

Molecular cloning of a human dentin sialophosphoprotein gene

Keni Gu, Syweren Chang, Helena H. Ritchie, Brian H. Clarkson, R. Bruce Rutherford

Department of Cariology, Restorative Sciences and Endodontics, School of Dentistry, University of Michigan, Ann Arbor, Michigan, USA

Gu K, Chang S, Ritchie HH, Clarkson BH, Rutherford RB. Molecular cloning of a human dentin sialophosphoprotein gene. *Eur J Oral Sci* 2000; 108: 35–42.

© Eur J Oral Sci, 2000

Dentin sialoprotein (DSP) and dentin phosphoprotein (DPP; phosphophoryn) are two principal dentin-specific non-collagenous proteins. DPP is extremely acidic and is rich in aspartic acid and serine. By virtue of this structure, DPP may bind large amounts of calcium and may facilitate initial mineralization of dentin matrix collagen as well as regulate the size and shape of the crystals. The function of DSP is not known. DSP and DPP are encoded by a single gene in both rat and mouse, and are uniquely expressed in odontoblasts and transiently in pre-ameloblasts. Because DSP and DPP are isolated from dentin as distinct proteins and appear to be present in different amounts, the nascent dentin sialophosphoprotein (DSPP) is likely cleaved to yield DSP and DPP. However, when, where and how the DSPP is cleaved into DSP and DPP is not clear. To further elucidate the structure and function of human DSP and DPP, we have cloned DPP and DSP cDNA by reverse transcriptase-polymerase chain reaction (RT-PCR) strategies, and then cloned and initiated characterization of a human dentin sialophosphoprotein gene. The genomic organization of human DSPP is very similar to that of mouse, containing five exons and four introns, suggesting it is a homologue of mouse dentin sialophosphoprotein (DSPP). Exons 1–4 encode for DSP, while exon 5 encodes for the C-terminus of DSP and the whole DPP. A 4.6-kb RNA transcript was detected on Northern blot analyses of total RNA extracted from immature (open root apices) human teeth using either a human DPP or DSP probe.

Keni Gu, Department of Cariology, Restorative Sciences and Endodontics, School of Dentistry, University of Michigan, Ann Arbor, MI 48109, USA

Telefax: +1-734-9361597
E-mail: rongguke@umich.edu

Key words: Human dentin sialophosphoprotein (DSPP) gene; DSP; DPP; DNA sequence; amino acid sequence

Accepted for publication October 1999

Dentin sialoprotein (DSP) and dentin phosphoprotein (DPP), comprising about 5% and 50%, respectively, are the principal noncollagenous proteins found exclusively in the extracts of dentin extracellular matrix (1–3). DPP is synthesized and secreted by odontoblast (4) and transiently expressed by pre-ameloblasts (5). It is extremely acidic and rich in aspartic acid (D) and serine (S), which are present principally as repeating sequences of DSS and DS of varying lengths located in distinct regions of the polypeptide chain. These repeated domains of aspartic acid and serine residues may be phosphorylated through the combinatorial action of casein kinase I and II (6, 7). Hence DPP, which by virtue of its extensive phosphorylation, may bind large amounts of

calcium, and thus play a direct role in the nucleation of hydroxyapatite onto dentin matrix collagen. It may also help regulate the crystal size and shape the hydroxyapatite (1, 6, 7). The function of dentin sialoprotein is unknown. Another enigma is that whilst a single gene, dentin sialophosphoprotein (DSPP), encodes for both DSP and DPP in both rat and mouse (3, 8–10), the mature dentin matrices contain DSP and DPP as distinct proteins in approximately 10:1 (DPP: DSP) molar ratios (3).

Less information is currently available regarding the molecular and cellular biology of human DSP and DPP. Recently, a 140-kDa phosphoprotein was isolated from human immature, but not mature permanent tooth apices (11, 12). The amino acid

composition of the 140-kDa protein was consistent with that described for other mammalian dentin phosphoproteins. The N-terminal sequence was determined to be aspartic acid-aspartic acid-proline (P) (11), but further sequencing was ambiguous due to the instability of the phosphorylated peptide chain. However, this protein was believed to be a human dentin phosphoprotein. Interestingly, the DPP gene has been localized to human chromosome 4q21 within the locus of dentinogenesis imperfecta type II (DGI-II) (10), an autosomal dominant genetic disorder. The DPP levels are reduced in individuals with DGI-II (13, 14). These data implicate DPP in the pathogenesis of DGI-II.

The nucleotide and deduced amino acid sequence of a human DPP cDNA has been determined using cloning strategies based on the reverse transcriptase polymerase chain reaction (RT-PCR) (15). This clone demonstrated the expected arrays of DSS and DS repeats, a high percentage of aspartic acid and serine residues, a deduced N-terminal sequence comprising D-D-P, and a C-terminal 19 amino acid sequence identical to mouse (10) and rat (9, 16). These data strongly support the conclusion that the cDNA (GenBank AF 094508) cloned by RT-PCR methods from human immature root total RNA codes for the human DPP. In addition, we have cloned a partial cDNA sequence directly upstream of, and contiguous to, the human DPP cDNA sequence. This partial sequence is very similar to the mouse DSP and DSP-DPP linker sequence (10), suggesting that DSP and DPP are coded by a single transcript, and there is a human DSPP gene analogous to mouse and rat DSPP. Finally, a whole DSP cDNA sequence was obtained by RT-PCR and 5'RACE (rapid amplification of cDNA ends).

As a step in understanding the role of DSP and DPP in human dentinogenesis, dentin repair and

remineralization, as well as human genetic diseases affecting dentin formation, such as dentinogenesis imperfecta type II, we endeavored to clone and characterize the human DSPP gene. Using human DPP cDNA to screen a human genomic P1 library, we obtained a human DSPP P1 genomic clone. Here, we report the molecular cloning of a human dentin sialophosphoprotein (DSPP) gene.

