Biochem. J. (1991) 276, 699–707 (Printed in Great Britain)

Domain structure and sequence distribution in dentin phosphophoryn

Boris SABSAY, William G. STETLER-STEVENSON,* Joseph H. LECHNER† and Arthur VEIS‡
Northwestern University, Division of Oral Biology, 303 E. Chicago Ave., Chicago, IL 60611, U.S.A.

Phosphophoryn (PP) is a protein unique to the mineralized matrix of dentin. It also has a unique composition, with aspartic acid and phosphoserine comprising > 85% of all amino acid residues. Because of this unique composition and high content of phosphoserine, it has been difficult to apply direct peptide sequencing procedures effectively. However, to understand its function, and to prepare suitable probes for screening cDNA libraries, some sequence distribution information is required. To this end, using bovine (b) and rat incisor (ri) PPs, partial mild acid hydrolysis has been used to cleave at the aspartic acid residues and generate free amino acids and small peptides. The nature of the released amino acids and peptides has been determined. Peptides have also been generated by limited digestion with trypsin. Some of the peptides have been purified by h.p.l.c. techniques and sequenced. About 90% of the bPP and riPP were resistant to trypsin, and the large resistant fragment was sharply depleted of the non-aspartic acid and non-phosphoserine [(P)Ser] residues. All peptides isolated were acidic, but the remaining residues (other than aspartic acid and serine) appeared to be collected in regions flanking the trypsin-resistant core. These data show directly the presence of regions $[Asp]_n$, $[(P)Ser]_m$ and $[Asp\text{-}(P)Ser\text{-}Asp]_k$ as prominent sequence features. A domain structure model is proposed.

INTRODUCTION

Phosphophoryn (PP) is the generic name of a group of compositionally unique, highly phosphorylated, proteins which are ubiquitous components of the dentin extracellular matrix (Dimuzio & Veis, 1978). Bovine dentin contains a single PP (Lee *et al.*, 1977; Stetler-Stevenson & Veis, 1983), bPP, whereas rat incisor (ri) dentin contains two or three phosphophoryns (Dimuzio & Veis, 1978; Butler *et al.*, 1983), riPP. The bPP has been thoroughly characterized (Stetler-Stevenson & Veis, 1983). It has a high and specific affinity for Ca^{2+} (Lee *et al.*, 1977; Zanetti *et al.*, 1981; Stetler-Stevenson & Veis, 1987), forms a ternary complex with Ca^{2+} and PO_4^{3-} (Lee *et al.*, 1983), and shows specific interactions with type I collagen fibrils (Stetler-Stevenson & Veis, 1986). In the rat-incisor system, autoradiographic (Weinstock & Leblond, 1973) and biochemical (Maier *et al.*, 1983) studies following pulse labelling, as well more recent histochemical (Takagi *et al.*, 1986; Nakamura *et al.*, 1985) and immunohistochemical localization studies (Nakamura *et al.*, 1985; McDougall *et al.*, 1985; Tsay & Veis, 1985; Rahima *et al.*, 1988) of bovine, rat and murine dentin have supported the conclusion reached by direct chemical analyses (Carmichael *et al.*, 1978; Jontell & Linde, 1983) that, in the process of secretion, the PP by-passes the predentin and is deposited directly at the mineralization front. All of these studies support the hypothesis (Veis & Sabsay, 1983) that at least one of the PP species might be involved in the initiation of the collagen-specific mineralization in the dentin matrix.

In order to explore the mechanism of the interaction of the PP with collagen, and to understand their role in the mineralization process, we feel that it is necessary to know the amino acid sequence of the PP. Direct attempts at sequencing have not been very successful, and only very short *N*-terminal sequences have been reported for two of the riPP species (Butler *et al.*, 1983). The approaches of molecular biology to this problem are just beginning to be used. A useful tool would be the construction of suitable nucleotide probes, but some direct sequence information is required. Preliminary evidence suggested that the bPP molecule might have distinct subdomains (Lechner *et al.*, 1981), and even in the absence of detailed sequence information, it seemed possible to analyse the problem from the point of view of the functions of the specific molecular domains. In the work reported here we have examined the concept that both bPP and riPP have domain structures. In addition, some partial direct sequence information is presented.

MATERIALS AND METHODS

Phosphophoryns

Bovine PP was isolated from unerupted third molars as previously described (Stetler-Stevenson & Veis, 1983). The amino acid composition, electrophoretic behaviour and phosphorus contents were in accord with data obtained previously. Foetal bPP (Termine *et al.*, 1980) was generously provided by Dr. J. D. Termine of the National Institute of Dental Research. Iodinated bPP was prepared using a commercial coupled, immobilized lactoperoxidase–glucose oxidase system (NEN; catalog no. NEZ0151) (Stetler-Stevenson & Veis, 1986). ^{125}I-bPP solutions were lyophilized and stored at −20 °C until use.

Freshly excised rat incisors were cleaned, demineralized and extracted as described by Rahima & Veis (1988). The riPP was collected from the EDTA extract by precipitation with 1.0 M-$CaCl_2$. The precipitated riPP was further purified by chromatography on a preparative Beckman Spherogel TSK DEAE-5PW h.p.l.c. column (21.5 mm × 15 cm) in 50 mM-Tris/HCl, pH 7.5. The riPPs were eluted in the ionic-strength range 0.48–0.56 M-NaCl. This peak contained several components with different molecular masses. The riPP fraction was chromatographed, on a preparative scale, over a Zorbax GF-250 XL Bio-Series gel-filtration column (21.2 mm × 25 cm), in 0.1 M-Na_2HPO_4, pH 7.5, or in 0.1 M-ammonium formate. The void-

Abbreviations used: PP, phosphophoryn; b, bovine; ri, rat incisor; f, foetal; PMSF, phenylmethanesulphonyl fluoride; TFA, trifluoroacetic acid.

* Present address: Laboratory of Pathology, NCI, NIH, Bethesda, MD, U.S.A.
† Present address: Department of Chemistry, Mount Vernon Nazerene College, Mount Vernon, OH 43050, U.S.A.
‡ To whom correspondence should be addressed.

volume peak was collected and re-run under the same conditions. No low-molecular-mass components were evident in the final fraction. The first half of the void peak was taken as the starting material for all further study.

Trypsin digestion

bPP (0.7 mg/ml), ^{125}I-bPP (0.45 mg/ml) and fetal bPP (fbPP) (2 mg/ml) dissolved in either distilled water or 1.0 M-NaCl/0.05 M-Tris/HCl, pH 7.5, were digested with trypsin (Sigma) at 4 °C at an enzyme/substrate ratio of 1:10 (w/w). The time course of the digestion was monitored by removing 20 μl aliquots at regular intervals and diluting them with an equal volume of concentrated electrophoresis sample buffer. Gel-electrophoretic analyses were then carried out. Alternatively, aliquots were removed at regular intervals and analysed by direct application on to a TSK G 3000SW column (0.75 cm × 30 cm) (LKB Instruments Co., Uppsala, Sweden) equilibrated in 0.1 M-NaCl/0.05 M-Tris/HCl, pH 7.5, and chromatographed at a flow rate of 0.5 ml/min. As a result of the time-course studies, limit digestions were carried out on a larger scale for 2 h at 4 °C. These reactions were stopped by the addition of phenylmethanesulphonyl fluoride (PMSF) to a final concentration of 0.2 mM.

