*Am. J. Hum. Genet.* 41:1035–1047, 1987

# Alleles at the *PRH*1 Locus Coding for the Human Salivary-Acidic Proline-Rich Proteins Pa, Db, and PIF

EDWIN A. AZEN,*,† HYUNG-SUK KIM,† PATRICIA GOODMAN,†,[1]
SHEILA FLYNN,†,[2] AND NOBUYO MAEDA†

*Department of Medicine and †Laboratory of Genetics, University of Wisconsin–Madison

SUMMARY

We cloned and sequenced the entire exon and intron structures of *Db* and *Pa* genetic determinants at the *PRH*1 locus. Their derived amino acid sequences and that previously determined for the *PIF* protein completely explain the electrophoretic phenotypes of the acidic proline-rich proteins (PRPs) Pa, Db, and PIF. Thus, the Cys substitution near Arg 106 in the Pa protein sterically interferes with proteolytic cutting at Arg 106 and accounts for the single-banded phenotype. In contrast, the Db and PIF proteins are proteolytically cut at Arg 106 and show a double-banded phenotype. The Db protein has an extra 21–amino acid repeat that accounts for its larger size compared with the equal sized Pa monomer and PIF proteins. Several amino acid substitutions account for the charge and mobility differences of the Pa, Db, and PIF proteins in isoelectric-focusing gels. These DNA/protein correlations, as well as the extremely similar genomic-DNA sequences that differ by <1%, establish that *Pa, Db,* and *PIF* are alleles at the *PRH*1 locus. On the basis of the DNA sequences, we conclude that *Db* and *Pa* alleles diverged more recently from a common precursor than did the *PIF* allele from its precursor.

## INTRODUCTION

The human *PRP* gene family consists of six closely linked genes (Azen et al. 1984; Maeda 1985) on chromosome 12p13.2 (Azen et al. 1985; Mamula et al.

Received March 19, 1987; revision received May 18, 1987.

Address for correspondence and reprints: Dr. E. A. Azen, 214 Genetics Building, University of Wisconsin, Madison, WI 53706.

1. Current address: Department of Neurology, University of California, San Francisco, CA 94143.

2. Current address: University of Minnesota Medical School, Minneapolis, MN 55455.

© 1987 by the American Society of Human Genetics. All rights reserved. 0002-9297/87/4106-0007$02.00

1985) coding for many salivary proline-rich proteins (PRPs) that show frequent protein polymorphisms (Azen, in press). The PRPs include many basic, acidic, and glycosylated types (Bennick 1982). On the basis of previous population and family studies of acidic PRP polymorphisms, it was postulated (1) that the Pa, Db, and PIF proteins are coded by three different but closely linked genes (*Pa*, *Db*, and *PIF*, respectively) that are part of a larger *PRP* gene complex and (2) that each gene is expressed as one productive and one nonproductive (null) allele (Azen and Denniston 1981; Goodman et al. 1985). In contrast, it was postulated that other acidic PRPs are encoded by three productive alleles (without nulls) at a different but closely linked locus (*Pr*) (Azen and Denniston 1981; Goodman et al. 1985).

Recent DNA studies (Maeda 1985) showed that there are six *PRP* genes that can be classified into two subfamilies. Four genes termed *PRB1*, *PRB2*, *PRB3*, and *PRB4* are of the *Bst*NI type, which means that the restriction enzyme *Bst*NI cuts the repetitive region (exon 3) of these genes frequently. Two genes, *PRH1* and *PRH2*, are of the *Hae*III type, which means that *Hae*III cuts the repetitive region (exon 3) of these genes frequently. The cDNA work (Maeda et al. 1985) showed that the six PRP genes can code for many more proteins by means of differential RNA processing and posttranslational proteolytic cleavages, thus adequately accounting for the many PRPs in the saliva.

After a reexamination of previous population and family data, Maeda (1985) noted (1) that there are neither any individuals having all three proteins nor any individuals lacking all three and (2) that the gene frequencies of $Pa^+$, $Db^+$, and $PIF^+$ sum to ~1.0 in different racial groups; thus she postulated that the acidic Pa, Db, and PIF proteins may be determined by three productive alleles without nulls at a single locus rather than coded by three separate genes. The *PRH1* locus was postulated to code for PIF, Db, and Pa proteins (Maeda et al. 1985), since (as will be described below) the *Bgl*II/*Xba*I-length polymorphism of the *PRH1* gene correlates with the protein phenotypes. The closely related *PRH2* locus was postulated as coding for the acidic Pr proteins (Maeda et al. 1985), which are very similar to Pa, Db, and PIF proteins and are determined by three productive alleles without nulls. If Pa, Db, and PIF proteins are allelic, they would be expected to show similar electrophoretic phenotypes; however, the Pa protein shows a single-banded phenotype, whereas the Db and PIF proteins show double-banded phenotypes (Azen and Denniston 1981). It was also known that the Pa protein (Friedman and Merritt 1975), alone among acidic PRPs, contains cysteine, although its amino acid sequence had not been determined (Azen 1977, 1978; Goodman et al. 1985). To explain these different phenotypes, Maeda (1985) postulated that the Pa protein may contain a cysteine substitution at the Arg 106 proteolytic cut site, thus preventing cleavage. This cleavage, as will be detailed below, probably accounts for the double-banded phenotypes for the Pr, Db, and PIF proteins (Azen and Denniston 1981).