Material and methods

Isolation of total RNA

Root tips were removed from freshly extracted human immature teeth and crushed into small shards. Approximately 50 mg of dentin shards were mixed with 1 ml of RNA STAT-60 (TEL-TEST B, Friendswood, TX, USA) and mechanically homogenized. The homogenate was mixed with 0.2 volume of chloroform and the phases separated. The RNA from the aqueous phase was precipitated with an equal volume of isopropanol at room temperature, centrifuged at $11,000 \times g$ at 4°C for 15 min, washed by 75% ethanol, and finally dissolved in diethyl pyrocarbonate-(DEPC) treated water and quantitated by ultraviolet (UV) spectrometry.

Primer design, RT-PCR, 5'RACE and sequencing

The primers used to obtain a human DSP cDNA sequence were based on the rat DSP sequence (17) and on the human DPP sequence (15). The primers for the 5'RACE were designed using Oligo (Primer Analysis Software, v.4; Wojciech Rychlik, Cincinnati, OH, USA).

cDNA was synthesized from total RNA of human root tips using RNA PCR Core Kit with random hexamers as primers for reverse transcription according to the manufacturer's protocol (Perkin Elmer Cetus, Foster City, CA, USA).

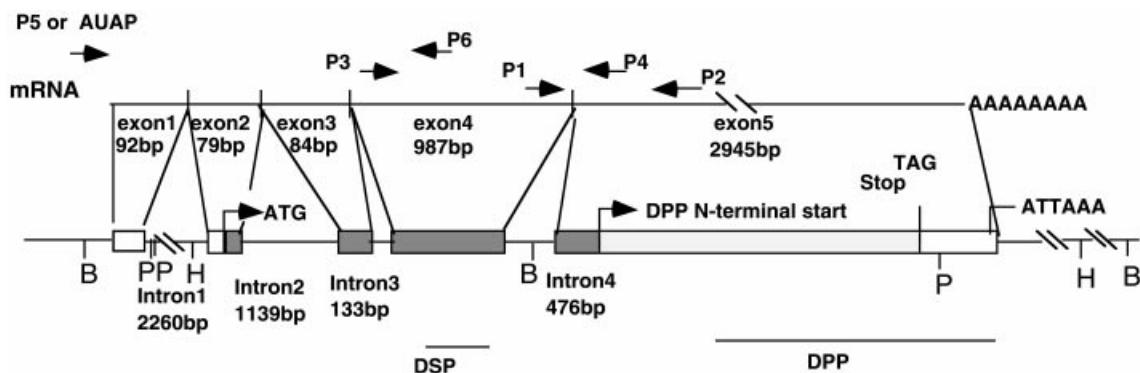


Fig. 1. Genomic organization of human dentin sialophosphoprotein gene. The opened box is for untranslated region. The closed box is for DSP including a signal peptide. The dotted box is for DPP. B: BamHI, H: HindIII, P: PstI. The sizes of the exons and the introns, positions of the primers used to obtain human DSP cDNA, and the locations of the probes used in Southern and Northern hybridizations as well as the cutting sites of restriction enzymes are depicted.