Trypsin digestions of the riPP (1 mg/ml) were carried out in 0.2 M-Na_2HPO_4 buffer, pH 7.5. Trypsin (1:100) was added and the digestion mixture was incubated at 35 °C for 20 min with gentle shaking. The reaction was stopped by heating the mixture to 100 °C for 15 min. This treatment corresponded to the limit digestion conditions for bPP at 4 °C described above. The digestion mixture was chromatographed on a Zorbax G 250 gel-filtration column, as described above, but in the analytical mode, and fractions were collected. These fractions were subsequently used for reverse-phase h.p.l.c. The trypsin digestion mixtures were subjected to gel electrophoresis.

Partial acid hydrolysis

Two methods were used for partial acid hydrolysis. This approach was based on the work of Schultz *et al.* (1962) who showed that, under mild acid conditions, aspartic acid could be specifically cleaved out of peptide chains. This was thought to be an ideal approach to the aspartic acid-rich PP.

Acetic acid hydrolysis. In the first series of experiments, which focused on the release of amino acids and small peptides from bPP, the acetic acid digestion procedure of Krippner & Nawrot (1977) was followed. bPP (7 mg) was dispersed in 10 ml of 0.25 M-acetic acid at room temperature. Intact PP is not readily soluble in dilute acetic acid, and the PP was a suspension at this point. A 1 ml aliquot was dried, then hydrolysed in 6 M-HCl for 22 h at 108 °C. The total amino acid composition was determined quantitatively on a JEOL 6AH analyser using ninhydrin detection. The remaining 9 ml in acetic acid was digested for 48 h at 108 °C. After digestion and cooling to room temperature a small amount of undissolved material, presumably undigested PP, was still present. This was removed by centrifugation prior to further analysis.

The partial hydrolysate was again divided into two portions, the first of which, a 1 ml aliquot, was analysed without further treatment on the amino acid analyser. This analysis yielded the content of free amino acids released. The remainder of the partial hydrolysate was chromatographed on a 1.2 cm × 130 cm column of Bio-Gel P-2 equilibrated with 1 mM-Tris/HCl, pH 8.0, containing 0.33 mM-EDTA. Elution was carried out at 1 ml/min, and the absorbance of the eluate was monitored at 230 nm. Fractions (2 ml) were collected. Fractions defined by absorbance were pooled, desalted and hydrolysed in 6 M-HCl as described above, and their amino acid compositions were determined. The Bio-Gel column was calibrated with small peptides of known composition, and the elution positions of aspartic acid, serine and phosphoserine were determined. Total column volume was measured by the elution position of ^3H_2O.

Formic acid hydrolysis. Dilute formic acid digestion is less efficient and hence yields somewhat larger peptides, some of which retain aspartic acid (Inglis, 1983; Rusenko, 1988). The riPP was dissolved in 2% formic acid at 200 μg/ml and hydrolysed at 2, 4 or 6 h at 108 °C. The peptides were run over the preparative DEAE column exactly as in the initial purification of the PP. After cleavage and removal of some of the aspartic acid residues, the residual peptides were less acidic and were eluted earlier in the gradient. The least acidic peptides were collected; these were passed over the Zorbax G-250 column and eluted with the volatile buffer (0.1 M-ammonium formate). The eluted peptides were clearly separated from the 'salt' peak. Final peptide purification was achieved by reverse-phase chromatography on a du Pont Poly F Bio-Series column (6.2 mm × 8 cm) using either of two elution systems: (1) Buffer A [0.1% trifluoroacetic acid (TFA) in water]; Buffer B (0.1% TFA in 100% acetonitrile); (2) Buffer A (0.47% ammonium formate in water); Buffer B (0.33% ammonium formate in 70% acetonitrile in water). The final purified peptides were collected for amino acid analysis and *N*-terminal sequencing analysis.

Gel electrophoresis

Electrophoresis analyses were carried out as previously described (Stetler-Stevenson & Veis, 1983), using the Laemmli (1970) system. The gels were stained with Stains All according to the method of Green *et al.* (1973). Autoradiography of ^{125}I-labelled PP and peptides resulting from tryptic degradation of the iodinated bPP was performed at −70 °C using XRP-5 film (Kodak) after drying the gels.

Amino acid analysis and peptide sequencing

Depending on the amount of material available, amino acid analyses were carried out on a JEOL 6AH analyser (ninhydrin) with post-column colour development, or using the Fmoc procedure of pre-column derivatization followed by reverse-phase chromatography. A Varian Micropak SP C18 Amino Tag column and Amino Tag buffers were used [Cunico *et al.*, 1986].

Peptide sequencing analysis was carried out by Dr. Ka-Leung Ngai at the Northwestern University Biotechnology Research Service Facilities, Evanston, IL, U.S.A., using an Applied Biosystems model 477A/120A apparatus.

RESULTS

The experiments reported here were carried out in three phases over a period of several years. The first studies used bPP as the starting material, since it appeared to be more homogeneous than the riPP, but, in order to generalize these data, the later experiments utilized well-fractionated riPP.

bPP

Partial acid hydrolysis. The sequence-related information desired from partial acid hydrolysis is based on the relative instability of the peptide bonds surrounding the aspartic acid residue as compared with all others in a peptide chain (Inglis, 1983). However, phosphoserine within a peptide sequence is also quite labile to mild-acid-hydrolysis conditions (Masters, 1985). In the earlier study by Krippner & Nawrot (1977) bPP was hydrolysed for 120 h at 110 °C in 0.25 M-acetic acid. Under these conditions, 80% of the aspartic acid was freed, and about 40% of the serine also appeared as the free amino acid. We explored

Table 1. Free amino acids released from bPP during partial acid hydrolysis in 0.25 M-acetic acid at 108 °C for 48 h

Results are given as nmol/sample. Equal samples of bPP were taken for total and partial digestions and equal amounts were injected into the analyzer. Each analysis was in duplicate. Norleucine was used as internal standard.