The acidic PRPs are well separated in isoelectric-focusing polyacrylamide gels, pH 3.5–5.2 (Azen and Denniston 1981). Whereas the Pr, PIF, and Db

acidic proteins have double-banded phenotypes, with each band of a pair staining with approximately equal intensity, the acidic Pa protein appears as a single band and is disulfide bonded, probably as a dimer (Azen 1977, 1978; Azen and Denniston 1981). Wong and Bennick (1980) previously determined the complete amino acid sequence for the Pr1 protein (also termed protein C). Wong et al. (1983) also showed that Pr1 is partially cleaved at Arg 106 by kallikrein, mainly in salivary glands. Thus, one band of the pair represents the intact Pr1 protein, 150 amino acids in length, and the other band represents the larger of the cleaved fragments, 106 amino acids in length. This would explain the double-banded phenotype for Pr proteins, although it is unclear why the presumed proteolysis does not go to completion. Since both the PIF and Pr proteins show identical amino acid sequences surrounding the Arg 106 residue (Maeda et al. 1985), it seems likely that the double-banded phenotype for the PIF protein is due to the same mechanism; however, it was not known whether this is also true for the Db protein, since its amino acid sequence, like that of the Pa protein, was not known.

We therefore wished to (1) test, at a molecular level, Maeda's hypothesis of allelism of the *Pa, PIF,* and *Db* genetic determinants (Maeda 1985) and (2) understand the basis for their different protein phenotypes (Azen and Denniston 1981). Genomic Southern blot analysis revealed a length polymorphism in *PRH*1 that correlates with the Db, Pa, and PIF protein phenotypes. We therefore cloned and sequenced the entire exon and intron structures of *PRH*1 for the *Db* and *Pa* genetic determinants. Their derived amino acid sequences and that previously determined for *PIF* (Maeda et al. 1985; Kim and Maeda 1986) completely explain the electrophoretic phenotypes of the proteins. These DNA/protein correlations, as well as the very similar genomic-DNA sequences of *Pa, PIF,* and *Db,* establish that their products are coded by alleles at the *PRH*1 locus.

## MATERIAL AND METHODS

### *Phenotypes of Db, Pa, and PIF Proteins*

Parotid saliva samples were electrophoresed in isoelectric-focusing polyacrylamide gels, pH 3.5–5.2, and phenotyped for the Db, Pa, and PIF proteins according to a method described elsewhere (Azen and Denniston 1981). According to Maeda's (1985) hypothesis, the Db, Pa, and PIF proteins are products of three different alleles (without nulls) at the *PRH*1 locus. Thus (according to Maeda's hypothesis), genotypes for *PRH*1 proteins of three individuals whose DNAs were used in cloning experiments include a *Db/Db* homozygote (also $Pa^-$, $PIF^-$), a *Pa/Pa* homozygote (also $Db^-$, $PIF^-$) and a *Pa/Db* heterozygote (also $PIF^-$). In the description of the double-banded electrophoretic phenotypes (Azen and Denniston 1981), the slow Db band and slow PIF band refer to the uncleaved proteins (representing the primary *PRH*1 gene products), whereas the fast Db band and fast PIF band represent the larger of the cleavage products of the slow proteins.

*Cloning Procedures*

Human genomic DNAs were prepared from the white blood cells or parotid glands of individuals whose salivas were also typed for acidic PRPs in isoelectric-focusing polyacrylamide gels. *Hin*dIII or *Xba*I libraries were constructed from complete genomic-DNA digests. The digests were ligated into either lambda-phage Charon 35 (Loenen and Blattner 1983) or Charon 3AΔ Lac (Blattner et al. 1977). Two probes were used for phage cloning and plasmid subcloning: *Hae*III 500, a 500-bp probe containing the *Hae*III-type repeat from a cDNA clone coding for *PRH*1, and *Hae*III 1.5, a 1.5-kb *Eco*RI/*Bgl*II fragment from the 5′ region (including exon 1) of the *PRH*2 locus (Kim and Maeda 1986). The phage libraries were first screened with the $^{32}$P-labeled *Hae*III 500 probe after hybridization at 68 C in 0.9 M sodium chloride/0.09 M sodium citrate/0.2% bovine serum albumin/0.2% Ficoll/0.2% polyvinylpyrrolidone/0.5% sodium dodecyl sulfate (SDS). The strongly hybridizing *Db* and *Pa* candidates were then each subcloned and isolated as two *Xba*I fragments (see restriction map of *PRH*1 at top of fig. 1). One of these fragments, herein termed the 5′ *Xba*I fragment, is ~1.9 kb and contains exon 1. This fragment hybridizes to the *Hae*III 1.5 probe. The other fragment, herein termed the 3′ *Xba*I fragment, is ~2.2 kb in length and includes the rest of the gene. It hybridizes to the *Hae*III 500 probe. DNA sequencing was performed according to the method of Maxam and Gilbert (1977) with slight modifications (Slightom et al. 1980), and all regions were sequenced in both directions. The sequences were analyzed using software provided by the University of Wisconsin Genetics Computer Group (Devereux et al. 1984).

