Identification and Characterization of the Carboxyl-terminal Region of Rat Dentin Sialoprotein*

Received for publication, July 14, 2000, and in revised form, October 18, 2000 Published, JBC Papers in Press, October 19, 2000, DOI 10.1074/jbc.M006271200

Chunlin Qin‡§, Richard G. Cook¶¶, Ralph S. Orkiszewski‖, and William T. Butler‡

From the ‡Department of Basic Sciences, The University of Texas-Houston Health Science Center, Dental Branch, Houston, Texas 77030 and ¶Department of Immunology and ‖Protein Chemistry Core Laboratory, Baylor College of Medicine, Houston, Texas 77030

Two acidic proteins, dentin sialoprotein (DSP) and dentin phosphoprotein (DPP), are present in the extracellular matrix of dentin but not in bone. These two proteins are expressed in odontoblasts and preameloblasts as a single CDNA transcript coding a large precursor protein termed dentin sialophosphoprotein (DSPP). DSPP is specifically cleaved into two unique proteins, DSP and DPP. However, the cleavage site(s) of DSPP and the mechanisms for regulating the cleavages are unknown. To identify the specific site(s) of DSPP that are cleaved when the initial translation product is converted to DSP and DPP, we performed a detailed analysis (Edman degradation and mass spectrometry) on selected tryptic peptides of a size originating from the COOH-terminal region of rat DSP. After cleavage with trypsin, the DSP fragments were separated by a two-dimensional method (size-exclusion chromatography followed by reversed phase high performance liquid chromatography). We characterized 13 peptides from various regions of DSP. The analyses showed that peptide Ile^{409}-Tyr^{421} was the major COOH-terminal fragment, ending at Tyr^{421} only 9 residues from the NH_{2} terminus of DPP. Peptide Gln^{385}-His^{406} represented a second, minor COOH-terminal peptide that terminated at His^{406}. Both of these residues are well beyond the COOH terminus predicted previously by two independent studies estimating that rat DSP contained 360–370 amino acids. Careful studies on two peptides showed that, among 9 potential casein kinase II phosphorylation sites, 2 serines were phosphorylated. We found that rat DSP was heterogeneous with respect to phosphorylation, because this same peptide sequence eluted in two discrete peaks, one with 2 phosphoserines and the other having 1. The finding that 3 lysines just preceding the COOH termini were modified by a 43-Da substituent (possibly a carbamoyl substituent) suggests that the lysines in this region were particularly susceptible to attachment of this substituent.

The dentin extracellular matrix is formed by highly specialized, postmitotic cells termed odontoblasts. These cells secrete a unique set of gene products similar to those expressed by osteoblasts in the formation of bone. The noncollagenous proteins include osteolectin, osteocalcin, osteopontin (OPN), bone sialoprotein, and dentin matrix protein 1 (Dmp-1), found in bone and dentin, and dentin phosphoprotein (DPP) and dentin sialoprotein (DSP), occurring in dentin but not in bone (1–5). Because these noncollagenous proteins are very acidic and are secreted into the extracellular matrix during the formation and mineralization of these tissues, it is generally accepted that they play key biological roles in the formation of dentin and bone (6); however, details concerning their precise functions are unknown.

Type I collagen is the most abundant organic constituent of dentin extracellular matrix, forming a fibrillar lattice for mineral deposition. DPP is the second most plentiful protein, accounting for as much as 50% of dentin noncollagenous proteins. The most unusual feature of DPP is the occurrence of large amounts of Asp and phosphoserine (7–10). Many of these residues are present in repeating sequences of (Asp-phosphoserine-phosphoserine), and (Asp-phosphoserine). Energy minimization modeling techniques (8) indicate that these repeating sequences of phosphoserine and aspartic acid assume extended backbone structures with relatively long ridges of carboxylate and phosphate groups on each side of the peptide backbone. These structures fit well with the purported function of DPP in the nucleation and modulation of hydroxyapatite crystal formation (6).

DSP, a sialic acid-rich glycoprotein first discovered in our laboratory (11, 12), accounts for 5–8% of the dentin noncollagenous proteins. This protein shares overall characteristics with other sialoproteins (OPN, bone sialoprotein, and Dmp-1) from bone and dentin, but the levels of sequence similarities are low. Nevertheless, the fact that the genes for DSP, Dmp-1, bone sialoprotein, and OPN are found at a similar chromosomal location (i.e. on human chromosome 4q21–4q23) suggests a type of ancestral relationship (13–16). DSP is a glycoprotein with 29.6% carbohydrate, including 9% sialic acid (12). The molecular mass for DSP, determined by analytical ultracentrifugation, was 52,570 Da. When the 29.6% carbohydrate content was taken into account, the molecular mass for the core protein was calculated to be 37,009 Da (12). From this molecular mass and the average residue mass, the number of amino acid residues was calculated to be 359 (12). With Edman degradation, the NH_{2}-terminal sequence was shown to be IPVPQLVPL (17). Cloning and sequence determination of a rat DSP cDNA were used to deduce the amino acid sequence (17). With the

* This work was supported by National Institutes of Health Grant DE 05092 (to W. T. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Chunlin Qin, Department of Basic Sciences, The University of Texas-Houston Health Science Center, Dental Branch, 6516 John Freeman Ave., DBB, Rm. 4.133. Houston TX 77039. Tel.: 713-500-4583; Fax: 713-500-4568; E-mail: cqin@mail.db.uth.tmc.edu.

The abbreviations used are: OPN, osteopontin; Dmp-1, dentin matrix protein 1; DPP, dentin phosphoprotein; DSP, dentin sialoprotein; DSPP, dentin sialophosphoprotein; MS, mass spectrometry; HPLC, high performance liquid chromatography; GdmCl, guanidium-HCl; PTH, phenylthiohydantoin.

