Dentin is the major mineralized extracellular matrix of the tooth. The organic components of dentin consist of type I collagen (90%) with 10% noncollagenous proteins, which are also components of bone. Two dentin proteins, dentin sialoprotein and dentin phosphoprotein, have been shown to be tooth-specific being expressed mostly by odontoblast cells. In this study, we screened a mouse molar tooth library for dentin sialoprotein and dentin phosphoprotein cDNA clones. Analysis of the clones resulted in characterization of a 4420-nucleotide cDNA that contained a 940-amino acid open reading frame. The signal peptide and NH2-terminal sequence of this protein codes for amino acid sequence consistent with that of dentin phosphoprotein. The noncoding region contains three potential polyadenylation signals, two of which were shown to be utilized. Northern blot analysis indicated the presence of two major transcripts of 4.4 and 2.2 kilobases in odontoblasts. Chromosomal mapping localized the gene to human chromosome 4. These data suggest that the previously identified dentin extracellular matrix proteins, dentin sialoprotein and dentin phosphoprotein, are expressed as a single cDNA transcript coding for a protein that is specifically cleaved into two smaller polypeptides with unique physical-chemical characteristics. Therefore, we propose that the gene be named dentin sialoprotein. The location of the human dentin sialoprotein gene on chromosome 4 suggests that this gene may be a strong candidate gene for the genetic disease dentinogenesis imperfecta type II.

During tooth formation instructive epithelial-mesenchymal interactions result in the cytodifferentiation of ectomesenchymal cells that line the dental pulp chamber into highly specialized cells termed odontoblasts. A consequence of odontoblast cytodifferentiation is the expression of specific genes products that form the dentin extracellular matrix (DECM). The organic components of dentin consist of mostly hydroxyapatite (70% by weight) with the remaining 12% composed of water. The organic components of dentin consist of mostly type I (approximately 86%), type I trimer, type III, type V, and type VI collagens, and several noncollagenous proteins. The noncollagenous DECM proteins include those proteins found in bone ECM, such as osteonectin, osteocalcin, osteopontin (OPN, also known as SSP1), bone sialoprotein (BSP), and dentin matrix protein 1 (DMP1) (1, 2). Two DECM proteins, dentin sialoprotein (DSP) (3, 4) and dentin phosphoprotein (DPP, also known as phosphophoryn) (5, 6), have been shown to be tooth-specific.

DPP is a 95-kDa glycoprotein identified first within the DECM (7) and further characterized by cDNA cloning (3) from a rat tooth library. This protein accounts for 5–8% of the DECM and has a high carbohydrate (30%) and sialic acid (10%) content. This protein has an overall resemblance to other sialoproteins like BSP and shares limited NH2-terminal sequence homology with other acidic phosphoproteins OPN, DMP1, and bone acidic glycoprotein-75. In situ hybridization studies have shown DSP expression to be tooth-specific confined to differentiating odontoblasts, with transient expression in prescriptive ameloblasts (4). DSP is the major noncollagenous DECM protein, representing as much as 50% of this fraction. DPP is strongly associated with the mineral phase of dentin, being soluble only after demineralization of the extracellular matrix. In vivo (8) and in vitro (9, 10) studies have demonstrated that DPP is synthesized by odontoblasts that form a single sheet of cells lining the dental papilla mesenchyme. Immunolocalization studies have shown DPP to be confined to the mineralized dentin layer of the DECM (5, 8), being secreted through the odontoblast cell process at the mineralization front (5, 11). DPPs from several species have been characterized; although differences in the number and molecular weight size of these phosphoproteins have been reported, all share several unique physical-chemical characteristics. These properties include the following: aspartic acid and serine comprise approximately 70–80% of the total amino acid residues (12); a high degree of phosphorylation (400/1000 residues phosphate or greater), usually as phosphoserine residues (13, 14); extreme anionic character (15), with the abbreviation used are: DECM, dentin extracellular matrix; BSP, bone sialoprotein; DMP1, dentin matrix protein 1; DPP, dentin phosphoprotein; DSP, dentin sialoprotein; DSPP, dentin sialophosphoprotein; OPN, osteopontin; PCR, polymerase chain reaction; SSP, single specific primer; bp, base pair(s); kb, kilobase(s); Pipes, 1,4-piperazinediethanesulfonic acid.
the reported isoelectric point for rat of 1.1 (16); and a strong affinity for calcium ions (17, 18) being preferentially precipitated by this ion (19). Although the biological function of DPP is not known, physicochemical properties suggest a function in the biomineralization of the dentin extracellular matrix. Most theories have centered on DPP acting as a nucleator or modulator of hydroxyapatite crystal formation (1). Studies in rat have suggested the DECM may contain multiple forms of the DPPs (20). These may represent classes or families of phosphoproteins that differ in their content of phosphoserine and primary amino acid sequences, or they are alternatively spliced DPP gene products. Currently the exact number of DPPs is not known, nor has the primary amino acid sequence been determined by cDNA cloning.