Amino acid	Total composition (6 M-HCl hydrolysis)	Released free amino acids (partial hydrolysis)
Lys	73.5	Trace
His	6.5	Trace
Arg	2.0	0
Asp	535	231
Thr	16	Trace
Ser	498	61
PSer	92	19
Glu	46	Trace
Pro	15	0
Gly	63	4.7
Ala	19	5.2
½-Cys	3.1	0
Val	12	0
Met	2.5	0
Ile	4.9	0
Leu	9.2	0
Tyr	Trace	0
Phe	4.2	0

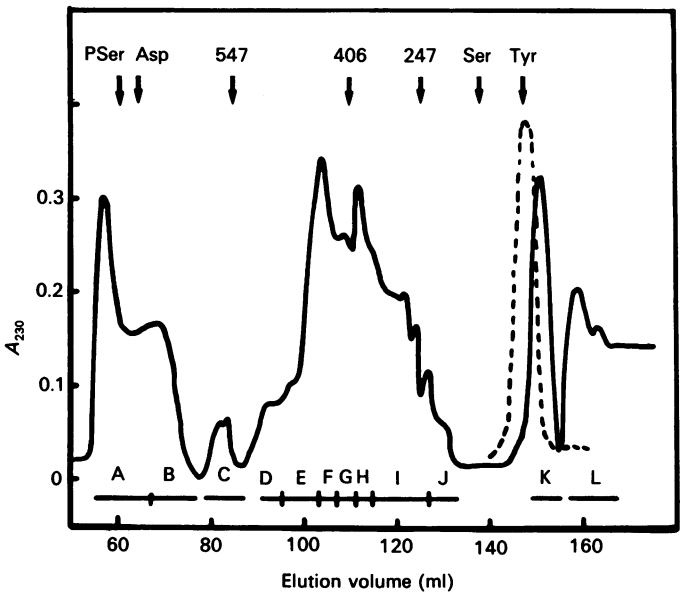

Fig. 1. Bio-Gel P2 chromatography of the small peptides released by the 0.25 M-acetic acid, 48 h, 108 °C hydrolysis of bPP

Elution was with 1.0 mM-Tris/HCl/0.33 mM-EDTA, pH 8.0. The arrows along the top indicate the elution positions of the specified free amino acids. The numbers indicate the molecular masses of standard di-, tri- and penta-peptides in this system. Note the anomalous elution of both free aspartic acid and phosphoserine in this system. Peaks A and B contain those amino acids as well as other anionic peptides. Each region denoted by the lettered bars was collected. The broken-line peak is the elution position of ^3H_2O.

hydrolysis times from 0 to 120 h and measured aspartic acid release compared with all other amino acids. Free amino acids other than aspartic acid (serine, phosphoserine, glycine and alanine) increased in amount up to about 48 h of hydrolysis and then remained constant in amount between 48 and 60 h. Longer hydrolysis times produced further non-specific increases. Therefore the release of each amino acid at 48 h was selected as the optimum for specificity of aspartic acid cleavage in 0.25 M-acetic acid.

Identical aliquots of fully hydrolysed and partially hydrolysed bPP, each containing an identical quantity of norleucine as an internal standard, were compared on the analyser (ninhydrin). As indicated by the data in Table 1, after 48 h of acetic acid hydrolysis, 43% of the total aspartic acid content was released as the free amino acid. Serine (combined serine + phosphoserine accounted for 13.6% of total content) was the next most abundant amino acid released, followed by alanine and glycine in much smaller amounts. These data suggest that PP contains a number of sequences of the type [Asp-Yaa-Asp] where Yaa is the freed amino acid. Serine (or phosphoserine) was the predominant occupant of the Yaa position, indicating the presence of some $[Asp\text{-}(P)Ser\text{-}Asp]_k$ sequences.

To examine some of the peptides produced by the acid hydrolysis, a 48 h hydrolysate was centrifuged and the supernatant was chromatographed over Bio-Gel P-2. The column was calibrated with small peptides of known size as indicated in Fig. 1. In a separate experiment the elution positions of each of the free amino acids (aspartic acid, serine and phosphoserine) were determined. The chromatogram in Fig. 1 showed that a wide range of fragments were obtained. As indicated, the free amino acids were eluted at positions away from the positions of several of the 'peptide' peaks.

The most interesting fragments were in the di- and tri-peptide range. Region J consisted entirely of serine residues, as the dimer (P)Ser–(P)Ser. The components in fraction I were of tripeptide size and were 75% (P)Ser, with an additional 10% lysine. The major components of region I must have been $[(P)Ser]_3$ and $[(P)Ser_2Lys]$. These data do not permit one to distinguish between serine and phosphoserine, hence the notation (P) is used to express this uncertainty. However, it should be recalled that most of the serine residues in bPP are phosphorylated.

Finally, it is important to note that the amino acids lysine, glutamic acid, threonine, proline, valine, leucine, isoleucine, phenylalanine and tyrosine are not likely to be placed in sequences bounded on both sides by aspartic acid, since none of these residues was released as free amino acid in the partial acid hydrolysates.

bPP: Tryptic digestion. Fig. 2 shows the time course of tryptic digestion of ^{125}I-labelled bPP in terms of the sizes of the fragments produced. These gel-electrophoretic data, in a 12% -(w/v)-polyacrylamide gel, show that the bPP underwent an immediate initial cleavage resulting in a small ^{125}I-labelled fragment and a residual, but still labelled, high-molecular-mass fragment. With longer digestion time, this high-molecular-mass fragment was cleaved to a slightly smaller, trypsin-resistant, but non-^{125}I-labelled, major fragment. There was a concomitant release of a second, small, ^{125}I-labelled peptide. The residual high-molecular-mass fragment yielded an M_r in the range 135000–140000, in a 5–15% gradient gel (Stetler-Stevenson & Veis, 1983), compared with the initial M_r of approx. 150000. The molecular masses of the small fragments were too low to be accurately determined in these gels.

A gel-filtration h.p.l.c. analysis of the 120 min digestion of ^{125}I-bPP at a high loading concentration was carried out (Fig. 3). Three small peptides (a, b and c) not present in the non-trypsin-