*Strategy for Cloning*

As a guide for cloning and sequencing, we used the previously published DNA sequence and restriction map (fig. 1, top) of a prototype *PRH*1 gene (Kim and Maeda 1986). This DNA sequence codes for the PIF protein, is ~3.5 kb in length, and contains four exons and three introns. The identifying numbers for the nucleotide differences shown in figure 1 refer to this published DNA sequence. Exon 1 is 64 bp in length and codes for the signal peptide and the first five amino terminal residues. Exon 2, located ~1 kb downstream from exon 1, contains only 36 bp and codes for the next 12 residues. Exon 3, located ~360 bp downstream from exon 2, encodes the repeated region where *Hae*III-type repeats occur tandemly five times. The termination codon occurs in exon 3. Exon 4, located after a long intervening sequence of 1,200 bp, contains the 3′-untranslated region and the poly-A addition signal sequence.

To isolate the *Db* allele, we screened two lambda-phage libraries. We prepared one genomic library from a presumed *Db*/*Db* homozygote by ligating a complete *Xba*I digest of DNA into lamba-phage Charon 3AΔ Lac. We screened this library with the *Hae*III 500 probe and isolated the 3′ *Xba*I fragment of the *Db* allele; however, when we screened this library with the *Hae*III 1.5 probe, we could not isolate the 5′ *Xba*I fragment. We therefore screened a *Hin*dIII genomic library (provided by E. Saitoh and prepared from a *Pa*/*Db* heterozy-

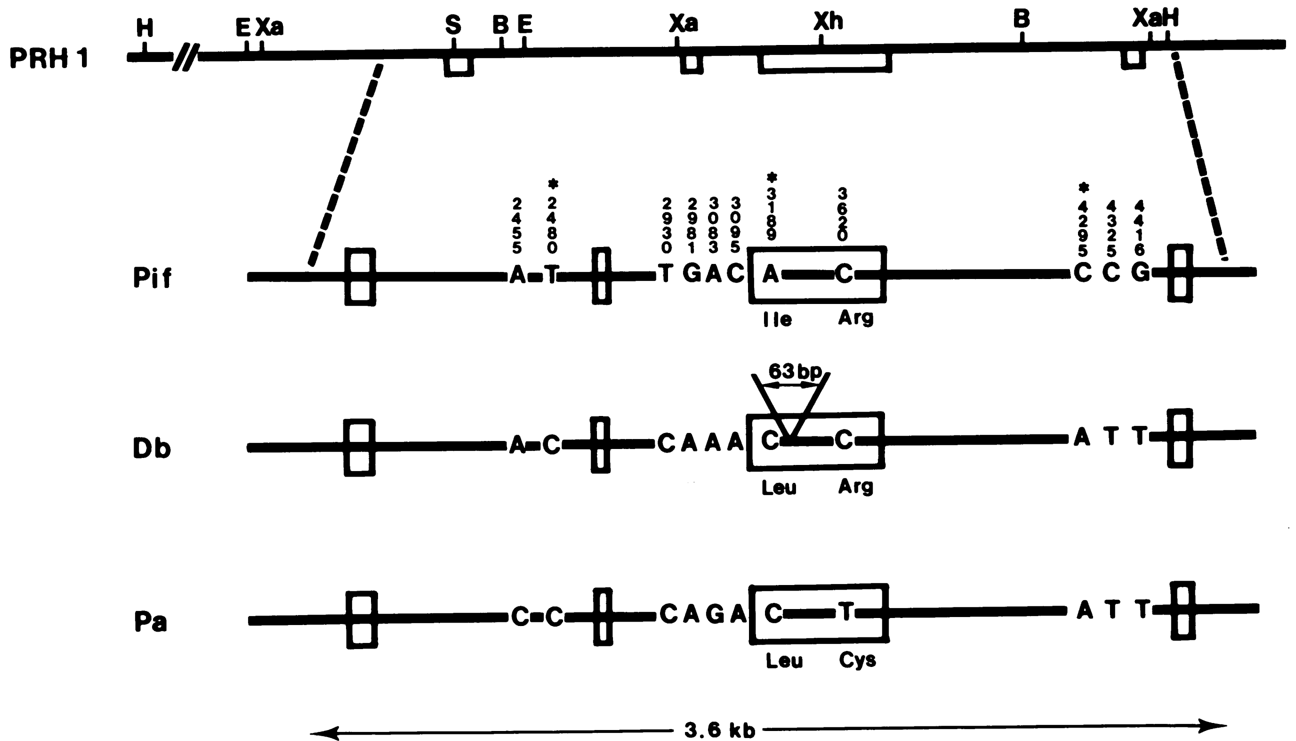


Fig. 1.—Comparison of DNA sequences for the *PIF*, *Db*, and *Pa* alleles at the *PRH1* locus. The restriction map and DNA sequence of the prototype *PRH1* gene (top) that codes for the PIF protein have been published previously (Kim and Maeda 1986). Only the nucleotide differences between *PIF*, *Db*, and *Pa* are shown, and these are numbered according to the previously published (Kim and Maeda 1986) sequence for *PIF*. Three of these nucleotide differences (starred at positions 2480, 3189, and 4295) cause polymorphic restriction-enzyme-site changes including *Dde*I, *Hph*I, and *Dde*I restriction-fragment-length polymorphisms, respectively. The boxes represent exons. The 63-bp insert in exon 3 of *Db* is also shown. H = *Hin*dIII; E = *Eco*RI; Xa = *Xba*I; S = *Sfi*I; B = *Bgl*II; and Xh = *Xho*I.