The purpose of this study was to identify the cDNAs for mouse DPP and DSP by screening a molar cDNA library in order to facilitate investigations related to the regulation of odontoblast cytodifferentiation and matrix-mediated biomineralization. Synthetic oligonucleotide polymerase chain reaction (PCR) primer set was designed to sequence of the rat DSP cDNA. Limited polypeptide NH2-terminal sequence of mouse DPP amino acid sequence for oligonucleotide probe construction was determined by this ion (19). Although the biological function of DPP is not known, physicochemical properties suggest a function in the biomineralization of the dentin extracellular matrix. Most theories have centered on DPP acting as a nucleator or modulator of hydroxyapatite crystal formation (1). Studies in rat have suggested the DECM may contain multiple forms of the DPPs (20). These may represent classes or families of phosphoproteins that differ in their content of phosphoserine and primary amino acid sequences, or they are alternatively spliced DPP gene products. Currently the exact number of DPPs is not known, nor has the primary amino acid sequence been determined by cDNA cloning.

The purpose of this study was to identify the cDNAs for mouse DPP and DSP by screening a molar cDNA library in order to facilitate investigations related to the regulation of odontoblast cytodifferentiation and matrix-mediated biomineralization. Synthetic oligonucleotide polymerase chain reaction (PCR) primer set was designed to sequence of the rat DSP cDNA. Limited polypeptide NH2-terminal sequence of mouse DPP (21), bovine (22, 23), and rat (20, 24) DPPs were used to generate degenerative oligonucleotide probes for library screening. After initial primary and secondary screenings, it was determined that some of the individually identified clones hybridized with both DPP- and DSP-specific primer sets or probes. Analysis of these clones resulted in characterization of a full-length mouse cDNA with a large single open reading frame containing coding information for both DSP and DPP. In this study, we report the full-length sequence of this primary transcript, complete amino acid sequence of DPP for the first time, the utilization of alternative polyadenylation signals, and chromosomal mapping of this tooth-specific gene that we have designated dentin sialophosphoprotein (DSPP).

**EXPERIMENTAL PROCEDURES**

**Mouse Molar cDNA Library Construction—**Newborn (day 19) Swiss Webster mice first and second mandibular and maxillary molars were dissected and frozen immediately on dry ice. Poly(A+) RNA was extracted from the molars using the FastTrack mRNA Isolation Kit (Invitrogen, San Diego, CA). A total of 5 μl of mRNA was converted to cDNA with Moloney murine leukemia virus reverse transcriptase using ZAP-cDNA Synthesis Kit (Stratagene, La Jolla, CA) according to the supplier’s instructions. The cDNA was size fractionated using Sephacryl S-400 columns. First and second strand synthesis cDNA was analyzed on alkaline agarose gels to determine the size range and presence of any 2'-structure. The double stranded DNA was polished and filled in with Klenow, and EcoRI adapters were ligated to the blunt ends. Digestion with XhoI allowed insertion into the vector in a sense orientation (EcoRI-XhoI) with respect to the lacZ promoter. The Uni-Zap vector was double digested with EcoRI and XhoI, and after ligation of cDNA into the vector arms, aliquots were packaged into Gigapack II Gold packaging extract (Stratagene) according to the packaging instructions.

**Generation of Oligonucleotide Primers and Probes—**Specific DPP and degenerative DPP oligonucleotide probes were constructed by the DNA Core Laboratory (University of Texas Health Science Center). Table I gives the sequence of all degenerative oligonucleotide primers used to initially screen for DPP constructed against rat, mouse, and bovine DPP sequences. Because the codon usage was not known initially, we constructed degenerative primers that contain all the possible codons. In the case of sequences containing serines residues, we elected to use an inosine in the third nucleotide position to decrease the total number of possibilities. All generated oligonucleotide sequence data were analyzed using PCR Primer Selection program (Epicerent Software, Pasadena, CA) to check for secondary structure, self-annealing, and primer dimer formation.

The DSP oligonucleotide primers designed were as follows: DSP 739–758 ACGGGATAGAGGAGGATGA and DSP 1161–1141 TTCCACT- GCCTTCCCAGA. A 2-μl aliquot of the molar library was used for PCR amplification with the following program: 94°C for 2 min followed by 94°C for 1 min, 58°C for 1 min, and 72°C for 1 min for 30 cycles and a final extension at 72°C for 2 min. The PCR reaction resulted in a single DNA product of 421 bp. The actin primer sequences resulted in a 145-bp product and were as follows: sense, 5'-CATCGTGGGCC- CGGCTCTAGGCACCA-3' and antisense, 5'-CGGTGGCCTTAGGTT- CAGGGGG-3'. The annealing temperature of these primers was 55°C, and the PCR amplification was performed for 30 cycles. DNA generated by these specific PCR amplification reactions was labeled to produce high specific activity probes by random oligonucleotide priming using the Prime-It RmT Random Primer Labeling Kit (Stratagene). Oligonucleotide probes were 3' end-labeled with [32P]ATP by using terminal transferase (Life Technologies, Inc.) according to the manufacturer’s instructions. Probes were purified from unincorporated nucleotides by purification on NucTrap columns (Stratagene) according to the supplier’s instructions.