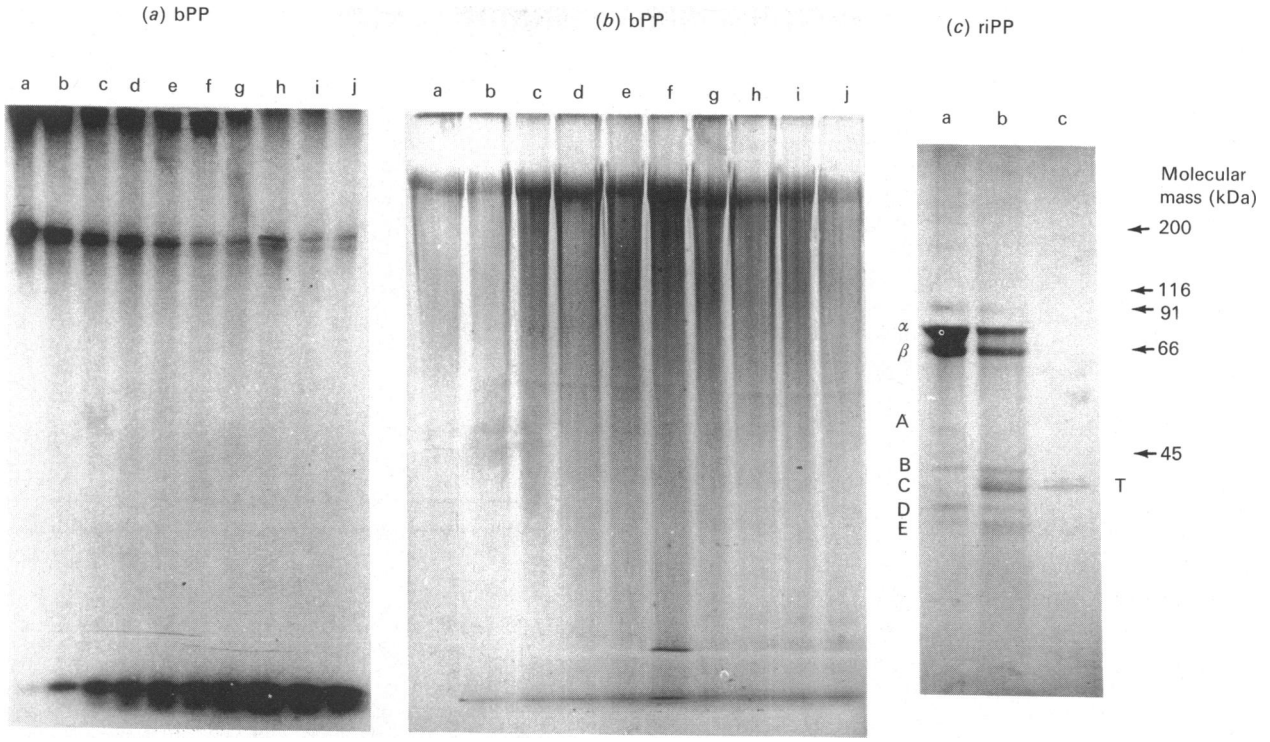

Fig. 2. Gel electrophoretic analysis of tryptic digests of bPP and riPPs

(*a*) Autoradiogram of ^{125}I-labelled bPP digests at a bPP:trypsin ratio of 10:1 as a function of digestion time. In 12% acrylamide/SDS gels. (*b*) Equivalent Stains All staining of the bPP digests of (*a*). In each gel: a, bPP substrate; b, zero time after addition of trypsin; c, 5 min; d, 10 min; e, 20 min; f, 30 min; g, 45 min; h, 60 min; i, 90 min; j, 120 min. (*c*) A limit digest of crude riPP (mixture of α- and β-riPP), Stains All staining, 5–15% gradient acrylamide/SDS gels: a, initial riPP; b, after digestion; c, trypsin alone at level in lane b. α, α-riPP; β, β-riPP; A, B, C, D, E are degradation bands; T, trypsin. Positions of molecular-mass markers are indicated by arrows.

treated bPP eluted at a position just ahead of the rather large trypsin peak. Two of these (b and c) had substantial radioactivity. These data were in general agreement with the results of the gel-electrophoresis data in that the bulk of the bPP was eluted at the column void volume. However, in gel-filtration chromatography the cleaved peptides appeared to have a higher molecular mass than trypsin itself.

The bPP obtained from fetal teeth (fbPP) is somewhat different from that in mature teeth, having a higher content of hydrophobic amino acid residues (Termine *et al.*, 1980). Nevertheless, tryptic digestion of fbPP, followed by h.p.l.c. gel filtration as a function of time (Fig. 4), showed the similar release of three small peptides and the retention of a high-molecular-mass trypsin-resistant fragment that was eluted in the void volume of the column. From the time course of the reaction and the relative heights of the peaks, it appeared that the peaks designated a and c were released very rapidly, whereas peak b was released more slowly. The relative migration positions of these showed no change with increased digestion times, and the intensity of the zero-time peptide band remained constant. Thus it is most likely that these peptides were cleaved independently.

Fetal bPP was digested for 2 h at 4 °C with trypsin at a 10:1 substrate:enzyme ratio on a semi-preparative scale. The high-molecular-mass trypsin-resistant fbPP fragment was collected from the void-volume peak of the h.p.l.c. gel-filtration column, and its amino acid composition was compared with the composition of the intact fbPP. These data, presented in Table 2, show that the tryptic digestion resulted in the removal of most of the non-polar residues from the high-molecular-mass fraction, leaving a core enriched in aspartic acid, phosphoserine and

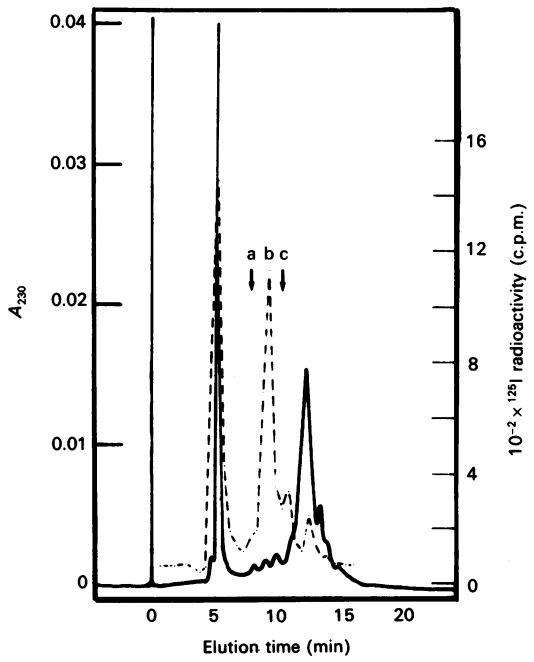

Fig. 3. Fractionation of a 120 min tryptic digest of ^{125}I-labelled bPP on a TSK 3000 SW gel-filtration column (bPP:trypsin = 10:1)

Elution was with 0.1 M-NaCl/0.05 M-Tris/HCl, pH 7.5. ——, A_{230}; – – – –, radioactivity.

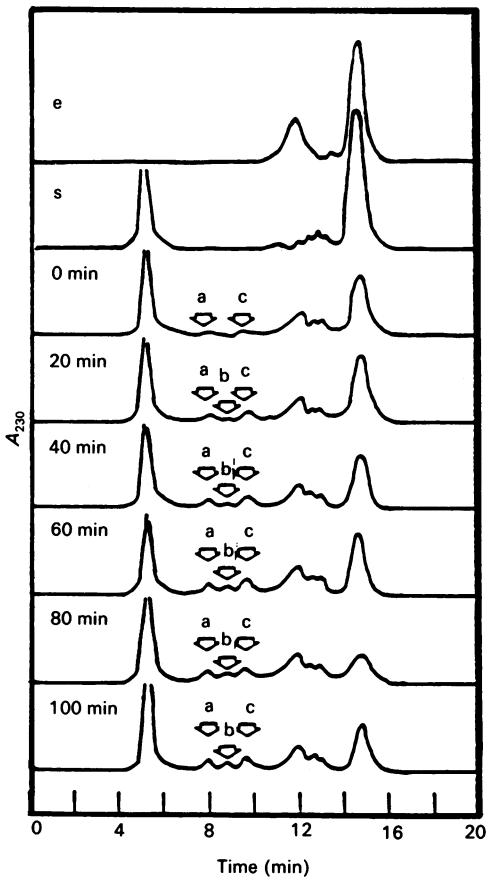

Fig. 4. Time course of the tryptic digestion of fbPP, followed by gel filtration h.p.l.c. on a TSK 3000 SW column

The fbPP was digested at a substrate:trypsin ratio of 10:1 at 4 °C for the indicated periods. Identical aliquots of the digestion mixture containing 7 μg of fbPP were injected into the column in 0.1 M-NaCl/0.05 M-Tris/HCl, pH 7.5. The column was eluted at a flow rate of 0.5 ml/min. The open arrows mark the elution of the three largest cleaved tryptic peptides. e, The enzyme blank, no fbPP; s, the substrate blank, fbPP alone. The cleaved peptides (a, b and c) emerge in a region free of any components of e and s. The trypsin is itself degraded during the reaction in the presence of fbPP, but not in its absence.

serine. The right-hand column of Table 2 shows the calculated estimated composition of the total cleaved peptides, based on a loss of 13% of the initial molecular mass by tryptic digestion. Although the small fragments are relatively enriched in non-polar residues (31.2% versus 6.0% in the intact fbPP), it is noteworthy that the small peptides contain substantial amounts of serine (or phosphoserine), glutamic acid, proline and glycine and remain acidic in character.