gote) with the *Hae*III 500 probe. The entire *Db* allele was recovered on a 10-kb *Hin*dIII fragment (see restriction map at top of fig. 1). The putative *Db* allele could be distinguished from the *Pa* allele by comparing an *Xba*I/*Xho*I digest of the 10-kb *Hin*dIII fragment with an *Xba*I/*Xho*I digest of a previously isolated fragment of the *Pa* allele. Thus, the putative 0.7-kb *Xba*I/*Xho*I Db fragment, which includes a portion of exon 3, is 63 bp larger than the corresponding Pa fragment, as will be discussed below. We then subcloned and isolated both the 5′ and 3′ *Xba*I fragments of the *Db* allele from the 10-kb *Hin*dIII fragment. We confirmed the identity of the 1.9-kb 5′ *Xba*I Db fragment by hybridizing it on a Southern blot to the *Hae*III 1.5 probe. We confirmed the identity of the 2.3-kb 3′ *Xba*I Db fragment by digesting it with *Xba*I/*Xho*I and showing the diagnostic 0.7-kb fragment as described above.

To isolate the *Pa* allele, we prepared a genomic library from a presumed *Pa*/*Pa* homozygote by ligating a complete *Hin*dIII digest of DNA into lambda-phage Charon 35 and then screened this library with the *Hae*III 500 probe. We confirmed the identity of the *Pa* candidate by digesting the DNA with *Bgl*II, *Eco*RI, and *Hin*dIII, hybridizing it with the *Hae*III 500 probe on a Southern blot, and noting the characteristic *PRH*1-type hybridizing bands. We then subcloned and isolated both the 1.9-kb 5′ *Xba*I fragment and the 2.2-kb 3′ *Xba*I fragment of the *Pa* allele and confirmed their identities by specific hybridization, on Southern blots, to either the *Hae*III 1.5 or *Hae*III 500 probes, respectively.

### *Hybridization to Genomic DNAs*

Total genomic DNAs were completely digested with *Bgl*II/*Xba*I, electrophoresed in 2.0% agarose gels, transferred to nitrocellulose according to the method described by Southern (1975), and hybridized to the $^{32}$P-labeled *Hae*III 500 probe by using conditions described elsewhere (Vanin et al. 1983). Prior to autoradiography, filters were washed in 0.45 M sodium chloride/0.045 M sodium citrate, pH 7/0.5% SDS for 2 h.

## RESULTS

### 1. *Size Comparisons of Acidic PRPs in SDS Gels*

The Db slow (larger Db band) and Pa dimer proteins were separately eluted from alkaline polyacrylamide gels; however, the Pr1 and PIF slow proteins electrophoretically overlap in the alkaline polyacrylamide gel and could not be separated prior to elution. When the mobilities of these eluted proteins are compared under reducing conditions in an SDS gel (fig. 2*A*), the slow PIF, Pr1, and Pa monomer proteins are of the same size (Pr1 and PIF in lane 2 and Pa in lane 4), whereas the slow Db protein is larger (lane 3). The Pr1 and the slow PIF proteins also comigrate in the SDS gel, since only a single band is seen (lane 2).

### 2. *Hybridizing Genomic Fragments from the PRH1 Gene in Individuals of Known Acidic PRP Phenotypes*

To look for DNA/protein-size correlations, we studied genomic hybridization patterns for the *PRH*1 gene in 13 individuals whose salivas were also typed

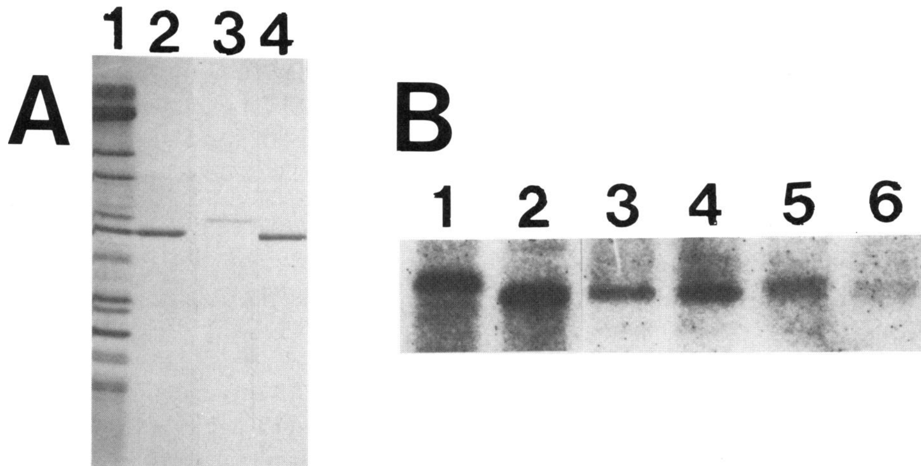


FIG. 2.—*A,* SDS gel electrophoresis of purified acidic PRPs. Lane 1, Parotid saliva; lane 2, Pr1 (protein C) and PIF proteins comigrate; lane 3, Db protein; lane 4, Pa protein. *B,* Hybridizing genomic fragments from the *PRH*1 gene in individuals of known acidic PRP phenotypes. The genomic DNAs were completely digested with *Bgl*II/*Xba*I, and Southern blots were hybridized to the $^{32}$P-labeled *Hae*III 500 probe. Lane 1, *Db/Db* homozygote; lane 2, *Pa/Pa* homozygote; lanes 3 and 4, *PIF/PIF* homozygotes; lanes 5 and 6, *Db/PIF* heterozygotes.