ATGCAAAAGCTCCAGGACAGTGGGCCACTTTCAGTCTTCAAGAGAAAGATAAGAAATTCGGAATTTTCAAAATCCTTTTGAAGCCCTTTAAG•CCATTGATTTATTATTTCCTAAAGAAA 120
 ATGAAGATAATTACATATTTTGGCATTGGCCAGTAGCAGGGCCATCCCA•GTTCCCTCAAGCAAAACCCTGGAGAGACATGTCGAAAATCCATGAAATTTGCATCTCCTAGCAGATCA 240
M K I I T Y V F C F W A V A W A I P V P Q S K P L E R H V E K S M N L H L L A R S 40
 AATGTGTCAGTACAG GATGAGTTAAATGCCAGTGGAAACCATCAAMGAAAGTGCTCCTGGTGCATGAAAGGTGATAGAGGAAGGCAAGAGAAATCCCAAGATGGTCCAAAGGGAGAGGG 360
 N V S V Q D E L N A S G T I K E S G V L V H E G D R G R Q E N T Q D G H K G E G 80
 AATGGCTCTAAGTGGCCAGAAGTAGGAGGGAGAGTTTCTACATATTCACATATAGCAAAAGAGAGGGAAATTTGAGGCTGGAATGGGGACACAGGAAAAGCAGAAACATATGTT 480
 N G S K W A E V G G K S F S T Y S T L A N E E E G N I E G W N G D T G K A E T Y G 120
 CATGATGGAATACATGGGAAAAGAAAACATCAGCAAGAAATGGCATCCAGGGACAAGTAAATCATTGACAATGCTGGAGCCACAACAGAAGCAACACTAATGAAATATCTGATAAG 600
 H D G I H G K E E N I T A N G I Q G Q V S I E I D N A G A T N R S N T N G N T D K 160
 AATACCCAAATGGGGATTGTGGCGATGCAGTCCAAATGAGGATGTCGGTGTGCAAGAAGATGGACCTCAAGTAGCTGGAAGCAATAACAGTACAGCAATGAGGATGAAATATT 720
 N T Q N G D V G D A G H N E D V A V V Q E D G P Q V A G S N N S T D N E D E I I 200
 GAGAATTCCTGTAGAAAAGCGGGTAATACAAGTAAATACCACTCAGATCAACAGCAAGAGAAAATGGGACTAAGGAAGCTGAGGTAACACCAGGCACCTGGAGAAAGTGCCTGGCTGGAT 840
 E N S C R N E G N T S E I T P Q I N S K R N G T K E A E V T P G T G E D A E G L D 240
 AATCCGATGGGACTCTAGTGGGAATGGAGCAGATGAGGATGAAGACAGGGTCTGCGTATGATGAAATGAAGAAGCAGGGAATGGAAAAGACAGTGAATAACAGCAAGGGCCAG 960
 N S D G S P S G N G A D E D E D E G S G D D E D E A G N G K D S S N N S K G Q 280
 GAGGGCCAGCACATGGGAAAAGAGATGATCATGATAGTACATAGGTCAAATTCGAGTAGTAAAGAAATATTATGACCCCTGAAGGCCAAGAAGATCCCAATTAATGAAAGTTGATGGAGAC 1080
 E G Q D H D S E S I G Q N S D S K E Y Y D P E G K E D P E G H E D G L D 320
 AAGACCTCCAGAGTGGAGAAATTCCTGGTATCCAGAAAGCAATGGCCGCAAGAAATAGAGGACACCCAGAAGCTCAACCATAGAGAAAGCAAAAGCGGTAGAAAATAGAAATCACC 1200
 K T S K S E E N S A G I P E D N G S Q R I E D T Q K L N H R E S K R V E N R I T 360
 AAAAGAATCAGAGCACATGCTGTGGGAAAGAGCCAAGATAAG•GGAATAGAAATCAAGGTTCCCAGCAGTGGCAACAGAAATATTACCAAAGAAGTTGGGAAAGGCAACCAAGGTTAAAGAG 132
 K E S E T K E S Q D K G I E I K G S G N R N I T K E V G K E D A E G L D 400
 GATAAAGGCAACATGGAAATGATCTGGGCAAGGCAATGTCAAGACACAAGGAGAGTGTCAACATAGAAGGACCTGGCCAAAATCAGAAACCAGAAAATAAGTTGGACACAGCAAT 1440
 D K G Q H G M I L G K G N V K T Q G E V V N I E G P G Q K S E P G N K V G H S N 440
 ACAGGTAGTGACAGCAATAGTATGAT 1560
 T G S D T G D G Y D S Y D S Y D S D D K S M Q G **D D P** N S S D E S N G N D E S 480
 AGTGACAAATACAGCAGTAGCCAGGAGATGCTTCTTAATCCTGATGAATCAAAAGATAATGGCAATGGCAGTACTCAAAAGGAGCAGAAGATGATGACAGTATAGTACATCAGAC 1680
 S D N N S S S R...D A S Y N S D E S K D N G N G S D S K G A E D D D S D S T S D 520
 ACTAATAATAGTACAGTATGGCAATGGTAACATGGGAAATGATGACAATGACAATCAGCAGTGGCAAAAGTTAAATCAGATAGCAGTACAGTATAGTATGATAGCAGCAATAGC 1800
 T N N S D N G N N D N D N D K S D S G K G K S D S S S S D S S D S 560
 AGTATAGTATGATGACAGCAGTACAGTACAGCAGTATAGCAGTACAGCAGTATAGCAGCAGTACAGCAGTACAGCAGTACAGCAGTATAGTATGATAGCAGCAAT 1920
 S D S S D S S D S D S D S S D S N S S S D S D S S D S S D S S D S S D S S D S S D S S D S S D S S D S S D S S S N 600
 AGCAGTACAGTATGATGACAGCAGTATAGCAGTACAGTATGAT 2040
 S S D S S D S D S S D S S D S D S D S D S K S D S S S K S E S D S D S S S D S 640
 GACAGCAGTACAGCAACAGCAGTACAGTATGACAACAGTATGATGACAGCAGCAGCAATAGCAGTAAACAGCAGTATAGTATGATGATGATGATGATGATGATGATGATGAT 2160
 D S S D S N S S D S S D N S D S S D S S S N S S S N S S S D S S D S S D S S D S S D S S D S S S S 680
 AGTGACAGCAGCAGTACAGTATGACAGCAGCAACAGCAGTATAGTATGAT 2280
 S D S 720
 AGTAGTGACAGCAGTAATAGTAACAGCAGCGATAGTACAGCAGCAACAGCAGCGATAGCAGTACAGCAGTATAGCAGTACAGCAGCAACAGCAGTACAGTACAGCAGTAC 2400
 S S D S S N S N S S D S S S N S S D S S S S D S S D S S D S S S D S S S S D S S N S S S D S S D S S D S S 760
 AGCAGCAACAGCAGTACAGCAGTATAGCAGTACAGCAGTATAGTATGAT 2520
 S S N S S D S S D S S D S D S S D S S D S 800
 AGTAGTGACAGCAACAGCAGTATAGCAGTACAGCAGTATAGCAGTACAGCAGTATAGCAGTACAGCAGTATAGCAGTACAGCAGTATAGCAGTACAGCAGTATAGCAGTAC 2640
 S D S S N S S D S S D S S D S S D S S D S 840
 AGCAGCCACAGCAGTATAGCAGTACAGCAGTATAGCAGTACAGCAGTATAGCAGTACAGCAGTATAGCAGTACAGCAGTATAGCAGTACAGCAGTATAGCAGTACAGCAGT 2760
 S S D S S D S S D S S D S S D S S N R S D S S S S S S S S S S S D S S D S S D S S D S S D S S S D S S 880
 GATAGTATGACAGCAGTACAGCAGTAAAGAGCAATAGCAGTACAGCAGTATAGCAGCAGCAGTATAGTACAGCAGTATAGCAGCAGCAGTATAGCAGCAGTATAGCAGTATAG 2880
 T S 920
 AGCAACAGCAGTATAGCAGTAAAGCAGTAAATAGTATGACAACAGCAATAGCAGTACAGCAGCAACAGCAGTACAGCAGTATAGCAGTACAGCAGTATAGTATAGTATAGCAGCAGC 3000
 S N S S S D S S E S S N S S D N S N S S S D S S S S S S S S S S S S S S S S D S S S D S S S S S S S S S S 960
 AATAGCGGTGACAGCAACAGCAGTACAGCAGTATAGCAATAGCAGCAGCAGCAGCAGCAGCAACAGCAGCAGTATAGCAGTACAGCAGTATAGCAGTATAGCAGTATAGCAGT 3120
 N S G S 1000
 AGTATAGCAGCAACAGCAGTATAGCAGTACAGCAGTACAGCAGTATAGCAGTAAATAGTAGTACAGCAGCAACAGCAGTACAGCAGCGATAGCAGTACAGCAGCGATAGCAGT 3240
 S D S S N S S D S S D S S D S S D S S S S D S S N S S D S S S S S S S S D S S D S S D S S D S S S D S S S 1040
 GACAGCAGTACAGCAGCAATAGCAGTACAGCAGTACAGCAGCAGCAGCAGTATAGCAGTACAGCAGTACAGCAGCAGCGATAGCAGTACAGCAGTATAGCAGCGAT 3360
 D S 1080
 AGCAGTACAGCAGCAGCAGCAGTACAGCAGTACAGCAGTAAAGAGCAGCAGCAGCGATAGCAGCAGCAGTACAGCAGCGATAGCAGCAGCGATAGCAGCAGCGATAGCAGCAGCAGC 3480
 S S D S S D S S D S S D S S E S S D S S S S D S S S D S S D S S S D S S D S S D S S D S S S D S S D S S D S S D S S 1120
 AGCGACAGCAGCGATAGCAGTACAGCAGCAATAGCAGTATAGCAGCAGCAGCAGTATAGCAGTACAGCAGCAGCAGCGATAGCAGCAGCAGCAGTATAGTATGATAGCAGT 3600
 S D S 1160
 GACAGCAGTACAGCAGCAGCAGCAGTACAGCAGCGCAGCAGTACAGCAGCGCAGCAGTACAGCAATGAAAGAGCAGCAGCAGTACAGCAGCGATAGCAGTACAGCAGCAAC 3720
 D S S D S S D S S S D S S D S S D S S D S S S D S 1200
 AGCAGTACAGCAGCGCAGCAGTATAGCAGTACAGCAGCAGTACAGCAGCAGTACAGCAGCAGCGATAGCAGTACAGCAGCAGTACAGCAGCAGTACAGCAGCAGTAC 3840
 S 1240
 AGTGAAGGCAGTACAGTAACCACTCAACCGTGATGATTAGAACAAGAAAAACCATAAGATTCCTTTTGTGAAAAGTTGGTAAATGGGATAGGAAAAAGAAATTTCCCAAGAAAGTA 3960
 S E G S D S N H S T S D D * 1253
 AAGAAAAGGGGAGATCATAAGACGATGTAACAAAAACAACCTGGGGAAATCAAACTAAACAGTTGGATTAGACCAAGACCTAATCTCCAGAGACAGCAGTCTGAAATGACATGA 4080
 CCTTTGTACATGCTTAAATATCAGTCTGAAAAATTTGTTAAAGGTGAATCTAAACATAAAAGAACATCATAATTTCTTAAACTCTACACAGAA 4187