The dentin extracellular matrix is formed by highly specialized, postmitotic cells termed odontoblasts. These cells secrete a unique set of gene products similar to those expressed by osteoblasts in the formation of bone. The noncollagenous proteins include osteolectin, osteocalcin, osteopontin (OPN), bone sialoprotein, and dentin matrix protein 1 (Dmp-1), found in bone and dentin, and dentin phosphoprotein (DPP) and dentin sialoprotein (DSP), occurring in dentin but not in bone (1–5). Because these noncollagenous proteins are very acidic and are secreted into the extracellular matrix during the formation and mineralization of these tissues, it is generally accepted that they play key biological roles in the formation of dentin and bone (6); however, details concerning their precise functions are unknown.

Type I collagen is the most abundant organic constituent of dentin extracellular matrix, forming a fibrillar lattice for mineral deposition. DPP is the second most plentiful protein, accounting for as much as 50% of dentin noncollagenous proteins. The most unusual feature of DPP is the occurrence of large amounts of Asp and phosphoserine (7–10). Many of these residues are present in repeating sequences of (Asp-phosphoserine-phosphoserine), and (Asp-phosphoserine). Energy minimization modeling techniques (8) indicate that these repeating sequences of phosphoserine and aspartic acid assume extended backbone structures with relatively long ridges of carboxylate and phosphate groups on each side of the peptide backbone. These structures fit well with the purported function of DPP in the nucleation and modulation of hydroxyapatite crystal formation (6).

DSP, a sialic acid-rich glycoprotein first discovered in our laboratory (11, 12), accounts for 5–8% of the dentin noncollagenous proteins. This protein shares overall characteristics with other sialoproteins (OPN, bone sialoprotein, and Dmp-1) from bone and dentin, but the levels of sequence similarities are low. Nevertheless, the fact that the genes for DSP, Dmp-1, bone sialoprotein, and OPN are found at a similar chromosomal location (i.e. on human chromosome 4q21–4q23) suggests a type of ancestral relationship (13–16). DSP is a glycoprotein with 29.6% carbohydrate, including 9% sialic acid (12). The molecular mass for DSP, determined by analytical ultracentrifugation, was 52,570 Da. When the 29.6% carbohydrate content was taken into account, the molecular mass for the core protein was calculated to be 37,009 Da (12). From this molecular mass and the average residue mass, the number of amino acid residues was calculated to be 359 (12). With Edman degradation, the NH_{2}-terminal sequence was shown to be IPVPQLVPL (17). Cloning and sequence determination of a rat DSP cDNA were used to deduce the amino acid sequence (17). With the

§ To whom correspondence should be addressed: Chunlin Qin, Department of Basic Sciences, The University of Texas-Houston Health Science Center, Dental Branch, 6516 John Freeman Ave., DBB, Rm. 4.133. Houston TX 77039. Tel.: 713-500-4583; Fax: 713-500-4568; E-mail: cqin@mail.db.uth.tmc.edu.

The abbreviations used are: OPN, osteopontin; Dmp-1, dentin matrix protein 1; DPP, dentin phosphoprotein; DSP, dentin sialoprotein; DSPP, dentin sialophosphoprotein; MS, mass spectrometry; HPLC, high performance liquid chromatography; GdmCl, guanidium-HCl; PTH, phenylthiohydantoin.
known NH₂-terminus, this analysis indicated that DSP contained 366 amino acids. The molecular mass calculated from the predicted 366 residues, when added to that for the carbohydrate mass, was ~53 kDa, identical to that determined by analytical ultracentrifugation (12, 17). The calculated amino acid composition for the predicted sequence was identical to that determined for the protein (12, 17). Thus, the length of DSP was determined by two independent methods, and the conclusions in each case of ~360 amino acids were in close agreement. It was shown later that a single-base mistake in the original rat DSP cDNA sequencing created a frame shift leading to an early stop codon following the coding sequence for residue 366 (10, 18). The corrected sequence creates an open reading frame and a sequence containing a 5′-DSP sequence and a 3′-DPP sequence (see below).

MacDougall et al. (10) and Feng et al. (13) discovered that the nucleotide sequences for mouse DSP and DPP reside on the same gene coding for a single cDNA transcript. This transcript would result in a translational product protein termed dentin sialophosphoprotein (DSPP) that would be specifically cleaved into two proteins, DSP and DPP, with unique physical and chemical characteristics. Analysis of the full-length cDNA revealed a 934-amino acid open reading frame corresponding to the sequence of rat DSP.

In situ hybridization with riboprobes specifically de-signed for DSP, DPP, and the “linker” (dsp-dpp) regions of the mouse DSPP revealed a 17-amino acid signal peptide. The signal peptide and the deduced sequence of the NH₂-terminal portion (amino acids 1–360 numbered from the NH₂-terminus of the secreted protein) of the mouse DSPP were 78% homolo-gous to the sequence of rat DSP.

The NH₂-terminal sequence of rat DPP determined by Edman degradation was Asp-Asp-Pro-Asn for rat HP2 (a highly phosphorylated DPP; Ref. 19), and in human it was shown to be Asp-Asp-Pro (20). The beginning of the DPP portion of mouse DSPP (10) was established from the rat NH₂-terminal sequence Asp-Asp-Pro-Asn (see Fig. 1). Beginning with the mouse Asp-Asp-Pro sequence, mouse DPP was shown to contain 483 amino acids. The amino acid sequence of this portion was highly similar to that of the rat DPP cDNA reported by Ritchie and Wang (7) and by George et al. (9). Thus, the 5′ region of this mouse DSPP clone is identical to that of rat DSP cDNA (17), and the 3′ region is the same as that of the rat DPP (7, 9). In situ hybridization with riboprobes specifically designed for DSP, DPP, and the “linker” (dsp-dpp) regions of the mouse DSPP nucleotide sequence showed a strict coexpression of the three probes in odontoblasts and preameloblasts (21). The fact that the dsp-dpp linker sequence remains identically coexpressed with DSP and DPP sequences suggests that formation of DPP and DSP molecules takes place after translation of a single transcript (21). Recently Gu et al. (22) cloned a human DSPP gene with genomic organization very similar to that of mouse DSPP. All these data support the hypothesis that DSP and DPP represent specific cleavage products of a large precursor protein.