for acidic PRPs (fig. 2*B*). The genomic DNAs were completely digested with *Bgl*II/*Xba*I, and the Southern blots of these DNAs were hybridized to the $^{32}$P-labeled *Hae*III 500 probe. Among the 13 DNA samples, one or two very closely spaced bands that are characteristic of the *PRH*1 gene are seen, and the band patterns can be classified into three types. Eight of the 13 DNA samples show a single hybridizing, smaller 1.6-kb band. The salivas of these individuals are Db-, and they are also either Pa+/PIF+ (five of eight), Pa−/PIF+ (two of eight), or Pa+/PIF− (one of eight). Thus, the presumed *PRH*1 genotypes of these three classes, according to Maeda's (1985) hypothesis, are either *Pa/PIF* heterozygote, *PIF/PIF* homozygote, or *Pa/Pa* homozygote, respectively. Three DNA samples of this type are shown in figure 2*B,* lanes 2–4. Five of the 13 DNA samples show a slightly larger band that differs from the smaller band by ~100 bp. The salivas of these five individuals are Db+. One of the five Db+ individuals (who is also Pa−/PIF−) has only the larger DNA band (fig. 2*B,* lane 1) and is a presumed *Db/Db* homozygote. The other four Db+ individuals show both DNA bands, which appear as a closely spaced doublet, with both bands being of approximately equal intensity. Two DNA samples of this type are shown in figure 2*B,* lanes 5 and 6. Three of these four Db+ individuals are also Pa−/PIF+, and one is also Pa+/PIF−. Thus, these four individuals with the doublet patterns are presumed Db heterozygotes (either *Db/PIF* or *Db/Pa*).

On the basis of these results, two of the three putative allelic proteins determined by the *PRH*1 locus (slow PIF and Pa monomer) are the same size and the third (slow Db) is slightly larger. Additionally, in 13 individuals this order of size for the salivary proteins correlates well with the order of size for the *Bgl*II/*Xba*I fragments of the *PRH*1 gene. Since the *Bgl*II/*Xba*I fragment that hy-

bridizes to the *Hae*III 500 probe includes exons 2 and 3 (see restriction map at top of fig. 1), the larger size of the Db protein may be due to a change in length of one of these exons. These DNA/protein correlations strongly support both the previous assignment of *PIF* to the *PRH*1 locus (Maeda et al. 1985) and our proposed strategy of cloning the *PRH*1 gene from individuals of selected acidic PRP types in order to characterize the *Db* and *Pa* genetic determinants.

### 3. *Comparison of DNA Sequences of the Three Putative PRH1 Alleles*

We did a preliminary mapping experiment to localize the region of length difference in the *Db* allele. Therefore, we compared the restriction-enzyme digests of 3′ *Xba*I fragments (containing exon 3) of the *Db* and *PIF* alleles. The combined results with *Bst*NI, *Acc*I, *Hinf*I, *Hph*I, and *Xba*I/*Xho*I localize the added length of the *Db* allele to exon 3 in the region of the tandem repeats.

We then sequenced ~3.6 kb of DNA for both *Db* and *Pa* alleles, including the entire coding regions and introns, and compared the sequences with that previously published for the *PIF* allele as shown in figure 1. A 63-bp insert occurs in the Db allele; otherwise the sequences are identical for the three alleles, with the exception of 11 single-nucleotide differences. Three of these nucleotide differences (starred at positions 2480, 3189, and 4295) cause polymorphic restriction-enzyme-site changes including *Dde*I, *Hph*I and *Dde*I restriction-fragment-length polymorphisms, respectively. Nine of the 11 single-nucleotide differences are in the introns, two in the first intron (positions 2455 and 2480), four in the second intron (positions 2930, 2981, 3083, and 3095), and three in the third intron (positions 4295, 4325, and 4416). Two of the 11 single-nucleotide differences are localized to exon 3 (positions 3189 and 3620). Pairwise comparisons of the DNA sequences show only three single-nucleotide differences between *Db* and *Pa* (plus a 63-bp length difference), whereas there are 11 between *Pa* and *PIF* and eight between *Db* and *PIF* (plus a 63-bp length difference).

### 4. *Comparison of Amino Acid Sequences of PIF, Db, and Pa Proteins*

In figure 3 the derived amino acid sequences of the Db and Pa proteins (starting at the amino terminal end) are compared with the previously published (Maeda et al. 1985) amino acid sequence for the PIF protein. The PIF and Pa monomer proteins are the same length (150 amino acids), whereas the Db protein is 21 amino acids longer owing to the extra 63-bp repeat in exon 3 (fig. 1). This result explains the relative sizes of the three proteins in SDS gels (fig. 2*A*). There are two nucleotide changes in exon 3 (fig. 1). The first (a C→A change at position 3189) causes a Leu→Ile (CTA→ATA) change in the PIF protein at residue 26. The second (a C→T change at position 3620) causes an Arg→Cys (CGT→TGT) change at residue 103 in the Pa protein. (For ease in comparing Pa, Db, and PIF proteins, this amino acid change is shown at position 124 in fig. 3, since the extra 21 amino acids in the Db sequence are also numbered.)