Fig. 2. The open reading frame (ORF) from the human DSPP gene and the deduced amino acid sequence. The putative signal peptide sequence is underlined by a single line. The putative N-terminal start of DSP is marked by a vertical arrow. The first three N-terminal amino acids of DPP is shown in bold and underlined by a broken line. The RGD sequence is underlined with a dotted line. The polyadenylation signals are underlined with a double line. The positions of the four introns are marked by the symbol of a diamond and the stop codon is indicated by an asterisk.

Table 1

Comparison of the deduced amino acid sequence from the hydrophobic signal peptide and the first 11 N-terminal sequences among human, rat and mouse DSP. The unmatched residues are underlined

	Signal peptide	N-terminus
Human Seq.:	MKII <u>Y</u> FCIWAVAWA	IPVPOSKPLER
Mouse DSP:	MKM <u>K</u> IIYICIWATAWA	IPVQLVPLER
Rat DSP:	M <u>K</u> T <u>K</u> IIYICIWATAWA	IPVQLVPLER

Human DSP cDNA sequence was obtained by several overlapping PCRs. The RNA PCR Core Kit (Perkin Elmer Cetus) was used. The concentrations of all reagents including the primers are per the manufacturer's protocol. First, PCR was performed using forward primer-1 (5'-GTA TTA CCA AAG AAT CTG GGA-3') and reverse primer-2 (5'-GCT ATC ACT GCT GTC ACT GTC-3') with the following program: 94°C for 30 s, 54°C for 30 s and 72°C for 2 min for 35 cycles. The above pair of primers produced a PCR fragment of approximately 700-bp. Next, a 850-bp PCR fragment was amplified with forward primer-3 (5'-GAG GAG AGA ACA GCA CAG CAA A-3') and reverse primer-4 (5'-CCC AAG ATC ATT CCA TGT TGT-3') following the program of 94°C for 30 s, 55°C for 30 s and 72°C for 2 min for 35 cycles. Lastly, a 5'RACE was performed using an abridged universal amplification primer (AUAP) or forward primer-5 (5'-GGC CAC GCG TCG ACT AGT AC-3') and reverse primer-6 (5'-TGA TGC TTA CTT GTC CCT G-3'), which amplified a 550- to 600-bp fragment, with the following program: 94°C for 30 s, 51°C for 30 s and 72°C for 2 min for 35 cycles. All the above PCR programs were finally followed by one cycle of 72°C for 10 min. An aliquot of the RT-PCR was resolved by gel electrophoresis (1% agarose) and the appropriately sized fragments cut from the gel. The DNA was purified (GeneClean Kit; Bio 101, La Jolla, CA, USA), subcloned into the vector pCR^{II-TOPO} (TA Cloning kit; Invitrogen, San Diego, CA, USA) and transformed into One ShotTM competent cells. The plasmids were isolated by Qiagen mini-prep (Chatsworth, CA, USA) and sequenced by a fluorescence-labeling technique (PRISMTM Ready Reaction, DyeDeoxyTM Terminator Cycle Sequencing Kit; Perkin Elmer-Applied Biosystems Division, Foster City, CA, USA).