However, the cleavage site(s) for the precursor protein DSPP and the mechanism for conversion to DSP and DPP are unknown. The NH₂-terminal sequences for rat (7, 9, 19) and mouse DPP (10, 13) establish that the cleavage sites precede (i.e. are NH₂-terminal to) residue 431 in rat or 435 in mouse (see Fig. 1). As stated earlier, in two independent studies, the rat DSP was estimated to contain ~360 amino acids, and the cleavage site of the DSPP precursor was predicted to be very near residue 366 (6, 12, 17, 18). On the basis of this conclusion and the site of the NH₂-terminal residue of DPP, Asp⁴³¹ (in rat), it was postulated that an intervening sequence of ~64 amino acids between DSP and DPP was cleaved out before or after secretion of DSPP (6).

Our primary objective in this study was to identify the specific sites of DSPP that are cleaved when the initial translation product is converted to DSP and DPP. DSP tryptic peptides were purified by a two-dimensional method, size-exclusion chromatography followed by reversed phase high performance liquid chromatography (HPLC). With the former method, we selected peptides that were 10–25 residues in length, a size expected of most of the tryptic peptides of rat DSPP between residues 366 and 431 (see Fig. 1). Fortuitously, two of these peptides were COOH-terminal ends, as indicated by COOH-terminal amino acids other than lysine or arginine. These two tryptic peptides, originating from residues Gln⁴³⁸-His⁴⁶⁰ and Ile⁴⁶⁰-Tyr⁴⁷¹ (as shown by cDNA deduced sequences), are those representing minor and major COOH-terminal ends, respectively. We also performed a detailed analysis on other tryptic peptides. In addition to sequences from the NH₂-terminal and central regions of DSP, our investigation led to a complete sequence analysis of the COOH-terminal region beyond residue 366 and to the identification of two phosphoserines.

EXPERIMENTAL PROCEDURES

Isolation of DSP—DSP was isolated from rat dentin by standard procedures as described (11, 12). Briefly, rat incisor pieces were extracted with 0.5 m EDTA in 4 m guanidine-HCl (GdmCl; Acros Organ-ics) containing protease inhibitors. Next, EDTA extracts were subjected to gel chromatography on Sephacryl S-200 in GdmCl. Using this procedure, the high molecular weight protein fraction (ES1) was separated from osteocalcin, which eluted in an included volume (ES2). ES1 was next chromatographed on diethylaminoethyl-Sephadex, eluted with a linear gradient formed from 750 ml of each starting buffer (50 mm Tris-HCl, 6 mm urea, pH 7.2) and 50 mm Tris-HCl, 6 mm urea, pH 7.2, containing 0.7 m NaCl. DSP eluted in a position in the gradient correspond-ing to ~0.2 m NaCl; this peak, referred to as fraction B (11, 12), was further purified by gel filtration on a Biogel A1.5m in 4 m GdmCl, Tris-HCl, pH 7.2. DSP eluted in one major peak, separated from lower molecular weight proteins (see Fig. 1 in Ref. 12). The purity of DSP was assessed by 5–15% SDS-polyacrylamide gel electrophoresis. The samples of DSP displayed one major protein band at about M, 95,000. In addition, two minor bands at M, 75,000 and 65,000 were often observed (see Ref. 12). Each of these minor DSP bands reacted with DSP antibodies on Western immunoblots (data not shown), and we believe that they represent fragments of DSP; their presence did not interfere with the studies reported here. Note that the presence of either GdmCl or urea in each step of this preparative procedure prevented artifactual degradation of DSP.

Trypsin Digestion—DSP was digested overnight at 37 °C with trypsin (Roche Molecular Biochemicals) at an enzyme-substrate ratio of 1:50 in 50 mm Tris-HCl, pH 8.0. The final concentration of trypsin was 20 ng/μl. In the absence of trypsin, the chromophore 1,1- tosylamido-2-phenylethyl chloromethyl ketone (Roche Molecular Biochemicals) was added to a final concentration of 100 μg/ml to inhibit possible chymotrypsin activity.

Peptide Purification—After trypsin digestion of DSP, HPLC separa-tion of peptides gave rise to a large number of overlapping peptide peaks that would be less useful in our structure determination. There-fore, peptides were separated using a two-dimensional approach to produce a series of pure peptides that could be unambiguously charac-terized. Another objective was to obtain peptides with sizes that were likely to emanate from the COOH-terminal region of DSP. Inspection of the sequence (Fig. 1) showed that trypsin cleavage at lysyl and arginyl bonds would result in several peptides of 11 to 22 amino acids from that area. This approach did not ensure that a COOH-terminal peptide would be in this size range. In the first phase, peptides were separated according to size on Superdex 75 HR 10/30 (Amersham Pharmacia Biotech) equilibrated and eluted at 0.5 ml/min in 4 m GdmCl (SigmaUltra-trix, pH 6.0). Using fast protein liquid chromatography, the eluant was monitored at 280 nm. DSP peptides separated into six peaks, and each peak was divided into three or four subfractions. Subfractions con-taining peptides with desired sizes were selected for a second phase HPLC separation. In the second phase, each selected subfraction was subjected to HPLC with a 2.1 × 250-mm C18 reversed phase column (Vydac). For optimal separations, two separate HPLC gradient systems were used, depending on the types of peptides to be purified. The first system used a 3–35% acetonitrile linear gradient in 0.1% trifluoro-aetic acid over 100 min at a flow rate of 300 μl/min; this gradient was...
used to separate most of the peptides. The second elution approach was to initially wash with water (0% acetonitrile) for 30 min before starting the 0–30% acetonitrile gradient (0.1% trifluoroacetic acid, 300 μl/min over 100 min). This latter gradient system was used to separate hydrophilic peptides because of difficulties in achieving purity; using the first approach, these peptides emerged from HPLC in short times and were of low purity. For all HPLC separations, the eluant was monitored at 218 nm, and peaks were collected by hand.

