J. Biochem. 121, 1025-1030 (1997)

Nucleotide Sequence of Gene PBI Encoding a Protein Homologous to Salivary Proline-Rich Protein P-B[1]

Satoko Isemura*,[2] and Eiichi Saitoh†

*Nippon Dental University Junior College at Niigata, Niigata 951; and †Department of Oral Biochemistry, School of Dentistry at Niigata, Nippon Dental University, Niigata, Niigata 951

Received for publication, November 20, 1996

The nucleotide sequence of gene PBI, the putative translational product of which is homologous to salivary proline-rich protein P-B, has been determined. PBI is 6.4 kb long and contains 3 exons. PBI appears to code for the precursor of a proline-rich protein which we have designated P-B1. The P-B1 precursor is composed of 134 amino acid residues including 22 residues of signal sequence, 23 residues of N-terminal sequence, five repeating units of 13 to 14 residues and 22 residues of C-terminal sequence. The signal sequence of this precursor is identical with that of the P-B precursor. The N-terminal 61 residue sequence of P-B1 has homology of 75% with the whole sequence of P-B (57 residues), when deletion of 4 residues is taken into account. P-B1 is 55 residues longer than P-B. The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL, and GenBank nucleotide sequence databases with the following accession number D89501.

Key words: D89501, genomic DNA, PBI, proline-rich protein, salivary protein.

Various kinds of proline-rich proteins (PRPs) including acidic, basic or glycosylated forms, have been isolated from human saliva and their chemical, physicochemical, and genetic features have been extensively studied (*1*). They are the major protein constituents of human saliva. Although they may play roles in the maintenance of teeth by inhibiting crystal growth of hydroxyapatite on the surface of teeth (*2*) and in the detoxification of dietary tannin by inhibiting absorption from the digestive organs (*3*), their physiological significance is not yet well understood.

The nucleotide sequences of cDNAs (*4*) and genes (*5*) of PRPs have revealed that polymorphism of PRPs is due to the presence of multiple genes belonging to the same family, alternative splicing of mRNA precursors, and post-translational modification.

Proline-rich protein P-B, isolated from saliva and characterized by us in 1979 (*6*), is similar to other PRPs in the high content of proline and the presence of repeating units, but the structure of the repeating unit GPGXX′PPPP in P-B, where X and X′ are any amino acid, is different from the unit PPGKPQGPPP in other PRPs. The nucleotide sequence of P-B cDNA (*7*) showed that P-B is not a degradation product of a larger protein, but a mature protein in itself. P-B cDNA has no significant homology with other PRP cDNAs or genes, suggesting that it may belong to a distinct PRP gene family.

Bovine PRP, the amino-terminal 57 residues of which are identical with those of P-B, was isolated from bovine developing enamel by Strawich and Glimcher (*8*). The apparent molecular weight of this protein is 31,000 as determined by SDS polyacrylamide gel electrophoresis. Since the molecular weight of P-B is 5,793, there could be other P-B like proteins in humans.

Although antibodies against other saliva proteins such as proline-rich protein C (*9*) and cystatins (*10*) have been reported, we have not been able to prepare antibodies against P-B in mice, rats, or rabbits. The unavailability of antibodies against P-B has hampered studies on this protein and other antigenically related proteins, if present, in human saliva, as well as studies on species specificity of P-B.

This study was initiated to isolate the P-B gene and other genes belonging to the same family to obtain information on their expression, intron-exon structures and family relationships.

MATERIALS AND METHODS

The following materials were purchased from the sources indicated: λEMBL3 human genomic library, human submaxillary gland cDNAs, Toyobo; *Escherichia coli* JM109 competent cells, size markers for electrophoresis, and PCR reagents, Takara Shuzo; dye terminator sequencing kit, Perkin-Elmer; ECL direct nucleic acid labeling system, ECL 3′-oligonucleotide labeling system, and ECL detection system, Amersham; enzymes and reagents used for molecular cloning, Toyobo and Takara Shuzo; custom-made oligonucleotides used as probes and primers, Asahi Emerse; DNA extraction kit, Wako Pure Chemicals.

λCharon 35 Human Genome Library—A λCharon 35 library was constructed from a complete *Hin*dIII digest of normal human genomic DNA by the method of Horn (*11*).

[1] This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan.

[2] To whom correspondence should be addressed. +81-25-267-1500, +81-25-267-1510

Abbreviations: C, carboxy; N, amino; PRP, proline-rich protein; USF, upstream stimulatory factor; nt, nucleotide.

Probe Preparation—A *Bam*HI/*Pst*I fragment (418 bp) of P-B cDNA prepared according to the method described previously (7) and a *Hin*dIII/*Eco*RI fragment (598 bp) of B41 (see below) were labeled according to the method described in the protocol of the ECL direct labeling system. Oligonucleotide probes were labeled according to the method described in the protocol of the ECL 3′-oligonucleotide labeling system.

Hybridization and Detection—Hybridization and detection were performed according to the method described in the protocol of the ECL direct nucleic acid or ECL 3′-oligonucleotide labeling system using hybridization reagents contained in the kit.

DNA Sequencing—The nucleotide sequences were determined by a DNA sequencer (Perkin Elmer 373S) using the dye terminator dideoxy method following the protocol of the manufacturer. The determined sequence was confirmed by re-analysis using custom-made primers designed to be hybridizable with the determined sequence. For confirmation of the sequence between the second and third *Sac*I sites, phage DNA of clone b21 was used as a template instead of the plasmid prepared from the PCR product, to exclude errors due to misamplification by PCR (see below).

