M^{null} and *PRB3*L or *PRB3*S, are noted. It cannot be determined whether these differences represent polymorphisms or are due to experimental sequencing error.

Discussion

The Gl protein, the major salivary glycoprotein, is rich in proline, glycine, and glutamic acid, contains about 40% carbohydrate, and has an isoelectric point of greater than 8.2. Levine et al. (1969) state that this glycoprotein accounts for 75% of the total carbohydrate in stimulated parotid saliva. The Gl protein is one of many related salivary PRPs as determined by amino acid compositions (Azen et al. 1979), peptide sequences (Shimomura et al. 1983), and immunological cross-reactivity to other salivary PRPs (Azen and Denniston 1980; Azen and Yu 1984). From family studies, the genetic determinant for the Gl protein was shown to be closely linked to those for other salivary PRPs (Azen et al. 1979). In the context of this paper, it is important to note that among human PRPs only the Pa 1 and Pa 2 proteins were known to be disulfide bonded (reviewed by Azen and Maeda [1988]). It was proposed that the Gl protein, among other heavily glycosylated salivary proteins, may have several functions in the oral cavity including protective formation of intraoral pellicles on hard and soft structures, lubrication at hard and soft tissue interfaces, selective adherence to bacteria, clearance of bacteria, and utilization as a microbial metabolic substrate (Cohen and Levine 1989).

Molecular-size allelic variants of the Gl proteins that were autosomally inherited were described in whites and blacks (Azen et al. 1979), in Japanese (Minaguchi et al. 1981), and in Chinese, Malays, and Indians (Shintani et al. 1990). The parotid salivary protein polymorphism termed Ph (Ikemoto et al. 1979) is the same as two slow-mobility Gl protein variants (Minaguchi and Bennick 1989). In addition to a number of expressed allelic variants of Gl, null alleles were also noted among whites, blacks and Japanese at population frequencies of approximately .046, .110, and .105, respectively. The molecular basis for these null alleles, however, was not clear.

More recently, through DNA/protein correlation studies, the Gl proteins were found to be coded by the *PRB3* gene (Lyons et al. 1988*a*). In particular, the relative order of molecular-size Gl protein variants corresponds with the same order of allelic length variants found at the *PRB3* locus (Lyons et al. 1988*a*). These correlations led to the assignment of Gl proteins to their respective *PRB3* allelic length variants in the following order of size: Gl 4/*PRB3*VL > Gl 1/*PRB3*L > Gl 2/*PRB3*M > Gl 3/*PRB3*S. Lyons et al. (1988*b*) sequenced the *PRB3*S and *PRB3*L alleles and determined that unequal and homologous intragenic recombination between tandemly repetitious exon 3 regions accounts for the allelic DNA length variants and their encoded Gl variant proteins.

As previously discussed, in the light of our recent better understanding of the *PRP* gene family, we reinvestigated saliva of subject A.F. to determine the true identity of the uncommon Pa 2 protein, since we doubted that it, like the common acidic Pa 1 protein, is encoded by the *PRH1* locus. We have found by electrophoretic analysis that subject A.F. possesses a slow-migrating Gl variant protein which (like Pa 1) is disulfide bonded. An acidic "Pa-like" protein was not found in

subject A.F., and we conclude that the Pa 2 protein is in fact an uncommon disulfide-bonded Gl protein variant. By electrophoretic analysis, we determined that this Gl protein variant is not the same as the eight previously reported Gl protein variants including four (Gl 1, Gl 2, Gl 3, and Gl 4) described in whites and blacks (Azen et al. 1979) and four (Gl 4, Gl 5, Gl 6, and Gl 7) described in Japanese (Minaguchi et al. 1981). In our electrophoretic comparisons we found that Gl 4 in Japanese (Minaguchi et al. 1981) is not the same as Gl 4 described by Azen et al. (1979). Therefore, we distinguished the two proteins by identifying them as Gl 4J and Gl 4 respectively. We have named the disulfide-bonded Gl protein in subject A.F. Gl 8 in accordance with the numerical sequence and the nomenclature discussed above. From Southern genomic analysis, subject A.F. shows a length polymorphism at the *PRB3* locus, but only one of the *PRB3* alleles is expressed as Gl 8, and the other allele, which should code for a different-size protein, is a null and is not apparently expressed according to our electrophoretic criteria. Thus, the *PRB3* protein phenotype of subject A.F. is Gl 8–0.