Sequence Analysis and Determination of Mass—Most purified peptides of desired sizes from HPLC were first sequenced by Edman degradation on Applied Biosystems ABI 473A and 477A sequencers using standard techniques. Next, they were further analyzed for molecular mass and sequence by mass spectrometry. For mass spectrometry, the samples were dried under vacuum and redissolved in aqueous solution containing 50% methanol and 1% formic acid. Aliquots of the solutions were deposited in metal coated glass nanoelectrospray capillary tubes for analysis on a PE Sciex API 3000 triple quadrupole mass spectrometer (Concord) equipped with a Protana nanoelectrospray source (Odense). The samples were analyzed by one or more of the following regimens. For determination of peptide molecular mass, full scan mass spectra were recorded using Q1 as the resolving analyzer in positive ion mode. Product ion spectra were acquired using Q1 to transmit the precursor ion of interest to the radio frequency-only collision cell Q2 under collisionally activated decomposition conditions. Nitrogen was the collision gas, and collision energies in the range of 20–50 eV were employed to achieve fragmentation. Phosphorylated peptides were detected by scanning Q1 for the precursor ion of interest to the radio frequency-only collision cell Q2. Product ions of desired sizes from HPLC were first sequenced by Edman degradation and mass spectrometry. For confirmation of the structures of most peptides, we compared the theoretical molecular mass with that determined by MS. Fig. 3 shows the data for the 13 peptides, which form the basis for our conclusions in this publication.

Identification of the COOH-terminal Peptides—The identification of COOH-terminal peptides, after tryptic digestion, was made by searching for peptides with a COOH-terminal amino acid other than lysine or arginine. Although we realized that such a peptide might terminate at any point and might not occur in the selected size range, we actually identified two COOH-terminal peptides, containing 13 and 22 amino acids. One of these peptides was peptide 13 (Fig. 3). Tandem MS indicated that it had 22 amino acids, ending at His406 and not at Arg408, as we expected from trypsin cleavage. The other peptide was peptide 12 (Fig. 3), which was longer than expected from the predicted sequence (1325.4 Da) agreed with that determined by MS. Fig. 3 shows the data for the 13 peptides, which form the basis for our conclusions in this publication.

Two-dimensional Separation of DSP Peptides—Direct HPLC analysis of peptides generated by digestion of DSP with trypsin resulted in poor separations (results not shown) attributable to the high degree of heterogeneity; using only this separation technique would make sequencing of most peptides impossible. To improve the separation, the trypsin digest of DSP was first applied to a Superdex 75 gel filtration column to separate peptides according to size. DSP tryptic peptides were separated into six major peaks on Superdex 75 (results not shown). Because there were still too many peptides in each major peak to obtain ideal purification and analysis, we subdivided each peak into one to three or four subfractions. Totally, we obtained 22 subfractions, some of which were selected for the next dimension: HPLC separation (Fig. 2). This two-dimensional separation technique enabled us to obtain a number of pure peptides suitable for sequencing by Edman degradation and mass spectrometry (MS).

Structures of Peptides from DSP—In our efforts to obtain peptides from the COOH-terminal region of DSP, we used the predicted cDNA sequence (Fig. 1) as a guide. We reasoned that, after treatment of DSP with trypsin, we should obtain peptides comprising 10–25 amino acids from residue 386 to the beginning of DPP (residue 431). To be more specific, we expected tryptic peptides of 22, 11, 13, 4, and 12 residues from the COOH-terminal region (see Fig. 1). From the trypsin digest we selected peptides of 10–25 amino acids by gel filtration and purified them by HPLC. In fact, we characterized 13 peptides arising not only from the COOH terminus but also from the NH2 terminus and the central regions of DSP (Figs. 1 and 3). These peptides, accounting for 167 amino acids, were sequenced by a combination of Edman microsequencing and MS peptide sequencing. For confirmation of the structures of most peptides, we compared the theoretical molecular mass with that determined by MS. Fig. 3 shows the data for the 13 peptides, which form the basis for our conclusions in this publication.

RESULTS

Rat DSPP cDNA Sequence—Fig. 1 shows the cDNA deduced rat DSPP sequence of the region from the NH2 terminus of secreted DSP to the established NH2 terminus of DPP. Because a full-length rat DSPP cDNA has never been reported, we based this cDNA sequence on the published data from three studies. The first two are by Ritchie et al. (17) and Ritchie and Wang (7), who reported the cDNA sequence for rat DSPP and DPP in two independent papers. The third reference is by George et al. (9) who published a cDNA sequence for rDmp3, representing the same region as the mouse DSPP. In formulating this sequence, we also referred to the mouse DSPP sequence reported by MacDougall et al. (10) for comparison. Amino acids are numbered from the NH2 terminus of secreted DSP, excluding the 17-amino acid signal peptide.