## DISCUSSION

Our molecular data show that the *PRH*1 genetic determinants that we have cloned and sequenced (designated as *Pa* and *Db* here and as *PIF* elsewhere

```
                      10                  20                  30
PIF   Q D L N E D V S Q E D V P L V I S D G G D S E Q F I D E E R
Db                                                      L
Pa                                                      L

                      40                  50                  60
PIF   Q G P P L G G Q Q S Q P S A G D G N Q D D G P Q Q G P P Q Q
Db
Pa

                      70                  80                  90
PIF   G G Q Q Q Q G P P P P Q G K P Q G P P Q Q - - - - - - - - -
Db                                              G G Q Q Q Q G P P
Pa                                              - - - - - - - - -

                     100                 110                 120
PIF   - - - - - - - - - - - - - G G H P P P P Q G R P Q G P P Q Q G
Db    P P Q G K P Q G P P Q Q
Pa    - - - - - - - - - - - - -

                    ↓ 130                 140                 150
PIF   G H P R P P R G R P Q G P P Q Q G G H Q Q G P P P P P P G K
Db          R
Pa          C

                     160                 170
PIF   P Q G P P P Q G G R P Q G P P Q G Q S P Q
Db
Pa
```

FIG. 3.—Comparison of PIF, Db, and Pa protein sequences. Derived amino acid sequences of the Db and Pa proteins (starting at the amino terminal end) are compared with the previously published (Maeda et al. 1985) amino acid sequence for the PIF protein. The arrow represents the position where Db and PIF proteins are cut by proteolysis to generate the double-banded electrophoretic phenotypes. The Arg→Cys substitution in the Pa protein is only three residues removed from the cut site and probably sterically interferes with cutting, thus causing the single-banded phenotype. The extra 21-amino acid repeat in the Db protein (residues 82→102) accounts for this protein's extra length.

[Maeda et al. 1985; Kim and Maeda 1986]) code for the Pa, Db, and PIF proteins, since these sequences completely explain the electrophoretic protein phenotypes, as will be discussed below. These DNA/protein correlations, as well as the extremely similar genomic-DNA sequences that differ by <1%, establish that *Pa, Db,* and *PIF* are alleles at the *PRH*1 locus. In contrast, the two closely related *Hae*III-type genes, *PRH*1 and *PRH*2, differ overall by 8.7% when each base mismatch and each gap is counted as a single difference (Kim and Maeda 1986).

On the basis of analysis of the amino acid sequences, we could completely explain the different electrophoretic phenotypes of the Pa, Db, and PIF proteins. Thus, the Pa monomer protein and the PIF protein are both 150 amino acids long, whereas the Db protein is 171 amino acids long. The added length of the Db protein is due to an additional 21 amino acids determined by an extra tandem repeat that is inserted in exon 3 of the *Db* allele. These length differences, inferred from the amino acid sequences, correlate well with the relative

sizes of the three proteins in SDS gels. We found an Arg→Cys substitution in the Pa protein at residue 103, very close to the Arg 106 proteolytic cut site (Wong et al. 1983). Therefore, it seems likely that this substitution either sterically interferes with proteolysis at residue 106 or disrupts the recognition specificity required for enzymatic cleavage and thus accounts for the single-banded pattern of the Pa protein. In contrast, the Db, PIF, and Pr proteins have homologous amino acids surrounding the Arg 106 cut site, and this accounts for their double-banded patterns. Our finding of a Cys residue in the Pa protein supports previous chemical and electrophoretic evidence that the Pa protein contains Cys and is disulfide bonded (Azen 1977, 1978; Goodman et al. 1985) as a dimer under nonreducing conditions in several electrophoretic-gel systems (Azen 1977, 1978; Azen and Denniston 1981). We found amino acid substitutions that could cause charge and subsequent electrophoretic-mobility differences among these proteins. Thus, the Db protein (with a Lys at residue 95 in the extra repeat) has one more positively charged amino acid than the PIF protein. Since Cys is substituted for Arg at residue 103 in the Pa protein, it, in turn, has one less positively charged amino acid than the PIF protein. The Leu→Ile substitution at residue 26 in the PIF protein does not alter its overall charge. These amino acid substitutions adequately explain electrophoretic differences in isoelectric-focusing polyacrylamide gels, since it has been shown elsewhere (Azen and Denniston 1981) that the slow Db protein is the most basic and that the slow PIF and Pa proteins are less basic in the order given (Db > PIF > Pa).