We have found the allele for the Gl 8 protein at a low frequency (around .008) in an Ashkenazi Jewish population but not in the general white, black, or Japanese population. There are a number of disease and benign genetic markers that are more prevalent among Ashkenazi Jews than other populations (Goodman 1979). A more extensive survey for the Gl 8 protein variant among the different Jewish populations would be interesting in order to better establish its frequency and distribution.

As stated before, the Gl protein is coded by the *PRB3* locus (Lyons et al. 1988*a*). We have not done amino acid sequencing of the Gl 8 protein but have determined its primary sequence from the decoded nucleotide sequence of its cognate *PRB3* allele. From DNA sequence comparisons, the smaller *PRB3* allele of A.F. contains a unique CGT→TGT mutation that leads to an Arg→Cys change at residue 15 in exon 3 and thus codes for Gl 8 and accounts for its disulfide-bonded property and probable heterodimer interaction with salivary peroxidase. This mutation was not seen in two previously sequenced *PRB3* alleles, *PRB3*L and *PRB3*S (Lyons et al. 1988*a;* present paper, fig. 5). The same type of mutation occurs in exon 3 of the $PRH1^2$ allele that codes for the cysteine-containing Pa 1 protein (Azen et al. 1987). From nucleotide comparisons, the exon 3 portion of the smaller *PRB3* allele (coding for the Cys mutation) is the same length as that of the *PRB3*S allele (Lyons et al. 1988*a,* 1988*b*), and both possess 11 tandem repeats. In accordance with the previous nomenclature based on allelic length (Lyons et al. 1988*a,* 1988*b*) we have named this allele $PRB3S^{cys}$.

From electrophoretic studies of salivary proteins of subject A.F. in two different gel systems (Fig. 2*A,* 2*C*), we found no evidence for a Gl protein product of the larger *PRB3* allele. It is possible, however, that a protein may have been produced from this allele, but either because of reduced amount or some modification related to the frameshift mutation it was not electrophoretically identified. From DNA sequence comparisons, the larger *PRB3* allele of A.F. has a unique insertion of a C nucleotide which is located precisely between tandem repeats 11 and 12. This mutation was not seen in two previously sequenced *PRB3* alleles, *PRB3*L and *PRB3*S (Lyons et al. 1988*a;* present paper, fig. 5). This causes a frameshift with a premature termination codon 42 nucleotides downstream of the insertion. It is likely that this frameshift mutation could account for the apparent lack of expression of this null allele. There are a number of other reported examples of small mutations, such as insertions and deletions, with premature termination of the protein and lack of expression that may be related to decreased mRNA stability. Some examples include β-globin (Orkin 1984), α_1-antitrypsin (Crystal 1989) and triose phosphate isomerase (Daar and Maquat 1988). Although exon 3 of the null allele is normal in all other respects including 5′ and 3′ splice junctions, we have not studied the entire allele including the other three exons, their splice junctions, and promoter region to exclude any additional mutations. It is interesting that the single nucleotide insertion occurs precisely between two tandem repeats. Perhaps the mutation occurred during the intragenic homologous recombination that is postulated to generate the different length polymorphisms at the *PRP* loci (Azen et al. 1984; Lyons et al. 1988*b*). From nucleotide comparisons, the exon 3 portion of the null allele is intermediate in length (14 tandem repeats) between *PRB3*L (15 tandem repeats) and *PRB3*S (11 tandem repeats). In accordance with the previous nomenclature based on allelic length (Lyons et al. 1988*a,* 1988*b*), we have named this allele $PRB3M^{null}$.