Two-dimensional Separation of DSP Peptides—Direct HPLC analysis of peptides generated by digestion of DSP with trypsin was followed by Edman degradation and mass spectrometry (MS). Most purified peptides were first sequenced by Edman degradation on Applied Biosystems ABI 473A and 477A sequencers using standard techniques. Next, they were further analyzed for molecular mass and sequence by mass spectrometry. For mass spectrometry, the samples were dried under vacuum and redissolved in aqueous solution containing 50% methanol and 1% formic acid. Aliquots of the solutions were deposited in metal coated glass nanoelectrospray capillary tubes for analysis on a PE Sciex API 3000 triple quadrupole mass spectrometer (Concord) equipped with a Protana nanoelectrospray source (Odense). The samples were analyzed by one or more of the following regimens. For determination of peptide molecular mass, full scan mass spectra were recorded using Q1 as the resolving analyzer in positive ion mode. Product ion spectra were acquired using Q1 to transmit the precursor ion of interest to the radio frequency-only collision cell Q2 under collisionally activated decomposition conditions. Nitrogen was the collision gas, and collision energies in the range of 20–50 eV were employed to achieve fragmentation. Phosphorylated peptides were detected by scanning Q1 for the precursor ion of interest to the radio frequency-only collision cell Q2. Product ions of desired sizes from HPLC were first sequenced by Edman degradation and mass spectrometry. For confirmation of the structures of most peptides, we compared the theoretical molecular mass with that determined by MS. Fig. 3 shows the data for the 13 peptides, which form the basis for our conclusions in this publication.

Identification of the COOH-terminal Peptides—The identification of COOH-terminal peptides, after tryptic digestion, was made by searching for peptides with a COOH-terminal amino acid other than lysine or arginine. Although we realized that such a peptide might terminate at any point and might not occur in the selected size range, we actually identified two COOH-terminal peptides, containing 13 and 22 amino acids. One of these peptides was peptide 13 (Fig. 3). Tandem MS indicated that it had 22 amino acids, ending at His406 and not at Arg408, as we expected from trypsin cleavage. The other peptide was peptide 12 (Fig. 3), which was longer than expected from the predicted sequence (1325.4 Da) agreed with that determined by MS. Fig. 3 shows the data for the 13 peptides, which form the basis for our conclusions in this publication.

Two-dimensional Separation of DSP Peptides—Direct HPLC analysis of peptides generated by digestion of DSP with trypsin resulted in poor separations (results not shown) attributable to the high degree of heterogeneity; using only this separation technique would make sequencing of most peptides impossible. To improve the separation, the trypsin digest of DSP was first applied to a Superdex 75 gel filtration column to separate peptides according to size. DSP tryptic peptides were separated into six major peaks on Superdex 75 (results not shown). Because there were still too many peptides in each major peak to obtain ideal purification and analysis, we subdivided each peak into one to three or four subfractions. Totally, we obtained 22 subfractions, some of which were selected for the next dimension: HPLC separation (Fig. 2). This two-dimensional separation technique enabled us to obtain a number of pure peptides suitable for sequencing by Edman degradation and mass spectrometry (MS).

Structures of Peptides from DSP—In our efforts to obtain peptides from the COOH-terminal region of DSP, we used the predicted cDNA sequence (Fig. 1) as a guide. We reasoned that, after treatment of DSP with trypsin, we should obtain peptides comprising 10–25 amino acids from residue 386 to the beginning of DPP (residue 431). To be more specific, we expected tryptic peptides of 22, 11, 13, 4, and 12 residues from the COOH-terminal region (see Fig. 1). From the trypsin digest we selected peptides of 10–25 amino acids by gel filtration and purified them by HPLC. In fact, we characterized 13 peptides arising not only from the COOH terminus but also from the NH2 terminus and the central regions of DSP (Figs. 1 and 3). These peptides, accounting for 167 amino acids, were sequenced by a combination of Edman microsequencing and MS peptide sequencing. For confirmation of the structures of most peptides, we compared the theoretical molecular mass with that determined by MS. Fig. 3 shows the data for the 13 peptides, which form the basis for our conclusions in this publication.

Identification of the COOH-terminal Peptides—The identification of COOH-terminal peptides, after tryptic digestion, was made by searching for peptides with a COOH-terminal amino acid other than lysine or arginine. Although we realized that such a peptide might terminate at any point and might not occur in the selected size range, we actually identified two COOH-terminal peptides, containing 13 and 22 amino acids. One of these peptides was peptide 13 (Fig. 3). This peptide, Ile409-Tyr421 was first partially sequenced by Edman degradation and then fully sequenced by tandem MS. The theoretical molecular mass (1325.4 Da) agreed with that determined by MS. The yield of peptide Ile409-Tyr421 was high, as indicated by peak areas and by the phenylthiodyantoin (PTH)-derivative yield during Edman degradation. The peak area ratio between peptide Ile409-Tyr421 and another tyrosine-containing peptide, Glu53-Arg62 originating from the NH2-terminal region, is 0.9 (Fig. 2, compare peptide 13 with peptide 2).

A second COOH-terminal peptide, Glu389-His106 (Fig. 3, peptide 12), was longer than expected from the predicted sequence if all lysines were fully cleaved. Because peptide 12 contained a modification of Lys8397, the lysyl bond was uncleaved, and we obtained a fragment 9 residues longer than expected. Edman microsequencing of peptide 12 identified 20 amino acids (Fig. 3). Tandem MS indicated that it had 22 amino acids, ending at His106 and not at Arg408, as we expected from trypsin cleavage. The yield of peptide 12 was considerably less than that of peptide 13, as indicated by peak areas and by the PTH-derivative yield during Edman degradation. From these observations we estimate that the ratio of peptide 13 to peptide 12 was ~3–7:1.