**Screening the Library—**The amplified mouse molar library was titered and plated with ~4 × 106 plaque-forming unit/150-mm plate. The phage were grown overnight at 37°C and transferred to nylon duplicate filters. All filters were denatured by submerging in 1.5 M NaCl and 0.5 M NaOH for 2 min and neutralized in 1.5 M NaCl and 0.5 M Tris-HCl, pH 8.0, for 5 min by again submerging. Filters are then rinsed in 0.2 M Tris-HCl, pH 7.5, 2 × SSC for 30 s and blotted on Whatman 3-MM filters. DNA is cross-linked to the filter in a Stratalinker UV cross-linker.
linker (Stratagene). Filters were prehybridized in 2 × Pipes containing 50% deionized formamide, 0.5% SDS, 100 mg/ml of sonicated salmon sperm DNA and 10 × Pipes for 2 h at 37 °C in a hybridization oven. Hybridization was performed in 2 × Pipes, 50% formamide, 0.5% SDS, 100 mg/ml salmon sperm DNA and 10 × Pipes over night at 42 °C overnight. Filters were washed in 0.1 × SSC containing 1.0% SDS at 55–65 °C, blotted on 3-MM paper, wrapped in plastic wrap, placed in a cassette with intensifying screens, and exposed to x-ray film (Biomax MR, Kodak, Rochester, NY) for 1–3 days of exposure at ~80 °C.

Secondary Screening—“Putative” positive plaques were excised, placed in 1 ml of SM buffer with 20 μl of chloroform, diluted, and plated at 100–250 plaques on a 100-mm NZY plates incubating overnight at 37 °C. Lifts were made and treated as previously outlined. After confirmation of insert by DNA sequencing, frozen stocks are prepared and stored at ~80 °C.

Conversion of Lambda Clones to Plasmids—The cDNAs were converted from the phage vector into the pBluescript SK− phagemid vector using M13 helper phage according to the Stratagene protocol.

DNA Sequencing Analysis—Both strands of the mouse DSPP cDNA and mouse subclones were sequenced by the dideoxynucleotide chain termination method using the Sequenase Version 2.0 sequencing kit (U. S. Biochemical Corp.). Rat and human DSPP cDNAs were also sequenced using this method with an internal DSPS primer (DSPP 1296–1279, TTTGGGCTATTCCTTTTG) or vector-specific T3 primer, respectively. In addition sequence was determined using Automated DNA sequencing (Applied Biosystems, model 370A) at the Human Genome DNA Core Laboratory (University of Texas Health Science Center) or by National Biosystems, Inc. (Plymouth, MN).

Protein consensus sequences and DNA self-alignment diagonal matrix plots were determined using the MacVector DNA sequence analysis software (Kodak). Amino acid sequence alignments (rat versus mouse) and the percentage of homology were determined using the GenePro 6.1 software program (Riverside Scientific Enterprises). Amino acid compositions and isoelectric points were calculated using the MacBiopsic Software (Perkin-Elmer Sciem Instruments, Thornhill, ON, Canada).

SSP-PCR of Mouse DSPP cDNAs 3′ Ends—To investigate the alternative utilization of the three potential polyadenylation signals, we used a novel SSP-PCR strategy (26). PCR amplification was performed using the vector-specific T7 promoter sequence that flanks the 3′ cloning site and DSPP-specific primers constructed to regions upstream of the potential polyadenylation signals. The PCR oligonucleotides primers used were as follows: DSPSP3537–3553, 5′-TACTAAGTCCCCAACCC-3′, and T7, 5′-GAATACGCTACTATAGGCG-3′. The PCR reaction was 4 min at 94 °C for 1 cycle, 1 min at 94 °C, 1 min at 52 °C, and 1 min at 72 °C for 30 cycles followed by 2 min for 72 °C for 1 cycle. PCR amplification products were separated on 2% agarose gel and stained with ethidium bromide.

Odontoblast Cell Line MO6G3 Culture—A culture of an established odontoblast cell line MO6G3 (27) was plated at a concentration of 5 × 10^5 cells/ml in a T-150-mm culture flask and grown at 33°C until 100–250 plaques on a 100-mm NZY plates incubating overnight at 37°C. Lifts were made and treated as previously outlined. After confirmation of insert by DNA sequencing, frozen stocks are prepared and stored at ~80 °C.

Northern Blot Hybridization—The poly(A)+ mRNA was electrophoresed in a 0.8% agarose-formaldehyde gel and transferred to Nytran nylon membrane by 1-h downward alkaline transfer using a TurboBlotter Rapid Downward Transfer System (Schleicher & Schuell) according to the supplier’s instructions. After transfer the membrane was UV cross-linked as described previously. RNA blot was prehybridized for 2 h at 68 °C in 6 × SSC, 5 × Denhardt’s reagent, 0.5% (w/v) SDS, and salmon sperm DNA at 100 μg/ml. Labeled DSPS cDNA probe was generated by random priming as previously outlined, denatured at 100 °C for 5 min, cooled, and added to the hybridization buffer. The blot was hybridized for 18 h at 68 °C, washed with four changes of 2 × SSC buffer containing 0.1% SDS for 15 min each at room temperature and then washed with two changes of 0.1 × SSC buffer with 0.1% SDS for 15 min at 60 °C. The blot was dried and exposed to x-ray film (Biomax MR, Kodak) at ~80 °C for 2 days.