DNA Preparation from Human Blood—Human genomic DNA was prepared from human leucocytes according to the protocol of the DNA extraction kit. As for other experimental techniques, standard protocols for molecular cloning were employed (*12*).

Isolation of Genomic Clone B41—B41 was cloned by screening 2.3×10^9 plaques of λ Charon 35 human genomic library using the P-B cDNA fragment as a probe as described above. Hybridization was performed by shaking in hybridization buffer containing 0.5 M NaCl at 42°C overnight. Membranes were washed with low stringency buffer [0.5×SSC (1×SSC is a solution containing 15 mM sodium citrate and 150 mM NaCl), 0.4% SDS and 6 M urea] at 42°C. Positive clone B41 was selected for further analysis.

Isolation of Genomic Clone b21—b21 was cloned by screening 4×10^4 plaques of λEMBL3 human genomic library using the *Hin*dIII/*Eco*RI B41 fragment as a probe. Conditions of hybridization and membrane washing were the same as in the case of B41 cloning.

Preparation of Plasmids for Sequencing—DNA of B41 was digested by *Hin*dIII and the resultant fragment of 20 kbp which was hybridizable with the P-BcDNA *Bam*HI/*Pst*I fragment was subcloned into PUC18. DNA of b21 was digested by *Sac*I and the fragments of 4.6 and 2.6 kbp which were hybridizable with an oligonucleotide corresponding to 13-30 nt and a *Bam*HI/*Pst*I fragment of P-BcDNA (7), respectively, were subcloned into PUC18. A 900 bp *Sac*I fragment located between the two *Sac*I fragments of 4.6 and 2.6 kbp was obtained by *Sac*I digestion of the PCR product amplified using b21 as a template and subcloned into PUC18. The primers used in PCR are oligonucleotides corresponding to 3637-3657 nt and to the complementary chain of 5218-5237 nt. For the numbering see Fig. 2.

Homology Search—Computerized homology search was performed by using FASTA search service.