Search for Other COOH-terminal Peptides—It is possible that a portion of DSP ends at residue 430, representing only...
fraction 13 (containing peptides with >22 amino acids), most abundant in subfraction 14 (with peptides of 17–25 amino acids), and almost absent in subfraction 15 (containing peptides with <22 amino acids). Because of its elution profile, we predicted that peptide 8 contains 22–25 residues. Edman degradation and MS confirmed that peptide 8 was Gly8270-Arg281 in which Ser281 is phosphorylated, making this peptide equivalent to 23 amino acids in size. 

Using these approaches with respect to size and hydrophilicity, we sequenced eight candidate peptides with a hydrophilic nature ranging in size from 10 to 25 amino acids. However, we did not detect any peptides that extend beyond Tyr421. Therefore, we concluded that the major COOH-terminal amino acid is Tyr422. Additionally, we did select for two phosphorylated fragments (Fig. 3, peptides 7 and 8) containing 22 amino acids; as predicted the hydrophilic nature led to early elution from HPLC. Sequence and MS data confirmed the fact that they were highly negatively charged.

### Posttranslational Modifications
We found two types of posttranslational modifications, phosphorylated serines and modified lysines. Two phosphoserines were present in the amino acid sequence Gly8270-Arg281 (Fig. 3, peptides 7 and 8). Phosphoserines were detected by a combination of Edman degradation and MS. Edman degradation of peptides containing phosphorylated serines resulted in gaps or disproportionately low yields of the PTH-Ser residues. With tandem MS, the phosphorylated serines showed a molecular mass of 167 Da (serine plus one phosphate). Additionally, the molecular mass of the peptides containing one or two phosphates increased in mass by 80 or 160 Da, respectively compared with the nominal mass of the cDNA deduced sequence. The sequence Gly8270-Arg281 showed heterogeneity with respect to phosphorylation. In peptide 7, both Ser281 and Ser288 were phosphorylated, as shown by MS sequencing and molecular mass (Fig. 3), whereas in peptide 8 only Ser281 was phosphorylated. The phosphorylation of Ser281 resulted in peptides longer than expected, presumably because the phosphate moiety converted the adjacent Lys384-Gln385 bond to a resistant bond.

Totally, we identified 12 lysines, and 3 of them located in the region directly preceding the COOH termini were modified with a substituent showing a molecular mass of 43 Da. The modification of Lys8284 resulted in two unexpected peptides displaying the same amino acid sequence corresponding to Asn831-Lys837 (Fig. 3, peptide 10 and 11). In these two peptides, the Lys8284-Glu8285 bond was uncleaved. In peptide 10, Lys8284 was modified by the 43 Da substituent, whereas in peptide 11, both Lys8284 and Lys8290 were modified in a similar manner. The modification of Lys8284 by the 43 Da substituent resulted in another unexpected peptide, Glu8285-His8286, in

---

*Fig. 3.* Sequence analysis of peptides isolated from DSP after treatment with trypsin. a, peptides we have sequenced, numbered (P4) starting from the NH2-terminal region (see Fig. 1). b, peptide sequence deduced from rat DSPP cDNA (cDNAseq; see Fig. 1). c, theoretical molecular mass (TMM; Da) of peptide sequence deduced from cDNA. d, peptide sequence determined by Edman degradation (EDSeq). e, Molecular mass determined by MS (MSMM); f, peptide sequence obtained by MS (MSMS). g, n.d., not determined. h, D differs from cDNA-deduced amino acid N in peptide 2. i, S represents phosphoserine. k, s indicates a low yield of PTH-Ser. l, f indicates an unidentified PTH-derivative after Edman degradation. m, K indicates that a lysine is modified by a 43 Da substituent.

---

C. Qin, R. G. Cook, R. S. Orksiszewski, and W. T. Butler, unpublished results.
which the Lys397-Ser398 bond could not be cleaved by trypsin. Note that peptides Asn381-Lys397 and Gln385-His406 had an overlap of 13 amino acids. After Edman degradation, the elution time for the PTH-derivative of this substituted lysine was similar to that for a succinylated lysine; however, tandem mass spectrometric analysis showed a mass of only 43 Da, differing from the 83-Da mass of a succinyl substituent. This mass of 43 Da fits a carbamoyl group, a substance resulting from reaction of the ε-amino group with cyanate in urea. At present we have no data to directly identify the 43-Da substituent.

**DISCUSSION**

The two tooth-specific proteins, DPP and DSP, isolated from dentin as distinct proteins with unique physical and chemical characteristics, are considered important in dentinogenesis (1, 2, 6). DPP, extremely rich in aspartic acid and phosphoserine, is believed to play key roles in the nucleation of hydroxyapatite onto dentin matrix collagen and the subsequent growth of the hydroxyapatite crystals (5, 6, 8), whereas the function of DSP is unknown. It is now well accepted that DPP and DSP are encoded by a single mRNA transcript and that the initial, larger protein, DSPP, contains sequences for both DSP and DPP (10, 13, 21, 22). The occurrence of one gene transcribing a single mRNA encoding both DSP and DPP must indicate that the primary translation product DSPP, giving rise to the individual proteins, is very heterogeneous with respect to phosphorylation. Taken together, we conclude that peptide Ile409-Tyr421 is not an artifactual product resulting from chymotrypsin contaminating the trypsin preparation. According to the manufacturer’s data sheet, the highly pure, sequencing grade trypsin used in the present study is free of chymotrypsin, a protease that preferentially hydrolyzes peptide bonds at the COOH terminus of Trp, Tyr, and Phe. Nevertheless, to avoid any problems arising from contamination by this enzyme, we added the chymotrypsin inhibitor 1-tosylamido-2-phenylethyl chloromethyl ketone to our trypsin digestions at a concentration of 100 μg/ml, five times as high as that of trypsin (20 μg/ml). Another finding that refutes this possibility is the presence of peptide Gln53-Arg62 (QVHSDGGYER) in which Tyr60 is not cleaved, suggesting again that the presence of chymotrypsin is highly improbable. Together, we conclude that peptide Ile409-Tyr421 is not an artifactual product resulting from chymotrypsin contaminating the trypsin, but rather that Tyr421 is one of the COOH termini of DSP.