In exon 3, the *Pa* and *PIF* alleles have five and the *Db* allele has six tandem *Hae*III-type repeats that are obviously related but vary slightly in sequence and length within each allele. To interpret the overall length difference between alleles, we have arbitrarily arranged the first three of the six repeats of the *Db* allele to maximize identity (fig. 4*A*). The extra repeat (labeled 2-1) can be interpreted as a composite of the second and first repeats. Thus, the 5′ 21-nucleotide portion of the extra repeat is identical to that of the second repeat but different from that of the first repeat. In contrast, the 3′ 28-nucleotide portion of the extra repeat is identical to that of the first repeat but different from that of the second repeat. Between these two regions is a 14-nucleotide region (boxed in fig. 4*A*) that is identical in the three repeats. Thus a change occurred, leading to an extra 63-bp repeat in the *Db* allele.

We believe it likely that this extra repeat was generated either by slippage during DNA replication (fig. 4*B*) or by homologous recombination between sister chromatids or chromosomes at the *PRH*1 locus (fig. 4*C*). Homologous recombination has also been responsible for frequent insertion/deletion-type rearrangements in *Bst*NI-type genes (Azen et al. 1984; Lyons et al. 1985). These rearrangements are more frequent in *Bst*I-type than in *Hae*III-type genes, possibly owing to a larger number (as many as 13) of tandem repeats in exon 3 of the former compared with the number in the latter (five or six repeats) (Maeda et al. 1985). We have not identified insertion/deletion-type rearrangements in the *Hae*III-type *PRH*2 gene that codes for acidic Pr proteins. The Pr1 and Pr2 allelic proteins (products of the *PRH*2 locus) are of the same length but

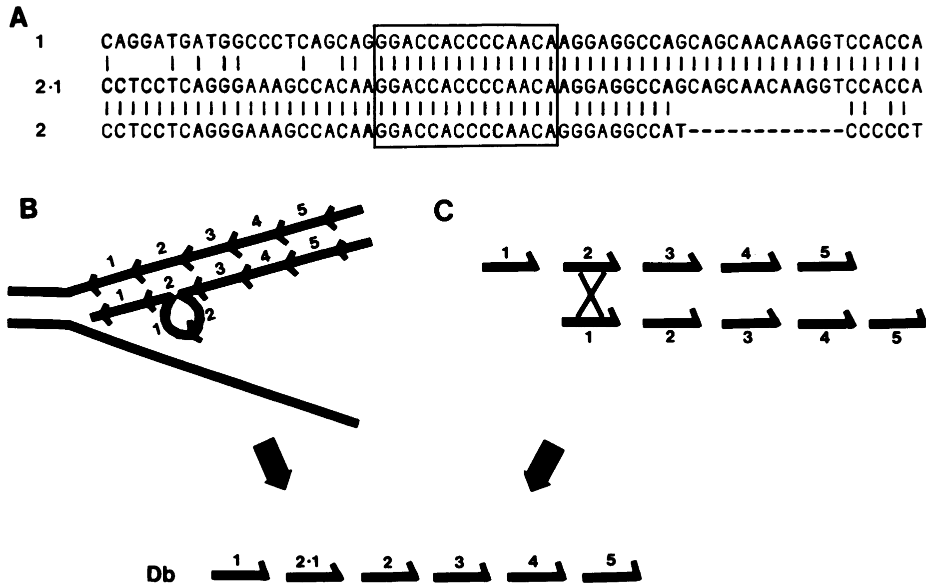


FIG. 4.—Interpretation and possible explanations for generation of the extra repeat in the *Db* allele. *A*, The first three of the six repeats of the *Db* allele are arranged to optimize homologies. The extra repeat (labeled 2-1) can be interpreted as a composite of the second and first repeats. The boxed portion is identical for the three repeats. *B*, Generation of an extra repeat by means of slippage during DNA replication. *C*, Generation of an extra repeat by means of homologous recombination between sister chromatids or chromosomes at the *PRH*1 locus.

have a presumed charge difference. In recent unpublished studies, we gel-purified another allelic protein produced by *PRH*2, the Pr1′ protein found primarily in blacks (Azen and Denniston 1974), and compared its size with that of the Pr1 protein by means of SDS gel electrophoresis. We found the Pr1′ protein to be the same size as the Pr1 protein; therefore, the electrophoretic difference between the Pr1′ and Pr1 proteins in nondenaturing polyacrylamide-gel systems (Azen and Denniston 1974, 1981; Azen 1978) is more likely to be due to a charge difference than to a size difference.

Only a limited statement can be made regarding the possible evolutionary interrelationships of the *PRH*1 alleles. When DNA sequences are compared and the number of single-nucleotide differences are considered (see fig. 1), *Db* and *Pa* alleles seem more closely related to each other than is *Db* to *PIF* or *Pa* to *PIF*. Thus, *Db* and *Pa* alleles may have diverged more recently from a common precursor than did the *PIF* allele from its precursor.

Several possible biologic functions reviewed by Bennick (1982) were elsewhere suggested for acidic PRPs. The acidic PRPs bind calcium at two phosphoserine residues, located in the PRPs' N-terminal portions, that mediate their adsorption onto apatitic surfaces; they also inhibit precipitation of salivary calcium and stabilize the supersaturated state with respect to the calcium phosphate salts that form dental enamel. These functions may be important in maintaining the protective and reparative environment for dental enamel and in

promoting dental health. Since the overall structures of the *PRH*1 and *PRH*2 acidic proteins are extremely similar (Maeda et al. 1985), it is likely that they share the functional properties described above. The Cys substitution in the Pa protein may have added functional significance, since it has elsewhere been shown (Azen 1977) that individuals possessing the Pa protein also have a modified form of salivary lactoperoxidase owing to probable disulfide bonding between the Pa protein monomer and lactoperoxidase. This observation may be clinically relevant, since recent epidemiologic evidence links the presence of the Pa protein in saliva to an increased occurrence of dental caries in children (Yu et al. 1986). This association may be biologically meaningful, since lactoperoxidase is part of an important antibacterial system in saliva that may relate to caries development (Hoogendoorn 1985).