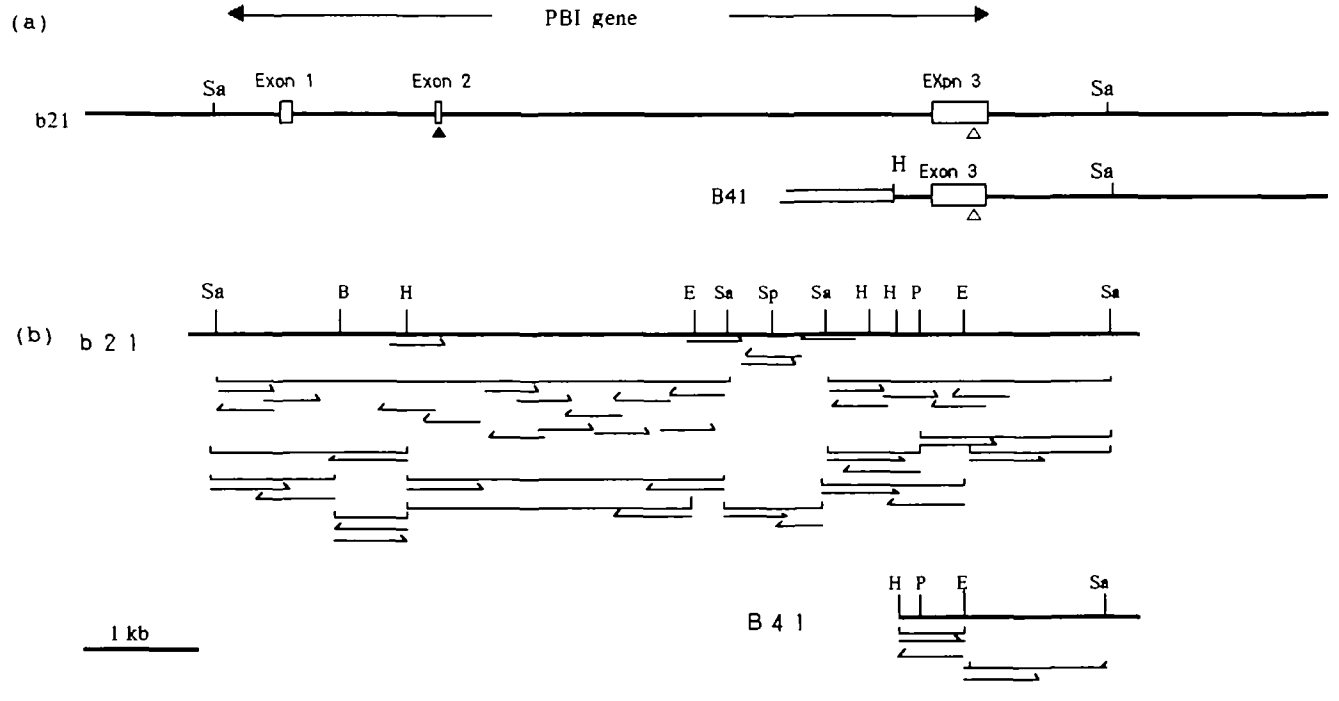

Fig. 1. **Sequencing strategy and restriction maps of PBI in clones B41 and b21.** (a) Gene organization of PBI. Exons are depicted by open boxes. Solid and open triangles are the positions of an initiation codon and a termination codon, respectively. The region between the two *Sac*I sites of b21 indicated here and the region between the *Hin*dIII and *Sac*I sites of B41 were sequenced. Sa, *Sac*I; H, *Hin*dIII; bold line, insert DNA; ⎓, phage DNA. (b) Restriction maps of PBI and sequencing strategy. Restriction enzymes: Sa, *Sac*I; B, *Bam*HI; H, *Hin*dIII; E, *Eco*RI; SP, *Sph*I; P, *Pst*I. The insert DNAs were cut by restriction enzymes and subcloned into PUC18 and then sequenced using M13 forward and reverse primers. Internal parts which could not be determined by M13 primers were sequenced using custom-made primers. The arrows indicate the length and direction of sequencing.

```
gagctcgagg atcatccagc ataggaaaaa ttatcctcag tgcctagatt tatcaaaaag acaaggctat aaatttccat tagacatcca cttcaaagta   -536
aaaatatctg tataattttc ccataatcaa atctttgcat gtcatcatat tttcatttaa ttaaaactct caagttaata acaacaggga ccacatgatt   -436
tttaaaacac ggttgttctc agttcttcac tgacaaaagg gaacttccta tagaaaatac aaaagatggc ccattaaata aaatagaaaa tacaaaagat   -336
ggcccattaa aattactcgg gatagaaaaa gttggtgctt tttctttaaa ataaagagga aatggtgcgc cacaggcccc agattgtttc cctgaagtga   -236
gaggatatta tttgttgaat ctgtttcccc aggaaacaac aaatattgaa caggtcaatg cttttcagaa atgagaccca aaaaatgcgg aaaaagactg   -136
ccaaattggc aagaacaact ttcctatctt tgtttgaaga gcatcagccc agaagttatt ttatttccaa gtagcctttc ttttactgcc aacagaaatc    -36
                                        1
ttttgcaata tataagaact gaagtttctt gtctcTTTTT CACCTTTATT TTCTGTCTTT CAACTGGCAA GAGTCATTTT GACCAGCAGG TTAATCAACT     65
CTAAGACAGA TCCTCACGCA AAGgtatgta cctggaaata ttaactgaag aaatttttttc tcagactttt catgagcttt ttataggaca ttcattatca    165
ccaaaaacat aagcaccttc accttatgta aaattgtaag tcgttaaaca tatcaaatct tgaaaaaatta tgttattcac atgtcagagt aaagcattca    265
tgttgggcat tgaggacaat tagtccaccc aagtacacaa ttgacttttc cccttgatgt tccttaaagt tcttgaaact tcataatgat agattattct    365
cagatgtact ctaaaagagg gaattgttcc acagtatatg tttattatta gtttttcttt tgtttttcaa acttcttata agaacggatc ctgggaaaag    465
taatccttaa gggaatagcc acacccacat acaattgaat atccagtgaa attttgtcca tataattctt tcaaaacttg cgttgttcag acaacacact    565
caacctatgt gcagagttga ctgtcaggaa gccatttgta caatggtgca tggagttgct ttcccagaac tcaagctatc aggtacttga aatatttcac    665
aaattgtgag ccaactgtga ataaaatatc tggactcctc catattctct gacttggtct ttgatctacc atatacggat ttttggaagc actttgagac    765
cacacatcac agactaaatt aatttagtgg aatgttgtta aaactaggag atatataata tgactctttt tttcttacac tacatcaaac agagtgattt    865
ctttcttttc cttatttttg agatatccca aggaattaca ctaacatatg atctagttaa tatgttaata tttcaaaccc aagcctatag aaatctataa    965
aacaatgtac ctgctattta aaaatgcaaa agtattcaaa tgtaaataat tgtacatgtc aaagtagtta cctctaaagg caaaaactgt agttacaaaa   1065
gcttctattg ttcaatcatt tatgaacaac tctttggcat tatctccagt atcagtttac aagtcaaaaa taaaaactac ttccactatt ttgaagtcaa   1165
attttataag cctactaatg ttgtcccaca gtaatactga atccaaaata atgttaggca aatttcctaa aaaggctaca gtagtatatc tgtactacat   1265
tatatccatg tgttgagaca ttttaaatag tacttttttaa taaaaaacaa tcatactgat cacctattgt gcttactttc agAGGCAACT GAAAGGATGA   1365
                                                                                                      M  K
AATCACTGAC TTGGATCTTG GGCCTTTGGG CTCTTGCAGC GTGTTTCACA gtaagtatca ttaatcacga tcacacatct ttatactttc tcattaacca   1465
  S  L  T   W  I  L   G  L  W  A   L  A  A   C  F  T
ttacttcaga tttcttttta cattaatgat gttacctttt cttatatatt agtaactatt aaccattgat gaaacactgt tctagtggtt aaatttaaat   1565
ttaaaatcta atttaattta aatagcccca tatgtctgat ggctacaata ctggacagct caagtctgta atatcagact gcattaaaca gataagtata   1665
aatgtaggtc cccatactct agcaattaca caataggtaa aactctataa aatgagatag tttaaagtag ttgttcctaa catatggagt cacctaactc   1765
acttattaaa aatatagtat taaagtcctc atccaagcct gaataatagg acctgggaag atgcattttg cagaggattt aggagaatct ttgcatgcta   1865
aattttgagg aaataataaa catttatagt aacacctacc atatattaaa aactaggtta aggacaacta cccttattgt ccttaaccta gtgtttcaga   1965
attttttaat gcatattttt aaaagtaata gaatgatact caatttcaac agaagggtca attaattcca gtgttacaaa gaggagatta attccagagt   2065
aacttctccc aaagtaccat aaaatctatt atatcaaatg ggaagatatg aaagctgatt ttttagaaga ctatgacttc cacaaagcag gcaacataaa   2165
attatatcat tctgaaacat gaaacataaa ataatacctc tttcttctac ttttaaaaac ttgaataaaa acaggacatt ctgagatgac tgaagtattc   2265
tgactttggc caccttcctg catgcattat gggtggacat aattttgat aagccaataa aacatattga catccgttaa atgtcatgtc ttaatgactt   2365