Genomic clone, Southern hybridization and Northern hybridization

Genomic clone. A human genomic P1 library (Genome Systems, St. Louis, MO, USA) was screened using a 2.2-kb human DPP cDNA

Table 2

Comparison of the amino acid composition of the deduced putative human DSP with a sialoprotein extracted from immature human root tips. Values are a percent of total

Amino acid	natural protein	cDNA
Asp	19	19.5
Ser	14	10
His	2	3.1
Thr	6.5	5.3
Pro	6.5	3
Val	4.7	5
Cys	0	0.2
Lys	3.7	2.2
Leu	4.4	7.4
Glu	12	11.4
Gly	14	14.3
Arg	2.5	2.7
Ala	6	3.8
Tyr	0.7	0.6
Met	0.25	0.6
Ile	2.5	4.7
Phe	1.5	0.4
Try	0	0.4

probe (Fig. 1). A positive P1 genomic clone was obtained and digested with different restriction enzymes including BamHI, HindIII and XbaI. Several positive subclones (pZerOTM-2 plasmid; Invitrogen) were then identified and sequenced.

Southern hybridization. Briefly, up to 30 µg of human genomic DNA was digested separately by BamHI, EcoRI, HindIII, NotI and PstI, resolved by 0.8% agarose gel, and transferred onto a nitrocellulose membrane. The membrane was probed by a 2.2-kb human DPP cDNA and re-probed with a 380-bp DSP cDNA from exon 4 (Fig. 1). After overnight hybridization at 42°C in 50% formamide with the probe concentration of 10⁶cpm/ml hybridization solution and probe specific activity of 0.5–1 × 10⁹ cpm/µg (³²P), the membrane was successively washed with 2 × SSC, 0.5% SDS at room temperature for 15 min, 0.1 × SSC, 0.5% SDS at 37°C for 10 min to 30 min, and with 0.1 × SSC, 0.5% SDS at 68°C for 10 min to 30 min and exposed to X-ray film for 24–48 h.

Northern hybridization. RNazolTM B (TEL-TEST B) was used to isolate total RNA from human third molar teeth with open root apices. Twenty µg of total RNA was applied to the RNA gel and analyzed by Northern blotting probed separately by human DPP cDNA and DSP cDNA according to published procedure (18). The condition of Northern hybridization is following: pre-hybridization for 2 h, and hybridization at 42°C overnight. The hybridization buffer contained 5 × SSC, 2 × Denhardt's, 0.1% SDS and 40%

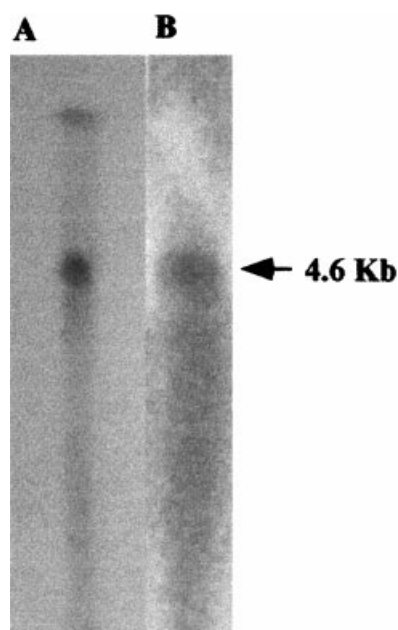


Fig. 3. Northern hybridization of human DSPP transcript expression in immature human third molar teeth. Twenty μg of human total RNA isolated from immature third molar teeth was analyzed. Panel A: ^{32}P -labeled human DPP as a probe. Panel B: ^{32}P -labeled human DSP as a probe. A single transcript of approximately 4.6 kb was recognized by both probes. Smearing was also evident below the major band in both hybridizations.

formamide, and the ^{32}P -labeled probes with a specific activity of 1×10^8 cpm/ μg . The membrane was washed with $2 \times \text{SSC}$, 0.1% SDS for 30 min at room temperature followed by $1 \times \text{SSC}$, 0.1% SDS for 20 min at 65°C .

Nested deletion of DPP subclone to facilitate sequencing in repetitive region

Erase-a-base system (Promega, Madison, WI, USA) was used to produce several new subclones with different lengths of DPP sequence to facilitate sequencing in the DSS repetitive region. The DPP-containing plasmids (pZERTM-2; Invitrogen) were digested by KpnI and SpeI, then subjected to digestion with Exonuclease III at 37°C from 30 s to 6 min at 30 s intervals. Finally, the plasmids were treated by S1 nuclease and re-ligated for transformation, isolation and subsequent sequencing.

Results

By using RT-PCR and 5'RACE cloning strategies, we successfully obtained a 1506-bp cDNA sequence, which was a 5' extension of human DPP cDNA (15). The primers were based on the rat DSP (17) and human DPP sequences (15) (Fig. 1) as well as a newly cloned partial human DSP

sequence. This newly cloned 1506-bp cDNA sequence contained a 1386-bp open reading frame (ORF), which preceded and was contiguous to the human DPP cDNA sequence (Fig. 2). The first 26 deduced amino acids of the N-terminus from this human cDNA sequence are nearly identical to the corresponding sequences from mouse and rat DSP (Table 1). These sequences contain a typical hydrophobic signal peptide and the first 11 amino acids of the N-terminus. In addition, the deduced amino acid composition of this cDNA is similar to that of a sialoprotein extracted from human root dentin (Table 2) (CHANG *et al.* report at the 1999 IADR meeting).