## ACKNOWLEDGMENTS

We thank G. Larson for valuable technical assistance. This work was supported by National Institutes of Health grants DE03658-22 and GM20069. This paper is 2936 from the Laboratory of Genetics of the University of Wisconsin–Madison. We appreciate the comments of K. Lyons and L. Sabatini on the manuscript.

## REFERENCES

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

———. 1978. Genetic protein polymorphisms in human saliva: an interpretive review. Biochem. Genet. **16**:79–99.

———. Genetic protein polymorphisms of human saliva. *In* J. Tenovuo, ed. Clinical chemistry of human saliva. CRC, Press Inc. (in press).

Azen, E. A., and C. L. Denniston. 1974. Genetic polymorphism of human salivary proline-rich proteins: further genetic analysis. Biochem. Genet. **12**:109–120.

———. 1981. Genetic polymorphism of PIF (parotid isoelectric focusing variant) proteins with linkage to the PRP (parotid proline-rich protein) gene complex. Biochem. Genet. **19**:475–485.

Azen, E. A., P. A. Goodman, and P. A. Lalley. 1985. Human salivary proline-rich protein genes on chromosome 12. Am. J. Hum. Genet. **37**:418–421.

Azen, E. A., K. M. Lyons, T. McGonigal, N. L. Barrett, L. S. Clements, N. Maeda, E. F. Vanin, D. M. Carlson, and O. Smithies. 1984. Clones from the human gene complex coding for salivary proline-rich proteins. Proc. Natl. Acad. Sci. USA **81**:5561–5565.

Bennick, A. 1982. Salivary proline-rich proteins. Mol. Cell. Biochem. **45**:83–99.

Blattner, F. R., B. G. Williams, A. E. Blechl, K. Denniston-Thompson, H. E. Faber, L. A. Furlong, D. J. Grunwald, D. O. Kiefer, D. D. Moore, J. W. Schumm, E. L. Sheldon, and O. Smithies. 1977. Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. Science **196**:161–169.

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

Friedman, R. D., and A. D. Merritt. 1975. Partial purification and characterization of a polymorphic protein (Pa) in human parotid saliva. Am. J. Hum. Genet. **27**:304–314.

Goodman, P. A., P. L. Yu, E. A. Azen, and R. C. Karn. 1985. The human salivary protein complex (SPC): a large block of related genes. Am. J. Hum. Genet. **37**:785–797.

Hoogendoorn, H. 1985. Activation of salivary peroxidase antimicrobial system: clinical studies. Pp. 217–227 *in* K. M. Pruitt and J. O. Tenovuo, eds. The lactoperoxidase system: chemistry and biological significance. Marcel Dekker, New York.

Kim, H.-S., and N. Maeda. 1986. Structures of two *Hae*III-type genes in the human salivary proline-rich protein multigene family. J. Biol. Chem. **15**:6712–6718.

Loenen, W. A. M., and F. R. Blattner. 1983. Lambda Charon vectors (CH32, 33, 34 and 35) adapted for DNA cloning in recombination-deficient hosts. Gene **26**:171–179.

Lyons, K. M., H. S. Kim, N. Maeda, and O. Smithies. 1985. Insertion/deletion polymorphisms in human salivary proline-rich protein genes. Genetics **110** (Suppl., pt. 2): s31.

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., H. S. Kim, E. A. Azen, and O. Smithies. 1985. Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system. J. Biol. Chem. **260**:11123–11130.

Mamula, P. A., N. A. Heerema, C. G. Palmer, K. M. Lyons, and R. C. Karn. 1985. Localization of the human salivary protein complex (SPC) to chromosome band 12p13.2. Cytogenet. Cell Genet. **39**:279–284.

Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.

Slightom, J. L., A. E. Blechl, and O. Smithies. 1980. Human fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between duplicated genes. Cell **21**:627–638.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.

Vanin, E. F., P. S. Henthorn, D. Kioussis, F. Grosveld, and O. Smithies. 1983. Unexpected relationships between four large deletions in the human beta-globin gene cluster. Cell **35**:701–709.

Wong, R. S. C., and A. Bennick. 1980. The primary structure of a salivary calcium-binding proline-rich phosphoprotein (protein C), a possible precursor of a related salivary protein A. J. Biol. Chem. **255**:5943–5948.

Wong, R. S. C., G. Madapallimattam, and A. Bennick. 1983. The role of glandular kallikrein in the formation of a salivary proline-rich protein by cleavage of a single bond in salivary protein C. Biochem J. **211**:35–44.

Yu, P.-L., D. Bixler, P. A. Goodman, E. A. Azen, and R. C. Karn. 1986. Human parotid proline-rich proteins: correlation of genetic polymorphisms to dental caries. Genet. Epidemiol. **3**:147–152.