Human DSPS cDNA Hybridization Probe—The mouse DSP primer sets were used to screen a human third molar tooth cDNA library constructed in the vector Uni-Zap II vector (Stratagene). Normal extracted third molar from a young (14-year-old male) without history of caries was used to construct the cDNA library obtained from the University of Texas Health Science Center Oral Surgery Department. The mandibular and maxillary third molars were at late crown formation with open forming roots and frozen immediately on dry ice. Poly(A+) RNA was extracted, and cDNA was synthesized as previously outlined. The library was plated and screened using the SB2 DSPF cDNA. A positive 1.9-kb cDNA clone was sequenced and shown to contain a partial open reading frame of repetitive aspartic acid and serine residues. This partial human DSPF cDNA clone was used for the chromosomal mapping studies.

Southern Blot Hybridization Analysis of Somatic Cell Hybrid DNA—Samples of genomic DNA (10 μg) isolated from hamster, mouse, and human were cut with four restriction enzymes (EcoRI, HindIII, BamHI, and MspI) in order to determine the restriction enzyme that could distinguish the DSPS hybridization signal between the three species. EcoRI was selected allowing discrimination of the mouse, human, and hamaster signals.

A panel of monochromosomal somatic cell hybrid clones was used for the assignment of the human DMP1 gene locus (BIOS Laboratories, New Haven, Connecticut) and mouse DMP1 in order to determine the restriction enzyme that could distinguish the DSPS hybridization signal between the three species. EcoRI was selected allowing discrimination of the mouse, human, and hamaster signals.

In order to isolate the full-length cDNA for the DECM proteins DSP and DPP, we first constructed a cDNA library for Swiss Webster mice newborn molars. We have previously shown that this developmental stage of tooth represents the most active dentinogenesis transcripts of the DECM protein, with minimal activity of the epithelial ameloblast (amelogenesis). The newborn molar library was initially plated and evaluated by 1) screening for an abundant mRNA sequence, actin, known to be present, and 2) screening with a total mixed cDNA probe used to construct the library. A labeled actin probe was generated by using actin-specific primers and adding radioactive dNTPs to the PCR amplification reaction mixture. Hybridization resulted in an estimated frequency of actin cDNA clones of 0.5%, which is within the range reported for other tissues. Hybridization with the total molar cDNA resulted in 87% positive clones, which is well within the expected value range of 50–95% hybridization for cDNA libraries.

In order to establish the presence of the DSP cDNA within the tooth library prior to library screening, an initial PCR amplification was performed. Primers specific for the NH2-terminal coding region of DSP were constructed based on the rat cDNA sequence (3). A 2-μl aliquot of the mouse molar library was used for PCR amplification with the DSP primer set resulting in a single DNA product of ~400 bp. This DSP amplification product was subcloned, sequenced, and confirmed to code for mouse DSP. This mouse DSP fragment was labeled and used to screen the library for a full-length cDNA. In parallel, the constructed DPP oligonucleotide probes were end-labeled, pooled, and used to screen the library.

Initial screening of 4 × 10^5 recombinants resulted in eight primary clones that were found to hybridize with both DSP- and DPP-specific probes. These clones were rescued to phage-mids and cDNA insert size determined by restriction enzyme digestion using EcoRI/XhoI or XbaI/KpnI and T3/T7 PCR amplification. The clone 5B2 was determined to have the largest insert size, and both strands were sequenced. The nucleotide sequence of this mouse cDNA is shown in Fig. 1. The DNA sequence, 4420 base pairs, was found to contain an open reading frame of 540 amino acids starting with a translation start site (ATG) at base 86 that had a ~3 adenine nucleotide representation of the Kozak initiation consensus sequence. A 17-amino acid leader hydrophobic sequence is present, suggesting targeting for the endoplasmic reticulum and secretion. The stop codon, at base 2905, begins an untranslated region of 1515 nucleotides with three polyadenylation signals (AAATAA) at bases 3602–3607, 3795–3800, and 4384–4389. Northern blot analysis of mRNA isolated from an odontoblast cell line (MO6G3) showed the 5B2 cDNA probe hybridized to two distinct transcripts of 4.4 and 2.2 kb (Fig. 2).
Amino Acid Analysis and Composition—The predicted complete translation product encodes an acidic protein with a calculated molecular weight (Mr) of 92,569 excluding the signal peptide. The signal peptide and NH₂-terminus (amino acids 1–387, bases 86–1246) has 75% homology with the published rat “complete” DSP cDNA clone RDSP2 (Fig. 3). The mouse DSP has two small deletions following amino acids 149 (1 amino acid) and 313 (2 amino acid), and an insert of five amino acids beginning at amino acid 338. The last four amino acids of the reported rat DSP cDNA vary between the two species with the mouse clone not containing the reported stop codon (TAA) of the rat sequence in this region.

The extended continuous open reading frame of the mouse 5B2 transcript contains an integrin binding Arg-Gly-Asp (RGD) sequence at amino acid 479 (base 1520), which is contained in other acidic phosphoproteins of bone and dentin such as BSP, OPN, and DMP1. In addition, the most 3’ portion of the open reading frame (amino acids 452–940, 1439–2905 nucleotides) codes for an extended region consisting of aspartic acid and serine residues, as well as sequences with homology to the DPP degenerative oligonucleotide probes used for the library screening.