Fig. 4 shows the amino acid sequence alignment of the region representing the DSP COOH-terminal portion and DPP NH2-terminal portion of rat and human DSPP. The cDNA sequence data for human DSPP are from Gu et al. (22). The rat DSP sequence is based on the studies by Ritchie et al. (17), Ritchie and Wang (7), and George et al. (9). It is worth noting that both Tyr421 and His406 are conserved between rat and human DSP. It is also interesting to note that the flanking region of Tyr421 shows a high level of conservation between rat and human.

Rat DSP has nine potential casein kinase II and four potential casein kinase I phosphorylation sites (17). Two phosphoserines were identified in this study. Rat DSP is heterogeneous with respect to phosphorylation. Peptide Gly370-Arg381 was eluted in two separate peaks demonstrating the same amino acid sequence but differing in the number of phosphates (Fig. 3, peptides 7 and 8). Two phosphoserines were identified in peptide 7 in which both Ser275 and Ser281 were phosphorylated, whereas in peptide 8, only Ser281 was phosphorylated. Phosphorylation of Ser281 resulted in an unexpected peptide, Gly370-Arg381, that was not cleaved at the COOH terminus of Lys282 by trypsin; this resistance to cleavage is undoubtedly attributable to the juxtaposed phosphoserine. Our previous studies have revealed that rat bone OPN, another sialic acid-rich protein, is very heterogeneous with respect to phosphorylation.
(23). We postulate that the other two sialic acid-rich proteins abundant in the mineralized tissue, bone sialoprotein and Dmp-1, will also be heterogeneously posttranslational modifications.

Some peptides that we purified contained lysyl bonds that were uncleaved by trypsin. Peptides 10 and 11 (Asp\textsuperscript{381}-Lys\textsuperscript{397}) and peptide 12 (Gln\textsuperscript{385}-His\textsuperscript{406}) were longer than expected because of lysine modification, which resulted in the refractory nature of the lysyl bonds. Peptide Asn\textsuperscript{381}-Lys\textsuperscript{397} was also eluted in two discrete peaks (Fig. 3, peptides 10 and 11). In peptide 10, Lys\textsuperscript{384} was modified by a substituent with a mass of 43 Da, whereas in peptide 11, both Lys\textsuperscript{384} and Lys\textsuperscript{390} were modified by the 43-Da substituents. The modification of Lys\textsuperscript{384} resulted in lack of cleavage of peptide Asn\textsuperscript{381}-Lys\textsuperscript{397} at Lys\textsuperscript{384} by trypsin. The third 43-Da substituent was found in peptide Gln\textsuperscript{385}-His\textsuperscript{406} (Fig. 3) in which the modified Lys\textsuperscript{397} could not be cleaved by trypsin. It is worth noting that all three lysines modified in the same manner are found in a region just preceding the COOH termini. Totally, we have identified 12 lysines, but only the 3 lysines directly preceding the COOH termini are modified. If these modifications occur within the cells synthesizing DSPP (i.e. they are biological), they may play key roles in signaling for the cleavage of DSPP precursor. On the other hand, the mass of the 43-Da substituent corresponds to that of a carbamyl moiety, a substance that results from reaction of amino groups with cyanates, formed from urea. Thus, these lysine modifications are likely to be artifacts, and their formation is probably related to the spatial structure of DSP, making the lysines in this region more accessible to modifications.

As stated earlier, DSP and DPP encoded by a single gene, DSPP, are found in dentin extracellular matrix as distinct proteins. Thus, the initial translation product must be proteolytically processed in a manner that has not been elucidated. The data presented here, along with the sequences deduced from cDNA, indicate that one of the major bonds proteolytically cleaved is Tyr\textsuperscript{421}-Asp\textsuperscript{422}, giving rise to the principal COOH terminus of DSP. A second major cleavage site is Gly\textsuperscript{430}-Asp\textsuperscript{431}, resulting in the NH\textsubscript{2} terminus of DPP. A third minor area of proteolytic hydrolysis appears to be His\textsuperscript{406}-Ser\textsuperscript{407}. Although the proteinase(s) catalyzing these scissions is unknown at this time, we speculate that tissue-specific enzymes, designed to activate DSPP by cleaving it into DSP and DPP, are involved. One candidate is a tooth-specific proteinase, enamelysin (matrix metalloproteinase 20), which is expressed by odontoblasts and ameloblasts (24, 25). Experiments using in situ hybridization and immunohistochemistry show that enamelysin transcripts are expressed before the onset of mineralization in sites where DSPP and ameloblastin translation products could be immunodetected (25). Enamelysin may be involved in cleaving protein substrates, including DSPP and ameloblastin, derived from odontoblasts and young ameloblasts, converting them from inactive precursors into their biologically active forms (25). Recently, it was shown that recombinant bovine enamelysin (recombinant matrix metalloproteinase 20) cleaved tyrosine-rich bovine amelogenin peptide at a site between Trp and Leu and leucine-rich bovine amelogenin peptide between Pro and Ala (26). It was also shown that recombinant porcine enamelysin cleaves recombinant porcine amelogenin at virtually all of the possible sites that have previously been described (27, 28); the authors concluded that the substrate specificity of enamelysin is broad, and a consensus target sequence cannot be defined. Thus, enamelysin appears to have broad enough specificity to suggest that it could catalyze the proteolytic cleavages necessary to convert DSPP to DSP and DPP. It is likely that the spatial structure rather than the primary amino acid sequence determines the cleavage sites of DSPP.