The finding of a high-molecular-mass fragment in both bPP and fbPP, enriched in serine and aspartic acid residues, and resistant to further tryptic cleavage, indicated that the smaller tryptic peptides must have been cleaved either sequentially from one end or independently from both ends of the intact bPP and fbPP molecules.

At the time these studies of bPP were carried out, we were not in a position to examine the sequences of any of the cleaved peptides. These data were, however, used as the baseline for our study of the riPP described below.

riPP

Preparation of a homogeneous starting material. The procedure

Table 2. Relative amino acid compositions of intact fbPP and the high-molecular-mass void-volume tryptic-digestion fragment (HMM-TR-fbPP)

Results are given as residues per 1000 amino acid residues and are the means of triplicate analyses. Values for HMM-TR-fbPP are calculated on the basis of the weight of the original fbPP, that is, corrected for decrease in molecular mass after digestion. The 'Difference' is the computed total composition of all small tryptic fractions from the difference in compositions [fbPP−(HMW-TR-fbPP)]. Values in parentheses represent the compositions as residues per 1000 for the total cleaved peptides.

Amino acid or amino-sugar	Initial composition of fbPP	Fraction composition after digestion	
		HMW-TR-fbPP	Difference
Polar			
Lys	49	38	11 (90)
His	11	1	10 (82)
Arg	4	Trace	4 (33)
Asp	359	360	0 (0)
Thr	10	1	9 (73)
Ser+PSer	461	442	19 (155)
Glu	20	4	16 (131)
½-Cys	1	Not detected	1 (8)
Pro	11	Trace	11 (90)
Non-polar			
Gly	32	11	25 (204)
Ala	11	5	6 (49)
Val	5	3	2 (16)
Ile	3	3	0 (0)
Leu	5	2	3 (25)
Tyr	2	Trace	2 (16)
Phe	2	Not detected	2 (16)
Glucosamine	2	Not detected	2 (16)
Galactosamine	5	Not detected	5 (41)
Total...	998	870	128 (988)
Non-polar (%)...	6.0	2.7	31.2

of Rahima & Veis (1988) was used to obtain a standard preparation of $CaCl_2$-precipitated, DEAE-purified, riPP. As usual, this preparation showed the presence of substantial molecular-mass heterogeneity on gel electrophoresis (Fig. 2*c*). The crude riPP was passed over a Zorbax-G250 column, and the void-volume peak was collected. This fraction was run once more over the G-250 column and the void-volume peak collected as the starting material for degradation studies. When re-run over the DEAE-column, a single peak corresponding to the main peak of the initial DEAE-chromatography was obtained. A single band was also obtained following gel electrophoresis and Stains All staining.

Partial acid hydrolysis. The procedure of Rusenko (1988), digestion in 2% formic acid for 4 h at 108 °C, was selected for partial acid hydrolysis in an attempt to obtain some larger peptides than was possible with the more drastic acetic acid cleavage. On the basis that aspartic acid would be removed selectively and that the remaining peptides would be overall less acidic, the hydrolysates were re-chromatographed over the DEAE h.p.l.c. column used for the initial isolation of intact riPP. Fig. 5 compares chromatograms of the intact purified riPP and the formic acid hydrolysate. Note that the intact riPP is eluted as a single sharp peak. The limited hydrolysis very consistently produced the same distribution of components shown in Fig. 5, indicating that there is a particular set of especially labile aspartic acid residues within the riPP. All of the major components in the

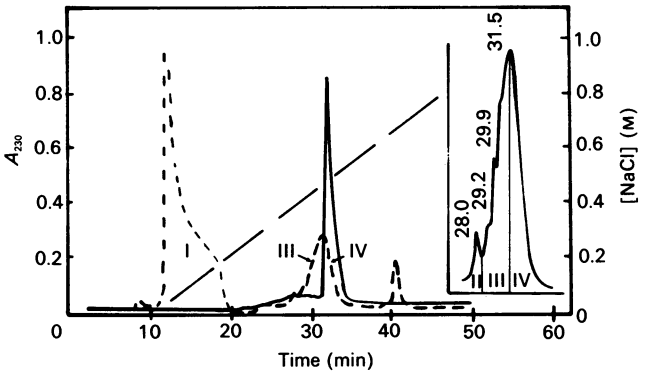

Fig. 5. DEAE-chromatography of purified and formic acid hydrolysed riPP on Beckman Spherogel TSK DEAE-5PW in 50 mM-Tris/HCl, pH 7.5, with an increasing NaCl gradient

——, Purified undegraded riPP; ----, riPP after partial hydrolysis. The inset shows the detail of the elution of the retained fractions after formic acid hydrolysis. The fractions collected are denoted by Roman numerals. These are also the prefix numerals for the origins of the fractions listed in Table 3. The very large initial peak is the result of the formic acid introduced with the sample; there is, however, a peptide component in fraction I.

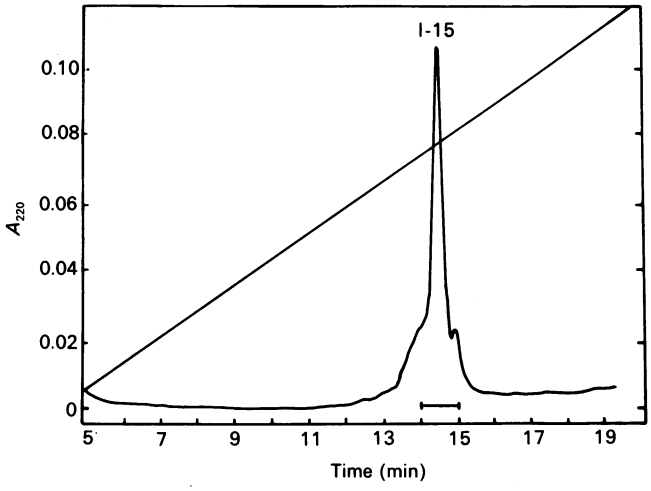

Fig. 6. Reverse-phase chromatography of DEAE-I on Zorbax Poly-F, eluted with ammonium formate plus an increasing gradient of acetonitrile

The central portion of the major peak (see the bar), designated as I-15, was collected for sequencing. Data are reported in Table 4. The diagonal line indicates the acetonitrile gradient.

hydrolysate were in the molecular-mass range between 17000 and 1300 (globular protein standards), on the Zorbax G250 gel-filtration column.