The total 940-amino-acid translation product, which we termed dentin sialophosphoprotein (DSPP), is an acidic protein (pI 5.4), rich in aspartic acid (18.9%) and serine (36.3%) residues (Fig. 4). The portion previously identified as DSP (amino acids 18–387, nucleotides 137–1246) is 370 amino acids with a composition consistent with that published for the rat cDNA sequence and DSP isolated from the DECM (Fig. 4). The portion identified as DPP we begin at base 1439 with NH₂-terminal sequence established for rat HP-2 through the stop codon at nucleotide 2905. This portion of the protein codes for 489 amino acids.

**Fig. 1.** DNA sequence of the mouse dentin sialophosphoprotein and deduced amino acid sequence. The amino acid and DNA sequence numbering is indicated on the right side. The signal peptide is underlined with a single line, the NH₂-terminal cleavage site is indicated by a horizontal line, and the stop codon indicated by an asterisk. The three polyadenylation signals are underlined with a double line. The RGD sequence is shown in bold, and DNA sequences with homology to the dentin phosphoprotein oligonucleotides are underlined with dotted lines; the names of DPP peptide sequences with homology are indicated above. The end of the DSP coding region is indicated by a delta, and the beginning of the DPP coding region is indicated by an arrow.

**Fig. 2.** The extended continuous open reading frame of the mouse 5B2 transcript contains an integrin binding Arg-Gly-Asp (RGD) sequence at amino acid 479 (base 1520), which is contained in other acidic phosphoproteins of bone and dentin such as BSP, OPN, and DMP1.
Acids (amino acids 452–940) with a calculated molecular mass of 47.8 kDa and is rich in aspartic acid (28.2%) and serine (57.3%) with a calculated pI of 3.3. The composition of this mouse DPP HP-2 region is nearly identical with the actual amino acid analysis of mouse DPP isolated from the DECM (Fig. 4).

**Potential Post-translational Modification Sites**—Protein sequence analysis revealed 18 N-glycosylation sites based on the conserved sequence NX(S/T), seven are within the DSP coding region, whereas the remaining eleven are in the DPP portion of the transcript. These potential sites are at amino acids 54, 61, 84, 130, 190, 246, 474, 494, 538, 586, 685, 763, 793, 811, 880, and 928. Potential casein kinase I and II phosphorylation sites are present within both the DSP and DPP regions. A total of 41 CK I sites are identified, six within the DSP region and the NH2-terminal region of the mouse DSP, with 34 sites with the DPP region. The last four amino acids are 100% homologous.

**Polyadenylation Signal Utilization**—The 3′ untranslated region of the DSPP contains three potential polyadenylation signals (AATAAA). The two signals not utilized in the clone 5B2 have GT-rich or T-rich segments within short distances downstream, which is consistent with utilization in many vertebrate genes. To test if these alternative signals are used we performed SSP-PCR. PCR amplification of the mouse molar cDNA library was performed using a DSPP primer upstream of the first polyadenylation signal and a vector-specific T7 primer. Amplification resulted in two detectable PCR products of ~900 and ~250 bp (Fig. 7). These would correspond to the utilization of the second and third polyadenylation signals. The use of the first polyadenylation signal was not evident in that no PCR products of ~100 bp were amplified.

**Human Chromosomal Mapping**—Southern blot analysis of DNA from mouse, human, and hamster DNA digested with four restriction enzymes revealed single hybridization bands of 10, 8.8, and 18 kb, respectively, for the enzyme EcorI when probed with a 32P-labeled human DSPP probe. Hybridization of EcorI-digested DNA from a human-rodent monochromosomal cell hybrid panel was performed in order to determine the chromosome locus for human DSPP. The 8.8-kb hybridization band identified for the human genomic DNA was present only in a lane that contained human chromosome 4 and 7 DNA (cell line 7A4AR) (Fig. 8) but was not detected in the chromosome 7 only sample (cell line 0A7AR). The cell line 0A2GR, which also contains fragments of chromosomes 8 and 4, did not hybridize to the DSP probe. These data indicate that the DSPP gene is located on human chromosome 4.

**DISCUSSION**

We have identified a mouse primary cDNA transcript that contains the coding information for both DSP and DPP; the two previously identified tooth-specific DECM proteins. Our data show for the first time that these dentin proteins are expressed as a single large transcript of 940 amino acids that is specifically cleaved into distinct polypeptides, which have been recognized as DSP and DPP by their unique physicochemical properties. We call the primary transcript and gene dentin sialophosphoprotein (DSP) to reflect this fact. This nomenclature allows the distinction to be made between the two major matrix proteins or polypeptide fragments (DSP and DPP) and the primary transcript or gene while retaining the current

![Diagram](image)

**Fig. 2. Northern blot analysis of the mouse dentin sialophosphoprotein transcripts.** Poly(A)+ RNA was isolated from mouse odontoblast cell line M06G3, separated on a denaturing agarose gel, and transferred to a membrane. Labeled mouse DSPP 5B2 cDNA was hybridized in A, and a housekeeping gene glyceraldehyde 3-phosphate dehydrogenase was hybridized in B.
nomenclature for these matrix proteins.