catcataaag ctgattttca acattttcac atcctcaaaa tacaagagaa aaacatttta ttcaggcttt tattatacaa atagtaactt tttacaagat   2465
aaattagaaa aacagagctt tccaataata taagcagttt aaatagaaat gagccttttt ccagttgata tgataatgat gattaataat tactgaacat   2565
tggaaatgtg aagggtagca tgtcctttgc atatacatat ttgtagatgg ggtaagtcat catattcagc ctgttacctt ttgagtgttt tctctgttat   2665
aggtaatttt ccaaatgcat tacatacatt atctcttctt taatcctcaa agcaagtctc actatttttа atctcactaa aatacaaata aatttagatt   2765
caaattatgc atgtcttagc tggattattg ctaaattgaa taatgtagga gattaacatt aattatagcc ctttgagagg ggagtgtttc ttcctaatat   2865
agtattttca tcattatcat cattttactt tagccagttg tggaaaggga gcttatcact gcgattgggt taagagaaca ggaaaagaat agatatttac   2965
tatgaagatt tttcattaca taatatttag atcaacatcg gtatatcttt aggaaaaggt actttcagct aatttttttaa ttgtctcaac aataaaaatg   3065
tgtgaaatat ttttccatac taaggatttg ttttcttgac aataaaaaat caaatcctga acttccacat aacctgacta gtccttggat tggcagccct   3165
tgggtcaaac accctgcata acaacttgaa ccatcactag ctagagaaga caaatgcaag gctggtccca cagaagtgcc ttggaatttt ttgtttgttt   3265
cattggtgat atgaataaaa cttttaccgg catgttaagg gcaaaagcca ggttgaagag caaatgggaa gtgagaagtg gggttggcaa atttagaata   3365
ctctttcata gagctttgct ttgcaggaaa gaggtgaaaa ggaggtcaca acaacaaggg tattggtggt tgaaactgaa tcttttataa acagggaaat   3465
atcaatttat atttatttct caaatgatgt cttcgacctc tctgtacgtg ttaggtgaaa gcggcaagat tagaatcaag aggtcaaatc aacgggccct   3565
acagtaattt gggtaaaagt aaatgacaga ttagaatgaa agcaagaaag atgataagaa gtggttagat ttagaattca gggagtggga tgaagagctg   3665
ataggatagg gagaaaaaaa attgagttga acataaatga tagtgttttt ctgaagaaaa tgagctaatt aaatggcaga tattacttgc agaaagtaaa   3765
agagaagact gacagagcca ggtccctcag aaaacaatta taaagaaaaa gaaactgtct agctatcaat gactccatgg ttgtgtggcc agtaaatcag   3865
ggaagttcgt ttctaattgc ctgaaatatc ttggtgaagc aggagaaaag ccagtttggt gggggtagat gagagctcag gatgaagtgg agaatgttta   3965
taaaattggc atcaacttgg aaaattacag gcctctgggg aagtgggaag agataatcta aggtaagata ctcaaaaaag gatttcaggc agcactgaag   4065
gacgctaact ttctgatggt acagtgcttg tgacatcctc tagcattgtt aagaacctgg aactaggaat ccaggagaaa gatggtttac cgtaatacta   4165
aagtttagga ttagcaaggg taatggccag acaaataggt gaagggactg agaaccacaa gaagtgaatg tagttttgtg ccctatataa accatctgct   4265
ccaagcaacc tggagttctt gccgtcagta gagtggaact tcattcttac cccagagcat ttgcatattt tccttctatt tggcatgctc cctccttatt   4365
ttgttatctt taagcctctt ttcaataact ccttatcaca gagacaatcc tttatcattt tcctatccag aactgatctc tttaaatccc taatgtattt   4465
gttccttgca ctacttttca gctatttttа gatttttaat ggagatgggg tcttattatg ttgtccaggt tggtcttgaa ctgctggcct aaagctatct   4565
acctacctca gcctcccaaa gtgctgggat tacaggcatg agccaccata tccagcctca gctattttca tttctaaaaa ttgtactcta tgtaaaccag   4665
acctgagata agcaccaccc taggcacttt acaggtgata tttcacaata gatgtgacaa caaccgcagg agatgggtac ttttatgaac atcattttcc   4765
tgataagaaa aagagactca aaaaggattc tatactcacc ctatctctat gtgccagaag gagagctcaa acactggcat gaaatcccag acactcattc   4865
cattcatttc acttcccttt agatcacatt tatccccccc tgtattgtaa ttactcagcc tgtggattca tagaagacat gtttgtaccc ctggcaaact   4965
gccaggttga cattcatgac atgaatgaac cagctccttc attagctcat tgcacaaaat agttgctcac atgtgtttca caattaatta atgttcaaag   5065
cccagaacaa ccctgaactc ttgaattaga cataaaaaga cgtgtctttc ccatgctgca aaggcattta tttccatttg gaatttggaa tgacagactt   5165
cactacaggc caaggataag atttttgctc attatctatt ttttccagtc aatactatcc aagctttcat tgagagccta ggtattcaga gggcaaaaga   5265
gaaacatttt tctttgatac agaagttcat acatcataac ttagatacat ttaacaattt tctattgctc aataacaaat taccactaat tagtagcctt   5365
taaaaactca ttattatctc agtttctttg gccataactc tgagcttggc ttaggtggct tctctgctca gagtctcaca ggctgaaaag gtgtgtccat   5465
tgctgtattt ttttttttctg cacctcagtg tcctttttta agcttgcatg gttgttggca ggattcagtc ccttggttgt tggactgaag cccctatttt   5565
tttgctgcct atgagccagg ggccactctc agctcctata gccagtctac catgtacttc cacatggccc tcgccacaac acggcatctt gattcaaagc   5665
ccagaggaaa gcatttgctg cagcttcaaa tctctctgat tccttctcac attgtagaac ctgttttaaa agggctcacc tgataaggtc aggccagcat   5765
gtgccagcag ggagtagaaa tcttgaggcc atctcagagt tctgcttacc atagaaacca ttcacccttа tatttaactg tgaaatatct gatttaaaat   5865
tatttacttc tttgtttcca cagCCTGGTG AGAGTCAAAG AGGCCCCAGG GGACCATATC CACCTGGACC ACTGGCTCCT CCTCCTCCAC CACGTTTTCC   5965
                      P  G  E   S  Q  R   G  P  R   G  P  Y  P   P  G  P   L  A  P   P  P  P  P   R  F  P
TTTTGGAACA GGATTTGTTC CACCACCCCA TCCTCCACCC TATGGTCCAG GGAGATTTCC ACCACCCCTT TCTCCACCCT ATGGTCCAGG GAGAATCCCA   6065
 F  G  T   G  F  V  P   P  P  H   P  P  P   Y  G  P  G   F  F  P   P  P  L   S  P  P  Y   G  P  G   R  I  P
```