The high absorbance of DEAE-I (Fig. 5) was the result of the elution of the formate ion in the hydrolysis buffer, but the peak did contain protein. This was isolated, as shown in Fig. 6, as a single peak, I-15, after reverse-phase chromatography. The amino acid composition and the *N*-terminal amino acid sequence were determined. Fractions DEAE-II, III and IV were treated in similar fashion, except that, because of the high salt concentration, they were first run over the Zorbax G-250 column using the volatile 0.1 M-ammonium formate system for elution. The main components from the gel filtration were then subjected to reverse-phase chromatography. The best-resolved peptides were selected for amino acid analysis and *N*-terminal sequencing. The amino acid compositions of the peptides are shown in Table 3 and the amino-terminal sequences in Table 4. The fraction designations reflect the DEAE-column-fraction origin and the approximate time of the elution from the reverse-phase chromatography.

Table 3 shows that each of the peptides isolated was different in composition; nevertheless, every peptide was clearly very acidic. The elution of the peptides on the DEAE column was in the order of total acidic residue content. Each of the three

Table 3. Amino acid compositions of some purified peptides from the formic acid partial hydrolysis of riPP

Data are expressed as amino acid residues/1000 total amino acid residues. Abbreviation: Pfn, peptide fraction no.

	Composition					
Amino acid	Intact riPP	Pfn... I-15	III-28	III-29	IV-31	IV-32
Lys	9.3	22	17	20	10	8.2
His	8.3	33	4.5	7.9	2.0	2.4
Arg	6.5	40	32	12	15	14
Asp	344	153	104	142	282	273
Thr	17	39	36	30	13	18
Ser	367	166	235	254	375	411
PSer	104	–	227	130	129	158
(Ser + PSer)...	(471)	(166)	(462)	(384)	(504)	(559)
Glu	61	133	68	59	35	41
Pro	14	40	37	59	17	19
Gly	43	147	64	66	29	25
Ala	10	Trace*	72	84	60	12
Val	3.2	111	16	13	9.1	5.0
Met	1.2	Trace	Trace	Trace	Trace	Trace
Ile	3.0	39	13	13	4.3	4.1
Leu	6.3	39	35	38	12	11
Tyr	0.7	25	20	28	3.1	Trace
Phe	2.6	11	14	16	3.6	Trace

* Cystine was present in trace amounts in the intact riPP, but was not detected in any of the peptides analysed.

Table 4. Results of *N*-terminal sequencing of some peptides produced by limited formic acid hydrolysis of riPP

Peptide	Sequence position...	Amino acid 1	2	3	4	5	6	7	8	9	10	11	12
I-15		Asp	Asp	Asp	Asp	-*							
IV-31		Asp	Asp	-†	Asp	-†	Asp	Asp	Asp	Asp	Asp	Asp	-*
IV-32		Asp	-†	-†	-†	Asp	Asp	-*					

* The yield of amino acids in all following cycles is very low after this point, probably indicating a loss of the peptide from the matrix during sequencing.

† These intervening very low yields are probably due to the presence of phosphoserine.

peptides produced by the limited formic acid hydrolysis, and thus far taken for sequence analysis, begins with an *N*-terminal sequence of several aspartic acid residues, indicating that, under these limited hydrolysis conditions, aspartic acid residues are most easily broken out from the peptide chain when they are bounded by other aspartic acid residues. These data, showing runs of several aspartic acid residues, clearly support the conclusion, drawn earlier from the acetic acid-hydrolysis data, that there are domains of contiguous aspartic acid residues.

Peptide I-15 is especially interesting, as it contains a collection of residues other than serine or aspartic acid that are not at all representative of the overall amino acid composition. That is, this peptide represents a single less-acidic domain, relatively rich in hydrophobic and basic residues, and especially rich in glutamic acid, glycine and valine, at the same level as aspartic acid and serine. The phosphoserine content seems very low. This peptide may be considered as a likely candidate for the construction of nucleotide probes or antibodies for screening of a cDNA rat odontoblast library.

Tryptic peptides. The void-volume peak from gel-filtration chromatography of riPP was reduced only slightly after trypsin digestion. However, as in the case of bPP, a limited set of low-molecular-mass peptides were obtained (Fig. 7). Fractions 11 and 12 in particular were examined further by reverse-phase chromatography using a TFA/acetonitrile gradient system. As shown in Fig. 8, the final retained components in peak 12 contained two major low-molecular-mass components; these fractions, designated '44' and '45', were taken for sequencing, with the results shown in Table 5. Peak 44 proved to be a 19-residue internal region from the active site of the bovine trypsin used for the cleavage. This peptide was not found in control digestions in the absence of the riPP. Peak 45, of comparable size, is clearly derived from riPP. It consists of a poly(aspartic acid) sequence and a serine/aspartic acid-rich region with a single tyrosine residue.

The next larger peptide, in fraction 11, had an *N*-terminal sequence consisting exclusively of a mixture of serine and phosphoserine residues throughout the first nine residues

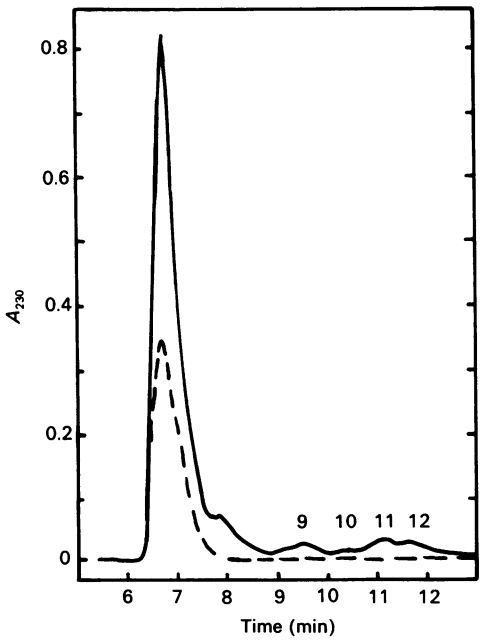

Fig. 7. Gel-filtration chromatography of DEAE-purified riPP before and after digestion with trypsin

----, Initial riPP; ——, trypsin degradation products. The smallest peptides, designated 9, 10, 11 and 12, were collected for further analysis.

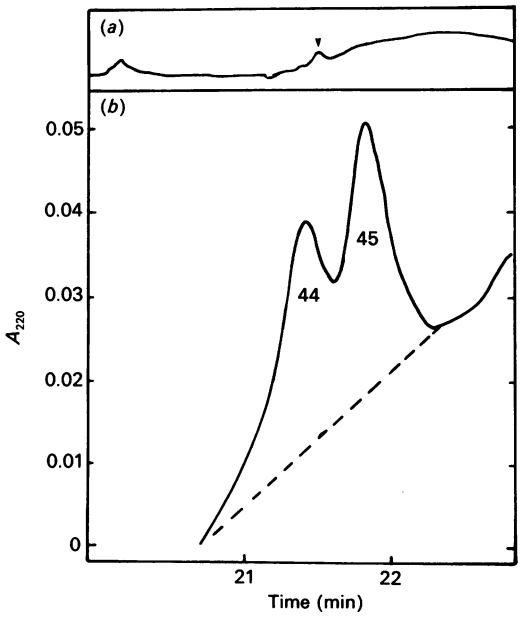

Fig. 8. Reverse-phase chromatography of tryptic peptides from peak 12, Fig. 7, from riPP

Chromatography was on Zorbax Poly-F, with a gradient of acetonitrile in 0.1 % TFA. (*a*). The entire chromatogram. The arrowhead denotes the region expanded in (*b*). Fractions designated 44 and 45 were taken for sequencing.