Ritchie et al. (3) reported the complete rat DSP cDNA sequence based on amino acid comparison analysis with DSP protein isolated from the DECM. This clone contain only 12 nucleotides after the identified stop codon (TAA), with no polyadenylation signals or poly(A) tail sequence reported. We have detected a single base sequencing error in the 3' end of their rat DSP cDNA sequence, which lead to the misinterpretation of a stop codon at nucleotide 1161. Our data that the DSP is part of a much larger transcript (4420 bp) are supported by their Northern blot analysis indicating that two transcripts at 4.6 and 1.5 kb are present in rat tooth RNA.

We have utilized a previously characterized mouse odontoblast cell line, MO6G3, for analysis of the DSPP transcript (27). The expression of DSPP has been shown by this cell line at both the transcriptional and translational levels, whereas the expression of DPP has been shown at a translational level. Our data showed two transcripts, the larger major 4.4-kb transcript is consistent with the DSPP clone characterized and reported in this paper. The slight variation in the size of the mRNA transcripts reported for rat and mouse may be due to the use of limited size markers (28 and 18S rRNA only) in the rat studies or due to species differences. At present, we cannot elucidate the biological significance of the presence of the two isoforms. However, these transcripts are most likely derived from a single gene by alternative splicing, because Southern blot analysis for the chromosomal mapping studies revealed the presence of a single gene. The minor transcript at 2.2 kb may be due in part to the use of alternative polyadenylation signals which differ by up to 833 nucleotides.

Ritchie et al. (7) has reported the “entire” structure of the rat DSP gene indicating 5 exons. Because the signal peptide and initiation codons are located in the first three exons, these second mouse DSPP transcript (2.2 kb) should differ by deletion of exons 4 and 5 or the other exons within the DPP region yet to be determined. Further characterization of the mouse and human DSPP gene structure is currently underway in our laboratory. The N-glycosylation and phosphorylation sites reported for the rat sequence are all conserved in the mouse cDNA DSPP clone. It is interesting to note that all of the phosphorylation sites are located in the transcript near the region containing the DPP coding information. In addition, a RGD sequence has been found in the DSPP mouse cDNA, as also determined for other acidic phosphoproteins such as DMP1, BSP, and OPN.

DNA sequence determination of the DSPP mouse cDNA has revealed the complete amino acid sequence of DPP for the first time. Determination of the amino acid sequence of this protein has been extremely difficult by conventional Edman degradation due the high level of phosphorylation. We have determined limited amino acid sequence for the dephosphorylated mouse protein fragments (45 and 40 kDa) by Edman degradation (28). The results showed NH₂-terminal sequence as follows: 45-kDa

---

**Amino Acid Comparisons**

| Amino Acid | DSPP | DSPP | DSP₃₁₀ | DSP₃₇₀ | rDSP₁DNA | rDSP² | DPP₄₈₉ | DPP₄₉₉ | mDSP³ | rDSP⁴ | bDSP⁴ |
|------------|------|------|--------|--------|----------|--------|--------|--------|--------|--------|--------|
| Alanine    | 27   | 29   | 20     | 54     | 29       | 24     | 4      | 8      | 10     | 14     | 11     |
| Cysteine   | 3    | 3    | 1      | 3      | 2        | 0      | 2      | 4      | --     | --     | --     |
| Aspartate  | 174  | 189  | 29     | 78     | 51*      | 59*    | 136    | 282    | 339*   | 374*   | 403*   |
| Glutamate  | 63   | 68   | 44     | 119    | 42       | 65+    | 14     | 29     | 36+    | 43+    | 37+    |
| Phenylyalanine | 3 | 3    | 2      | 5      | 1        | 3      | 0      | 2      | 4      | 0      |
| Glycine    | 76   | 85   | 56     | 151    | 59       | 56     | 14     | 29     | 29     | 52     | 34     |
| Histidine  | 20   | 22   | 13     | 35     | 11       | 12     | 3      | 6      | 3      | 10     | 10     |
| Isoleucine | 19   | 21   | 17     | 46     | 19       | 10     | 1      | 2      | 2      | 6      | 3      |
| Lysine     | 24   | 26   | 14     | 38     | 15       | 15     | 5      | 10     | --     | 56     | 13     |
| Leucine    | 14   | 15   | 12     | 32     | 11       | 13     | 0      | 0      | 4      | 10     | 2      |
| Methionine | 1    | 1    | 0      | 1      | 0        | 0      | 0      | --     | 2      | --     | --     |
| Asparagine | 48   | 52   | 26     | 70     | 10       | 10     | 18     | 37     | 3      | 7      | --     |
| Proline    | 21   | 23   | 17     | 46     | 16       | 0.5    | 2      | 4      | 6      | 9      | 4      |
| Glutamine  | 25   | 27   | 21     | 57     | 0        | 0      | 0      | --     | 2      | 4      | --     |
| Arginine   | 12   | 13   | 8      | 22     | 2        | 2      | 4      | 4      | 10     | 2      | --     |
| Serine     | 336  | 363  | 43     | 116    | 40       | 40     | 281    | 574    | 551    | 382    | 470    |
| Threonine  | 31   | 34   | 26     | 70     | 29       | 23     | 4      | 8      | 10     | 13     | 12     |
| Valine     | 20   | 22   | 19     | 51     | 16       | 12     | 0      | 0      | --     | 8      | --     |
| Tryptophan | 0    | 0    | 0      | 0      | 0        | --     | 0      | 0      | --     | --     | --     |
| Tyrosine   | 4    | 4    | 2      | 5      | 3        | 2      | 1      | 2      | 4      | 10     | 2      |

| Mr   | 92.6 | 38.1 | 47.8 |
| pl   | 4.0  | 5.4  | 3.3  |
| Range | 17-940 | 17-387 | 451-940 |