Fig. 2 (continued on next page)

```
CCATCCCCTC CTCCACCCTA TGGTCCAGGG AGAATTCAAT CACACTCTCT TCCTCCTCCT TATGGCCCAG GTTATCCACA GCCACCTTCC CAACCAAGAC   6165
P  S  P  P   P  P  Y   G  P  G   R  I  Q  S   H  S  L   P  P  P   Y  G  P  G   Y  P  Q   P  P  S   Q  P  R  P
CCTATCCACC TGGACCTCCA TTTTTCCCTG TAAATTCTCC AACTGATCCT GCCCTCCCTA CTCTTGCACC CTAAATACAG ACAACTGCAA CAGGTGCCAC   6265
  Y  P  P   G  P  P   F  F  P  V   N  S  P   T  D  P   A  L  P  T   L  A  P   *
CACCCACAAA AGACAACACT ACCCTCGTAA CTACTGCTTC TACTACCCAA AAATAAGAAT TTCAACACTA CTTCCAAGAG ACTTTTAGAT AAAATCACTT   6365
CCATTTTTGG ATGAGAATAA AGATTTCCAA Agcactgagc ttttgggaga aatatcttag aaattgtgaa acgatcccca tgaaccttta tatcagtagg   6465
ggaaaataaa gaattgagca acaatatgaa gtatccactg ttatcagagc caatagttta caccccagtt atgtcaccta aatgaatatt agtgctaaca   6565
g
```

Fig. 2. **Nucleotide sequence of PBI.** The numbers on the right refer to the nucleotide residues. The putative transcriptional site is numbered 1. The exon sequences are given in capital letters while the intron sequences are in lower case letters. Encoded amino acid residues are given as one-letter abbreviations. The TATA box, CCAAT box, and the binding site for USF are underlined. Polyadenylation signals are double-underlined. Sequences required for branching for splicing are shown in bold letters. Primers for PCR to check expression of PBI in salivary gland are marked by dots.