To confirm that a single gene encodes human DSP and DPP at the genomic level, a positive human genomic P1 clone (hDSPP-631P1) was identified by screening a human genomic P1 library with a 2.2-kb human DPP cDNA probe. This genomic clone carrying a 80- to 100-kb insert, was subsequently subcloned and sequenced. More than 10-kb DNA sequences were sequenced at the genomic level, which completely covered the whole region of human DSPP cDNA sequence determined by 3'RACE and 5'RACE methods. Four canonical splice-sites, defined by the bases 5'-GT and 3'-AG (19), were identified in the human DSPP gene, revealing the genomic organization of human DSP and DPP: a single gene containing 5 exons and 4 introns with the translational start codon ATG located in the second exon (Fig. 1). This newly cloned human gene contains a 3759-bp open reading frame coding for a protein with 1253 deduced amino acids (Fig. 2). The 5' portion of this open reading frame matched the human DSP cDNA sequence obtained by RT-PCR method, and the 3' portion contained the human DPP cDNA sequence (15).

Northern blot analyses performed with successive hybridizations by DPP and DSP probes, each revealed a prominent mRNA species of 4.6-kb (Fig. 3).

Human genomic DNA samples were digested with several restriction enzymes separately and analyzed by Southern blotting using low stringency and successive hybridization with the DPP and DSP probes (Fig. 4). The DPP probe detected a 15- to 20-kb BamHI fragment (Fig. 4A, lane 1), a HindIII fragment around 10 kb (Fig. 4A, lane 2), and an approximately 7-kb PstI fragment (Fig. 4A, lane 3) when 5–10 μg DNA was loaded into each lane. The DSP probe (Fig. 4B) detected a 5-kb BamHI fragment (lane 1), an approximately 10-kb HindIII fragment (lane 2), and an approximately 7-kb PstI fragment (lane 3). In contrast, when 20–30 μg of the BamHI digested DNA sample was loaded and hybridized with a DPP

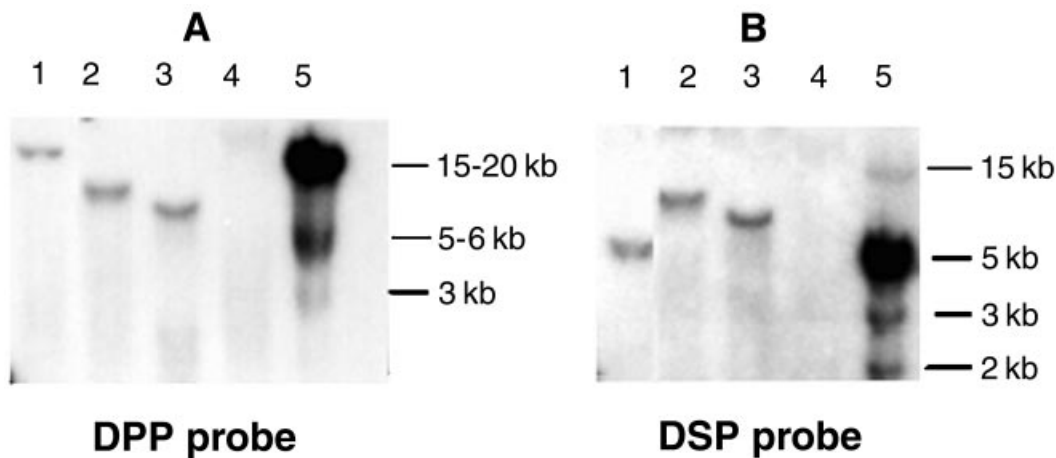


Fig. 4. Southern hybridization of human genomic DNA. DNA digests were probed by DPP (panel A) and DSP (panel B). Five – ten μ g of the genomic DNA sample was digested separately by BamHI, HindIII, PstI and NotI, and loaded in lane 1 to 4, respectively. Twenty – thirty μ g of the genomic DNA was digested by BamHI and loaded in lane 5.

probe (Fig. 4A, lane 5), two fragments of 5- to 6-kb and 3-kb in addition to the 15- to 20-kb major fragment were evident. Three additional fragments of approximate 15-kb, 3-kb and 2-kb were detected by the DSP probe (Fig. 4B, lane 5).

Discussion

The full coding sequence of human DPP was obtained using an RT-PCR cloning strategy (15). In order to determine the presence of a human dentin sialophosphoprotein (DSPP) gene, analogous to mouse and rat DSPP (3, 8–10), we have continued to analyze human immature root RNA by RT-PCR and 5'RACE methods. The primers were based on the rat DSP sequence (17), human DPP sequence (15), and a newly cloned partial human DSP sequence (Fig. 1). By overlapping 3 PCR cDNA sequences, we obtained a 1506-bp cDNA sequence with a 1386-bp ORF encoding for a protein rich in aspartic acid (19.5%), serine (10%), glutamic acid (11.4%) and glycine (14.3%). The extent of sequence similarity between this newly obtained human sequence and the mouse DSP sequence (10) is 69% at the DNA level (human, base pair #25–#1506 of Fig. 2 vs. mouse, base pair #1–#1438 of U67916) and only 52% at the amino acid level (human, amino acid residues #16–#462 of Fig. 2 vs. mouse, amino acid residues #18–#451 of U67916). All sequence comparisons were made using the Wisconsin Package Version 10.0, Genetic Computer Group (GCG) (Madison, WI, USA). However, the hydrophobic signal peptide and N-terminal 11 amino acids are nearly identical among human, mouse and rat (Table 1). Also, the deduced amino acid composition is similar to that of a sialoprotein extracted from human root dentin (Table 2). Although minor discrepancies exist in the amino

acid profile between the deduced amino acids and those from the dentin extracts, it could result from the difficulties in purifying the DSP in the dentin extract. Furthermore, this 1506-bp cDNA sequence is upstream of, and contiguous to, the human DPP cDNA. Therefore, we believe this newly cloned cDNA encodes for a human DSP, suggesting that, like rat and mouse, human DSP and DPP are translated from a single transcript.