- Aspartate + Asparagine
- Glutamate + Glutamine

1- Ritchie et al. (3)
2- Butler et al. (4)
3- MacDougall et al. (5)
4- Gorter de Vries et al. (6)

**FIG. 4.** Amino acid composition of the mouse dentin sialophosphoprotein and comparison of regions for cleavage products dentin sialoprotein and dentin phosphoprotein.

---

for the chromosomal mapping studies revealed the presence of a single gene. The minor transcript at 2.2 kb may be due in part to the use of alternative polyadenylation signals which differ by up to 833 nucleotides.

Ritchie et al. (7) has reported the “entire” structure of the rat DSP gene indicating 5 exons. Because the signal peptide and initiation codons are located in the first three exons, the second mouse DSPP transcript (2.2 kb) should differ by deletion of exons 4 and 5 or the other exons within the DPP region yet to be determined. Further characterization of the mouse and human DSPP gene structure is currently underway in our laboratory. The N-glycosylation and phosphorylation sites reported for the rat sequence are all conserved in the mouse cDNA DSPP clone. It is interesting to note that all of the phosphorylation sites are located in the transcript near the region containing the DPP coding information. In addition, a RGD sequence has been found in the DSPP mouse cDNA, as also determined for other acidic phosphoproteins such as DMP1, BSP, and OPN.

DNA sequence determination of the DSPP mouse cDNA has revealed the complete amino acid sequence of DPP for the first time. Determination of the amino acid sequence of this protein has been extremely difficult by conventional Edman degradation due the high level of phosphorylation. We have determined limited amino acid sequence for the dephosphorylated mouse protein fragments (45 and 40 kDa) by Edman degradation (28). The results showed NH₂-terminal sequence as follows: 45-kDa
Ser-Ser-Asp-Ser-Ser-Met-Ser-Ser and 40-kDa Ser-Ser-Ser-Ser. The mouse DSPP clone contains within the DPP coding region, six segments of polyserine residues (4–5 residues) consistent with the sequence determined for the 40-kDa fragment. The 45-kDa NH$_2$-terminal sequence has homology (7 out of 8 amino acids) with the highly repeated domain of Ser-Ser-Asp-Ser-Ser-Asp-Ser-Ser-Asp found throughout the DSPP cDNA. This sequence is located more upstream of the repeated serines segments and therefore would result in a larger polypeptide fragment.

Studies in rat have suggested that the DECM may contain multiple forms of the DPPs (20, 21). These may represent several classes or families of phosphoproteins that differ in their content of phosphoserine amino acid sequences or may represent alternatively spliced DPP gene products. The multiple rat DPPs have been termed related to their degree of phosphorylation highly (HP-1 and HP-2), moderately (MP) and low (LP) phosphorylated phosphoproteins (20, 21). These rat DPPs differ in their amino acid compositions with the HPs containing high levels of aspartic acid and serine residues, whereas MP and LP, although enriched in these two amino acids, contain higher levels of other amino acids such as glutamic acid, glycine, leucine, and arginine. NH$_2$-terminal amino acid sequence determination of the HP class performed by Butler et al. (20) generated two unique sequences: HP-1, Asp-Asp-Asp-Asn, and HP-2, Asp-Asp-Pro-Asn-Asp-Asp-Glu. The mouse DSPP transcript has homology with both the NH$_2$-terminal HP-1 and HP-2 sequence. We have set the HP-2 region as the beginning of the mouse DPP protein for calculation of molecular weight and amino acid composition because this sequence is found initially at amino acid 452. The HP-1 sequence is found more downstream beginning at amino acid 491 with 100% homology. The MP or LP protein may represent the linker protein region, which is found between the DSP and DPP coding regions, or another cleavage polypeptide, which occurs within the DPP region at the COOH terminus.

Additional, internal amino acid sequence has been determined using tryptic digestion of the dephosphorylated rat DPP

---

**FIG. 5.** Diagonal self-plot of the mouse dentin sialophosphoprotein cDNA sequence. The 4420-nucleotide DSPP cDNA sequence is presented on both the vertical and horizontal axes. The central diagonal represents the homology expected from aligning two identical sequences. The black box is produced by the tight clustering of small parallel lines between residues 1725 and 2750 indicating a region of extensive internal homology. This portion of the DSPP cDNA codes for the highly repetitive DPP protein.