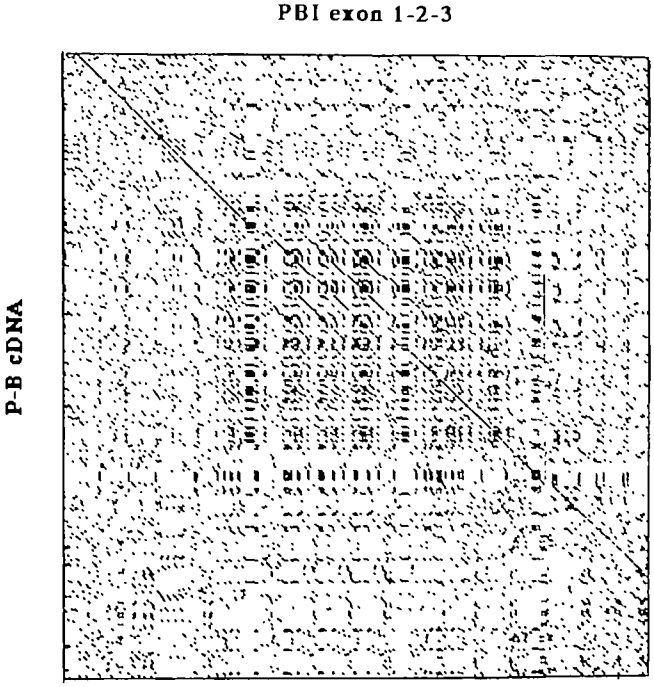

Fig. 3. **Comparison of PBI exons 1-3 and P-B cDNA by means of Harr plot.**

RESULTS

Sequence analysis of the plasmid subcloned from genomic clone B41 revealed that the 5′-terminal region of the insert of this plasmid was the 3′-terminal region of a gene closely related to, but not identical with that for proline-rich protein P-B. The upstream sequence of this P-B-related gene was determined by analyzing clone b21. Clone b21 covered the entire P-B related gene termed PBI, the approximate size of which was 6.4 kbp. Restriction maps of PBI and the sequencing strategy are shown in Fig. 1. The determined nucleotide sequence is shown in Fig. 2. In the light of the structure of P-B cDNA, this gene was considered to be composed of three exons. The size of the first exon was 88 bp but this value is approximate, since the transcriptional start site was not precisely determined. The putative transcriptional start point was assigned by considering that P-B cDNA has a T-rich region at its 5′ terminus (*7*) and that transcription usually starts at about 20 to 30 bases downstream of a TATA box (*13*). The first exon contained a 5′-untranslated region. The second exon was 68 bp long. This exon was composed of 14 bp of 5′ non-coding sequence and 54 bp of coding sequence corresponding to 18 amino acid residues of the signal sequence. The third exon was 508 bp long. This value is also approximate, since the polyadenylation site was not precisely determined. This exon contained 12 bp coding for 4 amino acid residues of the signal sequence, 336 bp coding for the entire mature protein (112 amino acid residues), and 160 bp of 3′ non-coding sequence from the stop codon TAA through the poly A addition signal AATAAA to the poly A addition site. Another AATAAA was found at 89 bp downstream of the first poly A addition signal. The sizes of the first and the second intron were 1,259 and 4,473 bp, respectively. Both introns satisfied the GT-AG rule (*14*) and had sequences homologous to the consensus sequence YNYRAY (Y=pyrimidine, R=purine, N=any base) required for branching for splicing (*15*) about 30 bp upstream of the 3′ end of each intron. The gene organization of PBI is summarized in Fig. 1.

The promoter sequence TATATAA (*13*) is found at 21-27 bp upstream of the putative transcriptional site. Another promoter sequence, CCAAT (*16*), was present at 127-131 bp upstream of the transcriptional site as an inverse complementary form. When the region upstream of the transcriptional site was computer-searched for transcription factor-binding sites, a USF (upstream stimulatory factor) binding site was found at about 400 bp upstream. The TATA box, CCAAT box, and USF binding site are shown in Fig. 2. GC boxes with a core consensus sequence GGGCGG, which have been shown to be present in several housekeeping genes (*17*), were not found.

Comparison of PBI Exons with P-B cDNA—Nucleotide sequence comparison of a combination of exons 1, 2, and 3 of PBI with P-B cDNA by means of a Harr plot is shown in Fig. 3. The result indicated that PBI is closely related to, but different from, P-B cDNA in terms of sequence. Here, we use the designation P-B1 for the gene product of PBI. A comparison of P-B1 precursor with P-B precursor is shown in Fig. 4. P-B1 precursor was composed of 22 residues of signal sequence and 112 residues of mature protein sequence. The mature protein sequence was composed of 3 regions, *i.e.*, 23 residues of N-terminal sequence, 5 proline-rich repeats of 13 to 14 residues, and 22 residues of C-terminal region. Each repeating unit usually starts with GPGXX′ (X,X′≠P) followed by a proline-rich peptide of several residues and ends with an aromatic amino acid. On the other hand, P-B precursor consists of 22 residues of signal sequence, 19 residues of N-terminal sequence, and 3