DPP appears to be more conserved in terms of sequence similarity among human, mouse and rat. Particularly, the regions in the repetitive sequences among human, mouse and rat are highly conserved, mainly consisting of varying lengths of repeat sequence of DSS. In addition, the C-terminal last 19 amino acids are identical. The percent similarity between the C-terminal amino acid sequence of the human DPP (residues #992–#1253, Fig. 2) and the 245 published amino acid sequence of rat DMP2 (one form of rat dentin phosphoproteins) (16) is 93%. The percent similarity of the human and mouse (10) DPP coding region is 84% at the DNA level (human, base pair #2589–#3882 of Fig. 2, vs. mouse, base pair #1627–#2890 of U67916), and 81% at the amino acid level (residues #767–#1253 of Fig. 2, vs. mouse, residues #456–934 of U67916).

To confirm that a single gene encodes for human DSP and DPP at the genomic level, a putative human DSPP genomic clone was identified by screening a human genomic P1 library with a human DPP cDNA probe. This genomic clone (hDSPP-631P1) was subsequently subcloned and sequenced. The boundaries between the exons and introns were identified. As expected, like mouse and rat, human dentin sialoprotein and phosphophoryn are encoded by a single gene with five exons and four introns spanning at least 8.2-kb in the human genome when the poly A signal

ATTAAA is used (Fig. 1). In addition, the 3' portion of exon 2 together with exon 3, the 4 and 5' portion of exon 5 encodes for DSP and DSP-DPP linker region while the remainder of exon 5 encodes for DPP (Fig. 1). Furthermore, the lengths of the exons and the introns of human DSPP are similar to those of mouse (20). For the whole genomic sequence of human DSPP, please refer to GenBank, accession number: AF163151.

This newly cloned human DSPP gene contains a 3759-bp open reading frame and 1253 deduced amino acids (Fig. 2). As determined by 3'RACE (15), the relatively rare poly A signal ATTAAA is most frequently used in human DSPP gene, rather than the typical AATAAA sequence, which is found 177-bp upstream. In addition, the 4.6-kb mRNA transcript on the Northern blot analyses (Fig. 3) supports the use of this poly A signal. This 4.6-kb DSPP transcript was detected on the Northern blot analyses of total RNA extracted from immature (open root apices) human teeth using either a human DPP or DSP probe (Fig. 3). However, these Northern analyses failed to detect a 1.5-kb to 2.5-kb mRNA species reported in mouse (5, 10) and rat (17), although smearing below 4.6 kb was evident (Fig. 3). This difference between human and mouse or rat could be explained by the lack of alternative splicing during processing of the human DSPP transcript or the sensitivity of detection. Also, since the RNA analyzed was prepared from teeth from a single individual, it is possible that splice variants or genetic polymorphisms may be shown to exist if RNA from several individuals had been analyzed.

The nascent human DSPP is rich in aspartic acid (20%) and serine (41%) with an isoelectric point of 3.4. The deduced molecular weight of the DSPP core protein is 124.6-kDa. The DSP cDNA sequence, obtained by RT-PCR and 5'RACE methods, matches perfectly the DSPP gene; however, the published human DPP cDNA sequence (15) does not exactly match the DPP portion of the DSPP open reading frame. The deduced amino acid composition profile from the published human DPP cDNA sequence contains one extra cysteine, one more arginine and one more glutamic acid as well as six less serine residues when compared to the human DSPP open reading frame coding for DPP. Therefore, the DPP portion of this DSPP gene contains 791 deduced amino acids, 3 amino acids more than the previously reported cDNA sequence (15). Also, several amino acid mismatches exist. The discrepancies are dispersed exclusively within the highly repetitive region and could result from erroneous PCR amplification. Another possible explanation for these discrepancies is individual variation or genetic polymorphisms.

DPP and DSP isoforms have been reported in rat (21) (RITCHIE and WANG, report at the 1999 IADR meeting). Genomic Southern blot analysis could yield information regarding the multiple genes of DSPP. Therefore, low stringent Southern blot analyses of human genomic DNA using human DPP and DSP cDNA probes were performed successively. As can be seen (Fig. 4), when 5–10 μ g DNA was loaded on each lane, a single band was detected by either probe in each of these three (BamHI, HindIII and PstI) digests. These results are consistent with the other Southern blot analyses from the human genomic P1 clone (hDSPP-631P1) that carries the whole of the human DSPP gene and its flanking regions, digested by EcoRI, BamHI, HindIII, etc. (nine restriction enzymes) probed by the same DSP and DPP probes (data not shown). However, when the 20–30 μ g genomic DNA sample was digested with BamHI and analyzed, the DPP probe revealed two extra fragments (Fig. 4A, lane 5). Also, three fragments in addition to a 5-kb major fragment were detected by the DSP cDNA probe (Fig. 4B, lane 5). NotI digestion produced a very faint band, as detected by both probes (Fig. 4A, B, lane 4) which was probably due to the insufficient transfer of the large fragments produced by NotI digestion. The above data suggest that only one DSPP gene encodes for human DSP and DPP. The extra fragments from the BamHI digestion could be non-specific sequences detected when large amounts of DNA are analyzed by genomic Southern hybridization. It is also possible the human genome contains DSPP homologous sequences that could be members of a DSPP superfamily.

