Short communication

Human dentin phosphophoryn nucleotide and amino acid sequence

**Keni Gu, Syweren R. Chang, Matt
S. Slaven, Brian H. Clarkson,
R. Bruce Rutherford, Helena
H. Ritchie**

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

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.* © Eur J Oral Sci, 1998

Dentin sialoprotein (DSP) and phosphophoryns (DPP) are major dentin-specific
non-collagenous proteins and are synthesized by odontoblasts. DPP are
extremely acidic, rich in aspartic acid and serine, possess a high affinity for
calcium and collagen, and are believed to function in dentin mineralization.
Whereas DSP and DPP are the products of a single gene in mouse and rat, an
analogous human gene has not been described. Using RT-PCR based cloning
strategies, we have cloned human DPP cDNA from immature molar root total
RNA. The open reading frame of this human DPP cDNA comprises 2364 bp
encoding 788 amino acids rich in serine (58%), aspartic acid (26%) and
asparagine (9%). These are mostly arranged as $(DSS)_n$ ($n = 1$–16), DS and NSS
motifs. The N-terminal sequence (DDP) matches that obtained from human
DPP extracted from the roots of immature teeth. The core protein of this human
DPP was calculated to have a molecular weight of 76,906 Da and a net charge of
–206 with an isoelectric point of 2.65. Of the serine residues, 53% can potentially
be phosphorylated by casein kinases I and II. Thus, this newly cloned human
cDNA, which encodes a protein with characteristics similar to rat and mouse
DPP, is identified as a human DPP.

Bruce Rutherford, Department of Cariology,
Restorative Sciences and Endodontics, School
of Dentistry, University of Michigan, Ann Arbor,
MI 48109-1078, USA

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

Mammalian dentin matrices contain collagen (principally type I) and several non-collagenous proteins (NCP). Principal among these are several isoforms of phosphoprotein (DPP) and dentin sialoprotein (DSP) comprising approximately 50% and 5% of dentin NCP, respectively (1–3). Interestingly, these two molecules which are present in dentin as distinct proteins are encoded by a single gene in both rat (3–5) and mouse (6). This gene, referred to as dentinsialophosphoprotein (DSPP (6)), comprises at least 5 exons in both mouse (7) and rat (3) genomes. Exons 1–4 encode DSP, while exon 5 encodes the mouse/rat DSP C-terminus as well as the entire mouse DPP. The structure and function of these distinct and unique dentin proteins, which are proposed to play crucial roles in tooth morpho-

genesis and the mineralization of dentin matrices, have recently been reviewed (1, 3, 8).

Limited data are available on human DSP and DPP. Human immature roots contain a highly phosphorylated phosphoprotein with an apparent molecular weight of 140 kDa by SDS-PAGE (9); this cannot be detected in dentin from fully mature (closed root apices) human permanent teeth (10). The amino terminal sequence of this phosphoprotein is aspartic acid (D) -aspartic acid (D) -proline (P) (9). Additional sequencing has not been possible due to the instability of the phosphorylated peptide chain (CROSSLEY et al. report 1089 at the 1996 IADR meeting). DSPP has been localized to human chromosome 4q within the dentinogenesis imperfecta type II (DI-II) locus (11). Additionally DPP levels are reduced in individuals with DI-II (12, 13). Taken together such findings support a role for DPP in dentin mineralization in

---

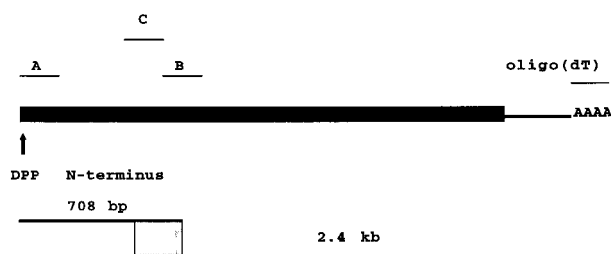* These authors contributed equally to this project.