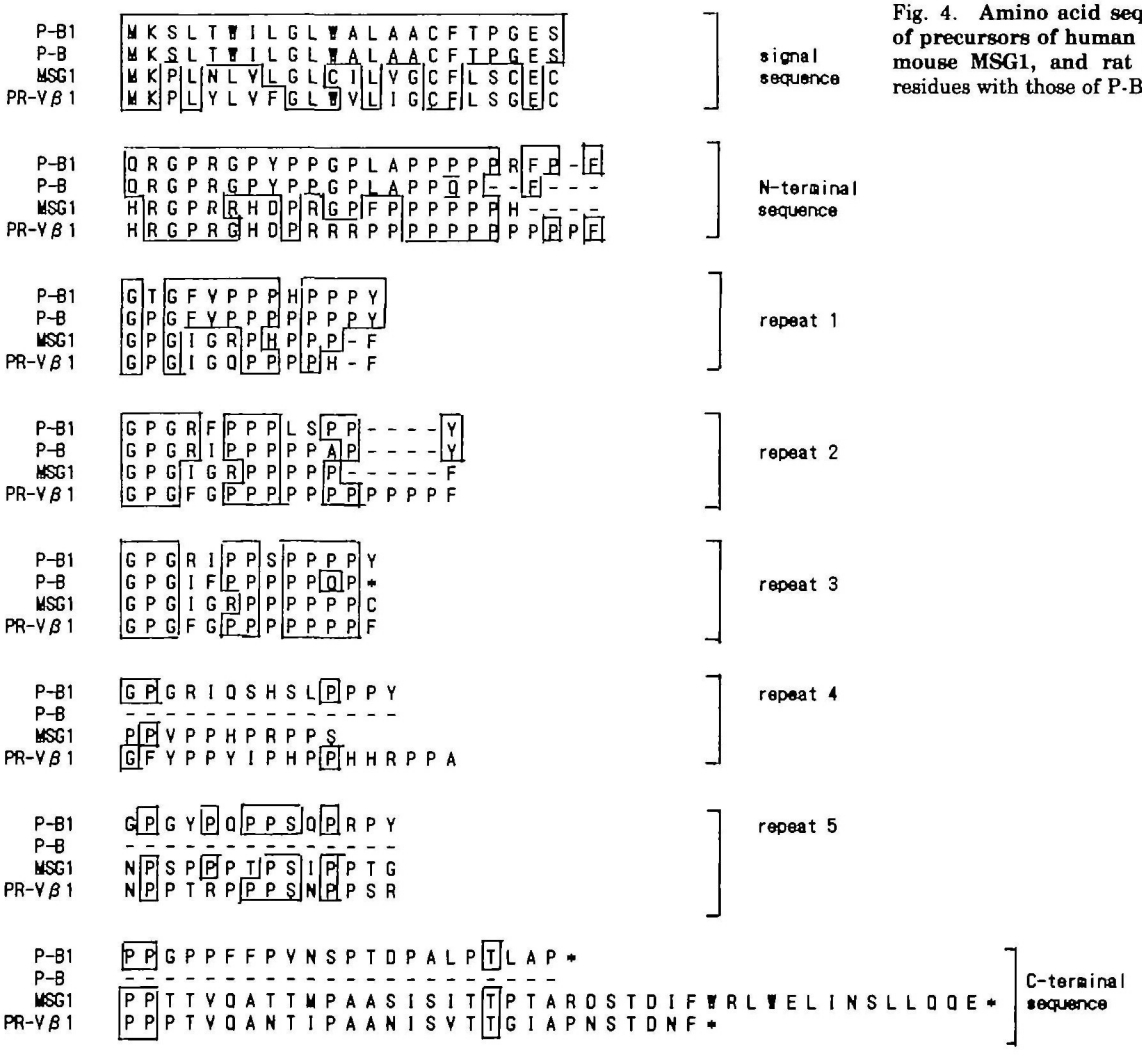

Fig. 4. **Amino acid sequence comparison of precursors of human P-B1, human P-B, mouse MSG1, and rat PR-Vβ1.** Identical residues with those of P-B are boxed.

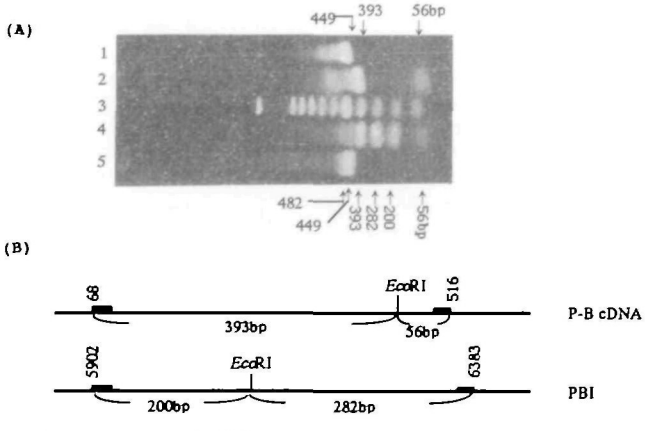

Fig. 5. **Location of PCR products on P-B cDNA and PBI and their analysis by agarose gel electrophoresis.** (A) Separation of PCR products and their *Eco*RI digests on 1.5% agarose gel. Lane 1, PCR product (template, human submaxillary gland cDNA); lane 2, *Eco*RI digest of PCR product in lane 1; lane 3, size marker of 100 bp ladder (100–1,000 and 1,500 bp); lane 4, *Eco*RI digest of the PCR product in lane 5; lane 5, PCR product (template, human genomic DNA). (B) Locations of expected PCR products and their *Eco*RI sites on P-B cDNA and PBI. Nucleotide numberings of P-B cDNA and PBI are based on Ref. 7 and this paper, respectively.

proline-rich repeats. The signal sequences and the N-terminal 16 residues of the precursors of P-B1 and P-B coincide completely. P-B1 is almost double the size of P-B.

Homology Search—Computerized homology search of PBI with known sequences in a data base revealed that P-B cDNA had the highest score and rat VCSβ1 gene (*18*) was next most homologous. VCSβ1 is the gene for the proline-rich protein PR-Vβ1 expressed in rat submaxillary gland. The gene organization of VCSβ1 gene is very similar to that of PBI (*18*). When the homology of the exons (exons 1–3) of PBI was searched in the data base, DNAs which gave the highest scores after P-B cDNA were those of proline-rich proteins MSG1 of mouse (*19*) and PR-Vβ1 (*18*) of rat submaxillary glands. The amino acid sequences of MSG1 and PR-Vβ1 are compared with those of P-B1 and P-B in Fig. 4.

Expression of PBI—In the previous study (*7*), we performed Northern blot analysis of human heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas by using the P-B cDNA fragment as a probe and found that only placenta gave a hybridized band. The present result that the same P-B cDNA fragment was hybridizable with PBI is considered to indicate that the placenta could be the expression site of P-B1. However, as the size of the positive

mRNA (5-8 kb) (7) is much larger than the expected size of P-B1 mRNA (0.8 kb), the mRNA of the placenta is clearly different from P-B1 mRNA. Some mechanism may exist by which the precursor mRNA of placenta is processed to the shorter P-B1 mRNA. Since human submaxillary gland cDNA was hybridizable with the P-B cDNA fragment and since the size of the positive band was about 800 bp (7), the human submaxillary gland may express P-B1 together with P-B. To determine whether PB-1 is expressed in the submaxillary gland, PCR was performed by using oligonucleotides shared by both P-B and P-B1 cDNAs as primer pairs and human submaxillary gland cDNAs as a template, and *Eco*RI digests of the PCR products were analyzed electrophoretically. Bands corresponding to 393 and 56 bp were observed, but those corresponding to 200 and 282 bp were not. The former two bands were expected to be derived from cDNA coding for P-B, while the latter two were expected to be detected if cDNA coding for P-B1 was present in human submaxillary gland cDNAs. When human genomic DNA was used in place of human submaxillary gland cDNAs as the template in PCR, the expected two bands from P-B and those from PBI were all detected (Fig. 5). It appears that P-B1 is not expressed in the human submaxillary gland.