Previously, null alleles were noted to occur at high frequency in several other basic PRP polymorphisms such as Pe and Pm (from *PRB1*), Ps (from *PRB2*), and Po, Con 1, and Con 2 (from *PRB4*) (reviewed by Azen and Maeda [1988]). Lyons et al. (1988*a*) showed from nucleotide analysis that some of the null alleles are actually productive alleles having alterations in proteo-

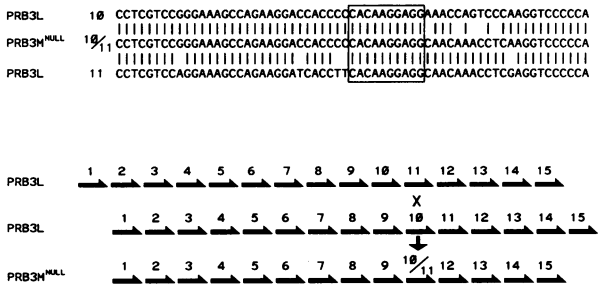

Figure 7 Model for the generation of the 14 tandem repeats in exon 3 of *PRB*3M^{null} by unequal and homologous crossing-over between *PRB*3L alleles. Repeats 10 and 11 (both from *PRB*3L) and composite repeat 10/11 (from *PRB*3M^{null}) are arranged to optimize homologies (top). The box encloses the region of crossover. In the hybrid tandem repeat 10/11 of *PRB*3M^{null}, the nucleotide sequence of the 5′ region resembles repeat 10 and the 3′ region resembles (with the exception of a single nucleotide mismatch) repeat 11 of the *PRB*3L allele. An overall model for the postulated recombination is also shown (*bottom*).

lytic cleavage sites of the precursor protein. In the absence of cleavage, the larger precursor peptides are not resolved electrophoretically with an absence of polymorphic smaller PRPs that are seen when cleavage occurs. Since the Gl protein (from *PRB*3) does not possess postulated proteolytic cleavage sites (Lyons et al. 1988*a*), this mechanism does not apply to the Gl protein. However, the nucleotide insertion with the consequences mentioned previously could explain the apparent lack of expression from the *PRB*3M^{null} allele and may apply to other null alleles among the four *PRB* genes.

As previously stated, Lyons et al. (1988*b*) determined that homologous and unequal intragenic recombination could explain the occurrence of frequent length polymorphisms at the *PRP* loci. We can account for the generation of the 14 repeats of the *PRB*3M^{null} allele according to this model as shown in figure 7 (bottom), where unequal pairing of two *PRB*3L alleles could lead to the formation of the *PRB*3M^{null} allele. In figure 7 (top) the crossover region is enclosed by a box, and in the hybrid tandem repeat 10/11 of *PRB*3M^{null}, the nucleotide sequence of the 5′ region resembles repeat 10 and the 3′ region resembles (with the exception of a single nucleotide mismatch) repeat 11 of the *PRB*3L allele.

It is interesting that products of two different *PRP* loci, Pa 1 from *PRH*1 and Gl 8 from *PRB*3, probably modify salivary peroxidase through heterodimer formation. The slower mobility of the V compared to the S form of modified peroxidase may be explained by the larger size of the complexing Gl 8 thiol monomer (229 amino acids) compared to the Pa 1 thiol monomer (150 amino acids). The possible functional significance of this complex formation on the antibacterial function of salivary peroxidase is unknown. However, it is known that salivary peroxidase contributes in several ways to the maintenance of oral health, primarily as a nonimmunoglobulin defense factor for oral microorganisms (Tenovuo and Pruitt 1984). Perhaps the postulated *PRP* components of the peroxidase complexes might lead to their localization by absorption to oral tissues and microbial surfaces.