In summary, like mouse (10, 20) and rat (3), the human dentin non-collagenous proteins DSP and DPP are encoded by a single gene, DSPP. This human DSPP gene comprising 5 exons and 4 introns contains an open reading frame of 3759-bp encoding for a polypeptide with 1253 deduced amino acid residues. The genomic organization is quite similar to that described in mouse (20), indicating that the gene structure of DSPP is highly conserved during evolution. Current experiments are focused on elucidation of the function of human DSP and DPP.

Acknowledgements – This work was supported by NIH grant DE11442 (HR), DE 12899(BHC) and NIH grant DE 12466 (RBR).

References

1. BUTLER WT. Dentin matrix proteins. *Eur J Oral Sci* 1998; **106**: 204–210.
2. BUTLER WT, BHOWN M, DIMUZIO MT, LINDE A. Noncolla-

- genous proteins of dentin. Isolation and partial characterization of rat dentin proteins and proteoglycans using a three-step preparative method. *Collagen Rel Res* 1981; **1**: 187–199.
3. RITCHIE HH, RITCHIE DG, WANG LH. Six decades of dentinogenesis research. Historical and prospective views on phosphophoryn and dentin sialoprotein. *Eur J Oral Sci* 1998; **106**: 211–220.
 4. WEINSTOCK M, LEBLOND CP. Radioautographic visualization of the deposition of a phosphoprotein at the mineralization front in the dentin of the rat incisor. *J Cell Biol* 1973; **56**: 838–845.
 5. BEGUE-KIRN C, RUCH J-V, RIDALL AL, BUTLER WT. Comparative analysis of mouse DSP and DPP expression in odontoblasts, preameloblasts and experimentally induced odontoblast-like cells. *Eur J Oral Sci* 1998; **106**: 254–259.
 6. VEIS A, SFEIR C, WU CB. Phosphorylation of the proteins of the extracellular matrix of mineralized tissues by casein kinase-like activity. *Crit Rev Oral Biol Med* 1997; **8**: 360–379.
 7. VEIS A, WEI K, SFEIR C, GEORGE A, MALONE J. Properties of the (DSS)_n triplet repeat domain of rat dentin phosphophoryn. *Eur J Oral Sci* 1998; **106**: 234–238.
 8. RITCHIE HH, WANG L-H. A mammalian bicistronic transcript encoding two dentin-specific proteins. *Biochem Biophys Res Commun* 1997; **231**: 425–428.
 9. RITCHIE HH, WANG L-H. Sequence determination of an extremely acidic rat dentin phosphoprotein. *J Biol Chem* 1996; **271**: 21695–21698.
 10. MACDOUGALL M, SIMMONS D, LUAN X, NYDEGGER J, FENG J, GU T. Dentin phosphoprotein and dentin sialoprotein are cleavage products expressed from a single transcript coded by a gene on human chromosome 4. *J Biol Chem* 1997; **272**: 835–842.
 11. CHANG SR, CHIEGO D JR, CLARKSON BH. Characterization and identification of a human dentin phosphophoryn. *Calcif Tissue Int* 1996; **59**: 149–153.
 12. MCCURDY SP, CLARKSON BH, FEAGIN FF. Comparison of phosphoprotein isolated from mature and immature human tooth roots. *Arch Oral Biol* 1992; **37**: 1057–1065.
 13. TAKAGI Y, VEIS A. Matrix protein differences between human normal and dentinogenesis imperfecta dentin. In: VEIS A. *The chemistry and biology of mineralized connective tissue*. Vol. II. New York: Elsevier North-Holland, 1981; 223–243.
 14. TAKAGI Y, VEIS A, SAUK JJ. Relation of mineralization defects in collagen matrices to non-collagenous protein components: identification of a molecular defect in dentinogenesis imperfecta. *Clin Orthop* 1983; **176**: 282–290.
 15. GU K, CHANG SR, SLAVEN MS, CLARKSON BH, RUTHERFORD RB, RITCHIE HH. Human dentin phosphophoryn nucleotide and amino acid sequence. *Eur J Oral Sci* 1998; **106**: 1043–1047.
 16. GEORGE A, BANNON L, SABSAY B, DILLON JW, MALONE J, VEIS A, JENKINS NA, GILBERT DJ, COPELAND NG. The carboxyl-terminal domain of phosphophoryn contains unique extended triplet amino acid repeat sequences forming ordered carboxyl-phosphate interaction ridges that may be essential in the biomineralization process. *J Biol Chem* 1996; **271**: 32869–32873.
 17. RITCHIE HH, HOU H, VEIS A, BUTLER WT. Cloning and sequence determination of rat dentin sialoprotein, a novel dentin protein. *J Biol Chem* 1994; **269**: 3698–3702.
 18. SAMBROOK J, FRITSCH EF, MANIATIS T. *Molecular Cloning, 2nd edition*. New York: Cold Spring Harbor Laboratory Press, 1989; 7.37–7.52.
 19. LODISH H, BALTIMORE D, BERK A, ZIPURSKY SL, MATSUDAIRA P, DARNELL P. *Molecular cell biology, 3rd edition* New York: Scientific American Books, 1995; 500–501.
 20. FENG JQ, LUAN X, WALLACE J, JING D, OHSHIMA T, KULKARNI AB, D'SOUZA RN, KOZAK CA, MACDOUGALL M. Genomic organization, chromosomal mapping, and promoter analysis of the mouse dentin sialophosphoprotein (Dspp) gene, which codes for both dentin sialoprotein and dentin phosphoprotein. *J Biol Chem* 1998; **273**: 9457–9464.
 21. GEORGE A, SRINIVASAN R, THOTAKURA S, VEIS A. The phosphophoryn gene family: identical domain structures at the carboxyl end. *Eur J Oral Sci* 1998; **106**: 221–226.