DISCUSSION

We have studied the P-B gene in order to elucidate its structure and to find out whether it belongs to a distinct gene family. The presence in bovine developing enamel of a proline-rich protein with an apparent molecular weight of 31,000, which is much larger than P-B (molecular weight 5,793), but whose amino-terminal 57 residues are identical with those of P-B (*8*), suggested that cloning might afford a gene different from the P-B gene. Indeed the gene isolated in the present work was not P-B gene itself but a homologous one. The putative translation product (P-B1) of this newly isolated gene has a very high amino acid sequence homology with P-B. P-B1 is also different from the protein found in the bovine developing enamel having the 57 residues at the amino-terminal identical with those of P-B.

The nucleotide sequence homology and similarity in gene organization suggest that the rat gene family to which VCSβ1 belongs (*18*) is the rat homolog of the human PBI gene family. This gene family may be ubiquitous among animal species, since Southern blot analysis using the P-B cDNA fragment as a probe revealed multiple positive signals in all the animals studied (7). It has been reported that P-B-related proteins of mice and rats are expressed in the salivary gland (*18, 19*). Although human P-B is expressed in saliva (7) , P-B1 appears not to be expressed in the human submaxillary gland. Since PBI contains the nucleotide motifs required for transcription, splicing, translation, and polyadenylation (*13-16*), this gene could be expressed under specific conditions, such as mechanical and/or chemical stimulation in the oral cavity. Comparison of the PBI gene with the P-B gene, which actively produces P-B in the human adult submaxillary gland should provide clues to understand the difference in expression between these two genes. The P-B gene structure is under investigation.

Prohormone-like protein SMR1, the gene product of the gene VCSα1, belonging to the rat homolog of the human PBI gene family, has the tetrapeptide Gln-His-Asn-Pro surrounded by two pairs of basic residues that represent potential processing sites. Peptides formed by digestion of SMR1 at these sites were detected in rat saliva and blood (*18*). Since PBI, like P-B, has no such sequence surrounded by two pairs of basic residues, these molecules may not be processed to produce hormone-like peptides.

REFERENCES

1. Bennick, A. (1987) Structural and genetic aspects of proline-rich proteins. *J. Dent. Res.* **66**, 457-461
2. Moreno, E.C., Varughese, K., and Hay, D.I. (1979) Effect of human salivary proteins on the precipitation kinetics of calcium phosphate. *Calcif. Tissue Int.* **28**, 7-16
3. Mehansho, H., Hagerman, A., Clements, S., Butler, L., Rogler, J., and Carlson, D.M. (1983) Modulation of proline-rich protein biosynthesis in rat parotid glands by sorghums with high tannin levels. *Proc. Natl. Acad. Sci. USA* **80**, 3948-3952
4. Maeda, N., Kim, H.-S., Azen, E.A., and Smithies, O. (1985) Differential RNA splicing and post-translational cleavage in the human salivary proline-rich protein gene system. *J. Biol. Chem.* **260**, 11123-11130
5. Kim, H.-S., Lyon, K.M., Saitoh, E., Azen, E.A., Smithies, O., and Maeda, N. (1993) The structure and evolution of the human salivary proline-rich protein gene family. *Mammalian Genome* **4**, 3-14
6. Isemura, S., Saitoh, E., and Sanada, K. (1979) Isolation and amino acid sequences of proline-rich peptides of human whole saliva. *J. Biochem.* **86**, 79-86
7. Isemura, S. and Saitoh, E. (1994) Molecular cloning and sequence analysis of cDNA coding for the precursor of the human proline-rich peptide P-B. *J. Biochem.* **115**, 1101-1106
8. Strawich, E. and Glimcher, M. (1990) Tooth enamelins identified mainly as serum proteins. *Eur. J. Biochem.* **191**, 47-56
9. Ito, S., Suzuki, T., Momotsu, T., Isemura, S., Saitoh, E., Sanada, K., and Shibata, A. (1985) Presence of salivary protein C and salivary peptide P-C-like immunoreactivity in the laryngo-tracheo-bronchial glands. *Acta Endoclinol.* **108**, 130-134
10. Isemura, S., Saitoh, E., Ito, S., Isemura, M., and Sanada, K. (1984) Cystatin S: A cysteine proteinase inhibitor of human saliva. *J. Biochem.* **96**, 1311-1314
11. Horn, B. (1979) *In vitro* packaging of λ and cosmid DNA in *Methods in Enzymology* (Wu, R., ed.) Vol. 68, pp. 299-309, Academic Press, New York
12. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
13. Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) The ovalbumin gene-sequence of putative control regions. *Nucleic Acids Res.* **8**, 127-142
14. Mount, M.S. (1982) A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**, 459-472
15. Sharp, P.A. (1987) Splicing of messenger RNA precursors. *Science* **235**, 766-771
16. Jones, K.A., Kadonaga, J.T., Rosenfeld, P.J., Kelly, T.J., and Tjian, R. (1987) A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. *Cell* **48**, 79-89
17. Dynan, W.S. (1986) Promoters for housekeeping genes. *Trends Genet.* **2**, 196-197
18. Courty, Y., Rosinski-Chupin, I., and Rougeon, F. (1994) A new proline-rich protein precursor expressed in salivary glands of the rat is encoded by a gene homologous to the gene coding for the prohormone-like protein SMR1. *J. Biol. Chem.* **269**, 520-527
19. Tronik-Le Roux, D., Senorale-Pose, M., and Rougeon, F. (1994) Three novel SMR1-related cDNAs characterized in the submaxillary gland of mice show extensive evolutionary divergence in the protein coding region. *Gene* **142**, 175-182