Acknowledgments

We greatly appreciate the efforts of S. Josvai and R. Kessel in collecting saliva samples, the cooperation of the Madison B'nai B'rith Hillel Foundation, and the helpful suggestions of K. Lyons and L. Sabatini. We thank B. Windle for providing us with lambda DL10. This investigation was supported by National Institutes of Health grant DE03658–25. This is paper 3106 from the Laboratory of Genetics, University of Wisconsin—Madison.

References

Azen EA (1977) Salivary peroxidase (SAPX): genetic modification and relationship to the proline-rich (Pr) and acidic (Pa) proteins. Biochem Genet 15:9–29

Azen EA (1978) Genetic protein polymorphisms in human saliva: an interpretive review. Biochem Genet 16:79–99

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

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

Azen EA, Denniston C (1981) Genetic polymorphism of PIF (parotid isoelectric focusing variant) proteins with linkage to the PPP (parotid proline-rich protein) gene complex. Biochem Genet 19:475–485

Azen EA, Denniston CL (1974) Genetic polymorphism of human salivary proline-rich proteins: further genetic analysis. Biochem Genet 12:109–120

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

Azen EA, Kim H-S, Goodman P, Flynn S, Maeda N (1987)

Alleles at the *PRH1* locus coding for the human salivary-acidic proline-rich proteins (PRPs) Pa, Db, and PIF. Am J Hum Genet 41:1035–1047

Azen EA, Lyons KM, McGonigal T, Barrett HL, Clements LS, Maeda N, Vanin EF, Carlson DM, Smithies O (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, Yu PL (1984) Genetic polymorphisms of Con 1 and Con 2 salivary proteins detected by immunologic and concanavalin A reactions on nitrocellulose with linkage of Con 1 and Con 2 genes to the SPC (salivary protein gene complex). Biochem Genet 22:1–19

Cohen RE, Levine MJ (1989) Salivary glycoproteins. In: Tenovuo JO (ed) Human saliva: clinical chemistry and microbiology. Vol 1. CRC, Boca Raton, FL, pp 101–130

Crystal RG (1989) The α_1-antitrypsin gene and its deficiency states. Trends Genet 5:411–417

Daar IO, Maquat LE (1988) Premature translation termination mediates triosephosphate isomerase mRNA degradation. Mol Cell Biol 8:802–813

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

Goodman RM (1979) Genetic disorders among the Jewish people. Johns Hopkins University Press, Baltimore

Ikemoto S, Minaguchi K, Tomita T, Suzuki K (1979) A variant protein in human parotid saliva detected by SDS polyacrylamide gel electrophoresis and its inheritance. Ann Hum Genet 43:11–14

Laemmli U (1970) Cleavage of structural proteins during assembly of the head bacteriophage T4. Nature 227:680–685

Levine MJ, Weill JC, Ellison SA (1969) The isolation and analysis of a glycoprotein from parotid saliva. Biochim Biophys Acta 188:165–167

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

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

Minaguchi K, Takaesu Y, Tsutsumi T, Suzuki K (1981) Studies of genetic markers in human saliva. VII. Frequencies of the major parotid salivary glycoprotein (Gl) system in a Japanese population. Bull Tokyo Dent Coll 22:1–6

Orkin SH (1984) The mutation and polymorphism of the human β-globin gene and its surrounding DNA. Annu Rev Genet 18:131–171

Poncz M, Solowiejczk D, Harpel B, Mori Y, Schwartz E, Surrey S (1982) Construction of human gene libraries from small amounts of peripheral blood analysis of β-like globin genes. Hemoglobin 6:27–36

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

Shimomura H, Kanai Y, Sanada K (1983) Amino acid sequences of glycopeptides obtained from basic proline-rich glycoprotein of human parotid saliva. J Biochem 93: 857–863

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

Tenovuo J, Pruitt KM (1984) Relationship of the human salivary peroxidase system to oral health. J Oral Pathol 13: 573–584

Towbin H, Staehlin T, Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and applications. Proc Natl Acad Sci USA 76:4350–4354

Windle BE (1986) Phage lambda and plasmid expression vectors with multiple cloning sites and *lacZ* α-complementation. Gene 45:95–99