Table 5. Results of *N*-terminal sequencing of some peptides from the partial tryptic digestion of riPP

Peptide	Sequence position...	Amino acid 1	2	3	4	5	6	7	8	9	10	11	12	13
12–44		Ser	Ile	Val	His	Pro	Ser	Tyr	Asn	Ser	Asn	Thr	Leu	Asn
		14	15	16	17	18	19							
		Asn	Asp	Tyr	Met	Leu	Ile‡							
		1	2	3	4	5	6	7	8	9	10	11	12	13
12–45		Asp	Asp	Asp	Asp	Asp	Asp	Tyr	Ser	Asp	Ser	Asp	Ser	Ser
		14	15	16	17	18	19							
		Asp	Ser	Asp	Asp	–†								
		1	2	3	4	5	6	7	8	9	10			
11		Ser	–*	–*	Ser	–*	–*	Ser	Ser	Ser	–†			

* Probably phosphoserine.
† Yields too low to continue.
‡ End of sequence.

(Table 5). In this case the repetitive yields of the serine in positions 1, 4, 7, 8 and 9 were virtually identical, but nothing further was detected through 20 cycles. It is most probably that residues 2, 3, 5 and 6 were phosphoserine.

The *N*-terminal regions of the other tryptic peptides have not been sequenced in any reasonable fashion as yet because, as in the case of fraction 11 and the peptides generated by the formic acid partial hydrolysis, the numerous phosphoserine residues cause the sequencing efficiency to fall off very rapidly. The next approach to sequencing from the peptides must include prior dephosphorylation. This may not be very profitable, because dephosphorylation does not go to completion and degradation very frequently accompanies this reaction.

DISCUSSION

The aims of this work were twofold: to develop enough unique sequence data to permit construction of nucleotide probes so that a rat odontoblast cDNA library could be examined, and to explore the distribution of domains of distinct sequences, as suggested by preliminary work (Lechner *et al.*, 1981).

Partial acid hydrolysis

Partial acid hydrolysis of bPP using acetic acid under optimal specific aspartic acid-residue-cleavage conditions released 43% of the aspartic acid (Table 1), far in excess of all other amino acid residues. These data lead to three immediate conclusions. First, if all other released free amino acids (Table 1) came strictly from selective aspartic acid cleavage in Asp-Yaa-Asp sequences, then 22% of the free aspartic acid must have arisen from other sequences. Considering that aspartic acid comprises 38% of the residues in the intact molecule, the excess aspartic acid released must have been from sequences Waa-Asp-Xaa, including (Asp-Asp-Asp) or $(Asp)_{n \geqslant 3}$. Secondly, the particular sequence or structure clearly affects the susceptibility of the aspartic acid to cleavage under these mild-acid-hydrolysis conditions. Less than half of the aspartic acid is released. Thirdly, although aspartic acid and phosphoserine are present in the intact bPP in near-equivalent amounts, the sequence Asp-(P)Ser-Asp is present and prominent, but can account for only 13% of the phosphoserine distribution. This is a much smaller amount than previously supposed.

The somewhat milder digestion of riPP with formic acid provides an additional insight into the residue distribution question, assuming some rather strong sequence identity between riPP and bPP as suggested by their similar overall compositions and common reactivity to the same antibody (Rahima & Veis, 1988). As indicated in Fig. 5, the riPP is broken down into a variety of peptides, most of which are eluted from the DEAE column at a lower ionic strength than is the intact riPP. The least acidic fraction, peak I, yielded a pure peptide following h.p.l.c. (Fig. 6), peak I-15, which had an amino acid composition considerably enriched in all residues other than aspartic acid and serine (Table 3), compared with the intact riPP. Nevertheless, I-15 contained aspartic acid and serine as the major amino acid constituents. On the other hand, DEAE fraction IV yielded a peptide following reverse-phase chromatography, IV-32, which was markedly enriched in serine and phosphoserine relative to riPP, and depleted in aspartic acid content. The [Ser + PSer]/Asp ratios were 1.37 for riPP, 1.08 for I-15, and 2.04 for IV-32. These data, as well as that for peptides III-28 and III-29, all indicate that the aspartic acid and serine residues within the intact riPP are not uniformly distributed and that there must be blocks of aspartic acid and serine (or phosphoserine). This argument is strengthened by the *N*-terminal-sequence data in Table 4. Peptides I-15 and IV-31 clearly have $[Asp]_n$ blocks. Further, the residue preceding residue 1 in each sequence must have also been an aspartic acid residue.

The reverse argument is also true. The composition of peptide I-15 (Table 3) shows that, at least in one region, there is an appreciable concentration of the other amino acids. On the basis of the composition, the minimum size of this peptide is approx. 35 or some multiple of 35 amino acids. Within the 35-mer, 12 or 13 of these are aspartic acid and serine, the remaining 22–23 being other residues.

Finally, it seems evident that aspartic acid is most readily cleaved by mild acid hydrolysis within sequences $[Asp]_n$.

Tryptic digestion

bPP and riPP do differ from each other markedly in terms of their relative lysine content, the bPP having almost four times more than riPP. Nevertheless, the tryptic digestion patterns of the two proteins are remarkably similar. Trypsin digestion is not very effective in reducing the molecular mass of either bPP or riPP (Fig. 2).

The data of Fig. 2*a*, depicting the time course of the digestion of ^{125}I-bPP, show quite definitively that only a select few of the lysine residues within bPP are trypsin-sensitive, and that those lysine residues must be in end regions of the molecule, since the major component has its apparent molecular mass reduced by < 10000, and two very-low-molecular-mass ^{125}I-labelled bands appear. The two small peptides appear at different rates. Gel-

filtration chromatography (Fig. 3) supports these data in the sense that most of the protein is eluted at the column void volume and only a low content of a few small peptides are seen. The Stains All staining (Fig. 2*b*), which does not show up the released peptides, corroborates the fact that the resistant high-molecular-mass fraction is highly phosphorylated.

fbPP behaves in a manner similar to the mature bPP in that there is a sequential release of a few small peptides and the retention of a major high-molecular-mass fragment (Fig. 4) during tryptic digestion. The void-volume peak (Table 2) after trypsin digestion is enriched in aspartic acid and serine, and retains 78% of the initial lysine content. The small peptides of fbPP would appear to contain all of the tyrosine and phenylalanine and most of the proline, glycine and glutamic acid residues.