**FIG. 6.** DNA sequence analysis of the rat dentin sialoprotein and mouse 5B2 cDNA. A, is the autoradiograph showing DNA sequencing ladders of rat and mouse cDNA using a internal dentin sialophosphoprotein oligonucleotide primer downstream of the reported "stop" site. The arrow indicates the region of the potential frameshift and DNA sequencing error. B, is the deduced amino acid sequence of the original rat DSP sequence (Rat') and corrected revised rat sequence compared with the mouse dentin sialophosphoprotein cDNA clone sequence.

**FIG. 7.** PCR analysis of the utilization of multiple polyadenylation signals for the mouse dentin sialophosphoprotein cDNA. A specific dentin sialophosphoprotein oligonucleotide was designed upstream of the first polyadenylation signal and used in a PCR reaction with a vector specific T7 primer located at the 3’ end of the cDNA inserts. Lane 1, low molecular weight DNA size marker with sizes indicated on the left. Lane 2, PCR amplification products generated by amplification of the total mouse cDNA library using primers DSPP$_{3537-3553}$ and T7. This demonstrates that the second and third polyadenylation signals are used.

**FIG. 8.** Southern blot analysis performed on DNA derived from somatic cell hybrids containing individual human chromosomes depicted in lanes 1–22, X, and Y in a mouse or hamster background using a human 1.9-kb DSPP probe.
generating the sequence Asp-Asp-Asp-Asp-Asp-Tyr-Ser-Asp-Ser-Asp-Ser-Asp-Asp-Asp. Furthermore, repetitive (Ser)n (Asp)n and (Ser/Asp)n residues blocks were also identified (22). Incomplete mild acid hydrolysis procedure has also suggested that bovine DPP contains similar repetitive blocks of (Asp-Ser(P))n, (Asp)n, (Ser(P))n, (Asp)-Y)n, where Y is phosphoserine, glycine, or alanine (29, 30). Automated Edman gas-liquid phase sequencing of bovine DPP resulted in a 23-amino acid sequence Ser-Asp-Pro-Asn-Ser-(Ser)-Asp-Glu-Asp-Asn-Gly-Asp-Ala-Ala-Asn-Ser-Asp-(Ser)-Asp-Ser with uncertainties existing at positions 6 and 20 (22). This sequence lends support to the existence of (Ser(P)-Asp)n repeats in the bovine DPP. A trypsin-digested peptide from bovine DPP has been identified that contained a NH2-terminal sequence of sixteen serines, some of which were phosphorylated (23). The most extended sequence obtained to date is for bovine DPP (150 kDa), reported by Crossley et al. (24) by conversion of the Ser(P) residues to S-propyl-cysteinyl residues by Ca2+-catalyzed β-elimination prior to Edman degradation analysis. A total of 50 amino acids were determined with NH2-terminal of Asp-Ser(P)-Pro-Asn-Ser-Asp(P)- followed by a repeating peptide of Asp-(Ser(P))n where n = 1–3. Sequences within the mouse DSPP cDNA translation share homology with all of the reported DPP sequences. The exception is that no regions of repeated aspartic acid residues are found within the mouse sequence. This could be due to errors in amino acid sequence analysis of the protein or species differences. The presence of highly repetitive blocks of aspartic acid and serine is very evident in the DNA self-plot of the mouse cDNA. A very unusual “black box” is apparent that is formed by the extreme number of close parallel lines representing regions of self-homology.

Human chromosome 4, where the DSPP gene locus has been located, contains the genes for several other dentin/bone extracellular matrix acidic phosphoproteins that have cell binding RGD sequence domains. These include BSP, SPP1 (OPN), and DMP1. DMP1 has been mapped to 4q21 by fluorescence in situ hybridization (31), BSP has been relocalized to 4q21-q23 (32) by PCR somatic cell hybrid mapping from 4q23-q28 (33) determined by in situ hybridization, and SPP1 has been localized to 4q21-q23 by somatic cell mapping (32, 34). DMP1, BSP, and SPP1 (OPN) have been mapped on a yeast artificial chromosome clone panel within a maximum region of 490 kb. This region of human chromosome 4 apparently contains a superfamil family gene locus of related acidic phosphoproteins that are important in the processes of biomineralization. Because the DSPP has a number of conserved features with these other proteins such as the RGD sequence, highly repetitive sequence, and potential role in mineralization, we predict that DSPP will also map within this region 4q21-q23.

The dental disease dentinogenesis imperfecta type II, which affects formation and mineralization of the DECM, has also been mapped to 4q21-q23, within a 3.2-centimorgan region surrounding the SPP1 (OPN) gene locus (35). Detailed analysis of the SPP1 gene (six coding exons) in affected individuals found no mutations eliminating OPN as a candidate gene (36). DPP was one of the first genes suggested as a likely candidate for dentinogenesis imperfecta type II. However, we initially reported that the DPP gene did not map to human chromosome 4 using a degenerate oligonucleotide probe based on the rat HP-2 NH2-terminal sequence (37). The oligonucleotide sequence constructed in that study, we know now, does vary from the mouse DPP sequence reported here. Our more recent data, in this report, clearly show that the DSPP gene is located on human chromosome 4. Therefore, DSPP is a strong candidate gene for dentinogenesis imperfecta type II due to its restricted pattern of expression within the tooth and potential role in biomineralization of the DECM.