We postulate that the conserved sequences around Tyr\textsuperscript{421}-Asp\textsuperscript{422} and Gly\textsuperscript{430}-Asp\textsuperscript{431} may form a structure readily exposed and susceptible to the proteinase(s) involved in this conversion. We envision that this area is open and accessible to the proteinase(s) involved in this activity. Clearly, definite answers to the questions concerning mechanisms involved in the proteolytic processing of DSPP require a complete analysis of the three-dimensional structures of DSP, DPP, and DPP.

Acknowledgment—We thank Jan C. Brunn for excellent technical assistance.

REFERENCES

1. Butler, W. T. (1995) Connect. Tissue Res. 33, 59–65
2. Butler, W. T., and Ritchie, H. H. (1995) Int. J. Dev. Biol. 39, 169–179
3. Linde, A., and Goldberg, M. (1993) Crit. Rev. Oral Biol. Med. 4, 679–728
4. Fujisawa, R., Butler, W. T., Brunn, J. C., Zhou, H. Y., and Kuboki, Y. (1993) J. Dent. Res. 72, 1222–1226
5. Veis, A. (1993) J. Bone Miner. Res. 8, 5493–5497
6. Butler, W. T. (1998) Eur. J. Oral Sci. 106, 204–210
7. Ritchie, H. H., and Wang, L. H. (1996) J. Biol. Chem. 271, 21695–21698
8. George, A., Bannon, L., Sabasy, B., Dillon, J. W., Malone, J., Veis, A., Jenkins, N. A., Gilbert, D. J., and Copeland, N. G. (1996) J. Biol. Chem. 271, 32869–32873
9. George, A., Srivivasan, R., Thotakurn, S., and Veis, A. (1998) Eur. J. Oral Sci. 106, 221–226
10. MacDougall, M., Simmons, D., Luan, X., Nydegger, J., Feng, J., and Gu, T. T. (1997) J. Biol. Chem. 272, 835–842
11. Butler, W. T. (1987) Methods Enzymol. 255, 290–303
12. Butler, W. T., Bhown, M., D’Souza, R. N., Farah-Carson, M. C., Happonen, R. P., Schroenersher, R. E., Seyer, J. M., Somerman, M. J., Foster, R. A., Tomana, M., and VanDijk, S. (1992) Matrix 12, 343–351
13. Feng, J. Q., Luan, X., Wallace, J., Jing, D., Oshihama, T., Kulkarni, A. B., D’Souza, R. N., Kokaz, C. A., and MacDougall, M. (1998) J. Biol. Chem. 273, 9457–9464
14. Aplin, H. M., Hirst, K. L., Crosby, A. H., and Dixon, M. J. (1995) Genomics 30, 347–349
15. MacDougall, M. (1996) Eur. J. Oral Sci. 104, 227–233
16. Crosby, A. H., Lyu, M. S., Lin, K., McBride, O. W., Kerr, J. M., Aplin, H. M., Fisher, L. W., Young, M. F., Kokaz, C. A., and Dixon, M. J. (1996) Mamm. Genome 7, 149–151
17. Ritchie, H. H., Hou, H., Veis, A., and Butler, W. T. (1994) J. Biol. Chem. 269, 211–220
18. Butler, W. T., Bhown, M., Dimuzio, M. T., Cohran, W. C., and Linde, A. (1983) Arch. Biochem. Biophys. 225, 178–186
19. Chang, S. R., Chiego, D. J., and Clarkson, B. H. (1996) Calcif. Tissue Int. 58, 149–153
20. Be`gue-Kirn, C., Ruch, J.-V., Ridall, A. L., and Butler, W. T. (1998) Eur. J. Oral Sci. 106, 254–259
21. Gu, K., Chang, S., Ritchie, H. H., Clarkson, B. H., and Rutherford, R. B. (2000) Eur. J. Oral Sci. 108, 35–42
22. Neame, P., and Butler, W. T. (1996) Connect. Tissue Res. 35, 145–150
23. Bartlett, J. D., Simmer, J. P., Xue, J., and Margolis, H. C. (1996) Gene (Amst.) 183, 123–129
24. Be`gue-Kirn, C., Kresbach, P. H., Bartlett, J. D., and Butler, W. T. (1998) Eur. J. Oral Sci. 106, 963–970
25. Li, W., Machule, C., Gao, C., and DenBesten, P. K. (1999) Eur. J. Oral Sci. 107, 352–359
26. Ryu, O. H., Fincham, A. G., Hu, C. C., Zhang, C., Qian, Q., Bartlett, J. D., and Simmer, J. P. (1999) J. Dent. Res. 78, 743–750
27. Bartlett, J. D., and Simmer, J. P. (1999) Crit. Rev. Oral Biol. Med. 10, 425–441
Identification and Characterization of the Carboxyl-terminal Region of Rat Dentin Sialoprotein
Chunlin Qin, Richard G. Cook, Ralph S. Orkiszewski and William T. Butler

J. Biol. Chem. 2001, 276:904-909.
doi: 10.1074/jbc.M006271200 originally published online October 19, 2000

Access the most updated version of this article at doi: 10.1074/jbc.M006271200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 28 references, 5 of which can be accessed free at
http://www.jbc.org/content/276/2/904.full.html#ref-list-1