Fig. 2(*c*) shows the digestion of a crude preparation of riPP, that is, one with both major PP constituents, α and β (Dimuzio & Veis, 1978). Neither band is reduced appreciably in apparent size. The lower-molecular-mass Stains All-stained bands (A–E) are naturally occurring degradation products which also appear when riPP is stored for a long time; A and D are diminished upon trypsin treatment, and C and E are enhanced. Nevertheless, and in spite of its much lower content of lysine, small peptides can be detected by h.p.l.c. in the tryptic digests of highly-purified riPP from which all of the naturally-occurring degradation peptides had been removed (Figs. 7 and 8). Peptide 11 has an *N*-terminal region comprised exclusively of at least nine serine and phosphoserine residues, whereas peptide 12–45 begins with a block of six aspartic acid residues.

A domain model for phosphophoryns

The similar behaviour of bPP and riPP in the partial acid- and tryptic-hydrolysis systems, as well as their overall immunological and compositional similarities, suggest that both molecules may well have a similar structure and distribution of molecular domains as depicted in Fig. 9. This model has three essential features. First, there must be a number of blocks of $[Asp]_n$ and $[(P)Ser]_m$ distributed among aspartic acid- and serine-rich regions which also contain some $[Asp\text{-}(P)Ser\text{-}Asp]_k$ sequences as prominent sequence elements. Secondly, although every molecular region is acidic, a major portion of the non-aspartic acid and non-serine residues is located within sequences near the ends of the molecule and can be cleaved from the central region of the molecule with trypsin. As indicated in Table 2, all of the tyrosine in bPP may be in such regions. This may explain much of the controversy in the early literature, where Linde & colleagues (see, e.g., Jontell & Linde, 1983) claimed that 'pure' PP was free of tyrosine and that those showing tyrosine were working with impure preparations; in fact the reverse is correct, and their early studies were on degraded preparations. Most of the glutamic acid, proline and glycine are also in the end-region sequences. Finally, the central portion of the molecule is almost entirely aspartic acid, (phospho)serine and lysine, with the lysine being protected from tryptic digestion by the high concentration of surrounding acidic residues. The tryptic-digestion-resistant central region retains all of the Ca^{2+}-binding properties of the molecule (Stetler-Stevenson & Veis, 1987), but, as has been discussed elsewhere, it retains its affinity for binding to collagen, whereas it does not retain antigenicity to the polyclonal antibody to riPP (Tsay & Veis, 1985; Rahima *et al.*, 1988; Rahima & Veis, 1988). On the basis of this model, it should now be possible to begin an exploration of the functions of the individual domains.

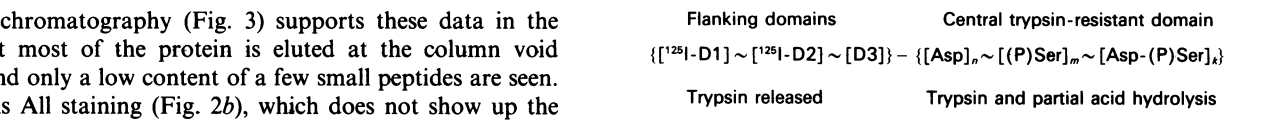

Fig. 9. A domain structure model for the PPs as applied particularly to bPP

No order is implied for the three flanking domains (D1, D2 and D3) other than they are at either end of the central trypsin-resistant domain. Likewise, no order is implied for the aspartic acid- and serine-rich regions within the central domain.

This work was supported by the National Institute of Dental Research, National Institutes of Health (grant DE-01374).

REFERENCES

Butler, W. T., Bhown, M., DiMuzio, M. T., Cothran, W. C. & Linde, A. (1983) Arch. Biochem. Biophys. **225**, 178–186
Carmichael, D. J., Chovelon, A. & Pearson, C. H. (1978) Calcif. Tissue Res. **17**, 263–271
Cunico, R. L., Mayer, A. G., Wehr, C. T. & Sheehan, T. L. (1986) BioChromatography **1**, 6–13
Dimuzio, M. T. & Veis, A. (1978) Calcif. Tissue Res. **25**, 169–178
Green, M. R., Pastewka, J. V. & Peacock, A. C. (1973) Anal. Biochem. **56**, 1625–1633
Inglis, A. S. (1983) Methods Enzymol. **91**, 324–332
Jontell, M. & Linde, A. (1983) Biochem. J. **214**, 769–776
Krippner, R. D. & Nawrot, C. F. (1977) J. Dent. Res. **56**, 873
Laemmli, U. K. (1970) Nature (London) **277**, 680–685
Lechner, J., Veis, A. & Sabsay, B. (1981) in The Chemistry and Biology of Mineralized Connective Tissues (Veis, A., ed.), pp. 395–398, Elsevier/North-Holland, New York
Lee, S. L., Veis, A. & Glonek, T. (1977) Biochemistry **16**, 2971–2979
Lee, S. L., Glonek, T. & Glimcher, M. (1983) Calcif. Tissue Int. **35**, 815–818
Maier, G. D., Lechner, J. H. & Veis, A. (1983) J. Biol. Chem. **258**, 1450–1455
Masters, P. M. (1985) Calcif. Tissue Int. **37**, 236–241
McDougall, M., Zeichner-David, M. & Slavkin, H. (1985) Biochem. J. **232**, 493–500
Nakamura, O., Gohda, E., Ozawa, M., Miyazaki, H., Murakami, I. & Daikuhara, Y. (1985) Calcif. Tissue Int. **37**, 491–500
Rahima, M. & Veis, A. (1988) Calcif. Tissue Int. **42**, 104–112
Rahima, M., Tsay, T.-G., Andujar, M. & Veis, A. (1988) J. Histochem. Cytochem. **36**, 153–157
Rusenko, K. W. (1988) Ph.D. Dissertation, Clemson University, Clemson, SC
Schultz, J., Allison, H. & Grice, M. (1962) Biochemistry **1**, 694–698
Stetler-Stevenson, W. G. & Veis, A. (1983) Biochemistry **22**, 4326–4335
Stetler-Stevenson, W. G. & Veis, A. (1986) Calcif. Tissue Int. **38**, 135–141
Stetler-Stevenson, W. G. & Veis, A. (1987) Calcif. Tissue Int. **40**, 97–102
Takagi, Y., Fujisawa, R. & Sasaki, S. (1986) Connect. Tiss. Res. **14**, 279–292
Termine, J. D., Belcourt, A. B., Miyamoto, M. S. & Conn, K. S. (1980) J. Biol. Chem. **255**, 9769–9772
Tsay, T.-G. & Veis, A. (1985) Biochemistry **24**, 6363–6369
Veis, A. & Sabsay, B. (1983) in Biomineralization and Biological Metal Accumulation (Westbroek, P. & de Jong, E. W., eds.), pp. 273–284, D. Reidel, Dordrecht
Weinstock, M. & Leblond, C. P. (1973) J. Cell Biol. **56**, 838–845
Zanetti, M., de Bernard, B., Jontell, M. & Linde, A. (1981) Eur. J. Biochem. **113**, 541–545

Received 7 September 1990/1 November 1990; accepted 27 November 1990