Fig. 1. Cloning of human DPP cDNA from immature molar tooth root total RNA by RT-PCR. A 708 bp PCR product was generated from human odontoblast and dental pulp cDNA pool with forward primer A (5'caaggagacgatcccaac3') corresponding to the rat DSP link sequence QG and N-terminal DPP sequence DDPN and reverse primer B (5'gctactgtcgctactgtc3') corresponding to rat DSSDSS (5). A second PCR fragment with a size of 2.4 kb was generated by 3'RACE with an oligo(dT) primer (5'ggccacgcgtcgactagtact$_{17}$3'), primer C (5'acagcagcaaatcagagagc3') derived from human DPP and an abridged universal amplification reverse primer. The 708 bp and 2.4 kb PCR fragments were subcloned into TA vector pCR$^{II\text{-}TOPO}$ and DNA sequences were determined. The gray zone represents the 215 bp overlapped region of these two cDNA clones, which was confirmed by aligning several RT-PCR-derived clones which used primers based on the human sequence.

human teeth. Therefore, it is important to clone and sequence human DPP to further our understanding of its role in human dentinogenesis, dentin repair and remineralization, and in abnormalities such as dentinogenesis imperfecta type II.

Using reverse transcriptase-polymerase chain reaction- (RT-PCR) based cloning strategies, we have cloned human DPP cDNA and determined its sequence. Here we report the first nucleotide and deduced amino acid sequence for the human DPP.

Materials and methods

Isolation of total RNA

Root tips were removed from freshly extracted human immature teeth (apices approximately 1 mm or greater diameter) and crushed into small shards. Approximately 200 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 followed by phase separation. 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 and finally dissolved in DEPC-treated water and quantitated by UV absorbance spectrometry.

Primer design and RT-PCR

Human odontoblast and dental pulp cDNA pool was synthesized from total RNA obtained from the immature molar teeth root tips using random hexamers and reverse transcriptase according to the manufacturer's protocol (Perkin Elmer Cetus, Foster City, CA, USA). PCR was then performed using forward primer A (5'caaggagacgatcccaac3'), and reverse primer B (5'gctactgtcgctactgtc3'; Fig. 1), based upon rat DPP (5) with the following program: 94°C for 30 s, 51°C for 15 s and 72°C for 2 min for 35 cycles. To obtain a cDNA sequence containing the polyA region, 3'RACE (rapid amplification of cDNA ends; Gibco-BRL) was performed with AMV reverse transcriptase and an oligo(dT)-adaptor primer (5'ggccacgcgtcgactagtact$_{17}$3'; Gibco BRL,) to synthesize a cDNA pool, and followed by PCR with a forward primer C (5'acagcagcaaatcagagagc3'; Fig. 1) and an abridged universal amplification reverse primer (auap) with the following program: 94°C for 30 s, 54°C for 15 s and 72°C for 3 min for 35 cycles. The primer pairs A/B and the upper primer C (Fig. 1) for 3'RACE were prepared using Oligo^{TM} (Primer Analysis Software v.4, Wojciech Rychlik, Cincinnati, OH, USA).

Sequencing

An aliquot of the PCR mixture was resolved by agarose gel electrophoresis, and the appropriately sized bands were cut from the gel. The DNA was purified (Geneclean Kit; Bio 101, La Jolla, CA, USA), subcloned into the vector pCR$^{II\text{-}TOPO\ TM}$ and transformed into One Shot^{TM} competent cells (TA Cloning Kit; Invitrogen, San Diego, CA, USA). The plasmid was isolated by the Qiagen (Chatsworth, CA, USA) mini-prep protocol and sequenced by a fluorescence-labeling technique (PRISM^{TM} Ready Reaction DyeDeoxy^{TM} Terminator Cycle Sequencing Kit; Perkin Elmer-Applied Biosystems, Foster City, CA, USA).

Results and discussion

An RT-PCR based cloning strategy was selected to clone human DPP from total RNA prepared from immature root tips. Root tips were selected because this tissue provides a good natural source of undegraded human DPP (9). Using primer pair A/B (Fig. 1), a 708 bp PCR product was generated, subcloned and sequenced. The portions of this sequence that contained the rat primer sequences (primers A/B) were corrected by the construction of human gene specific primers delineating regions, which spanned the primers. This corrected 708 bp fragment contained a deduced N-terminal sequence (Fig. 2) which is identical to both the rat DPP N-terminal sequence
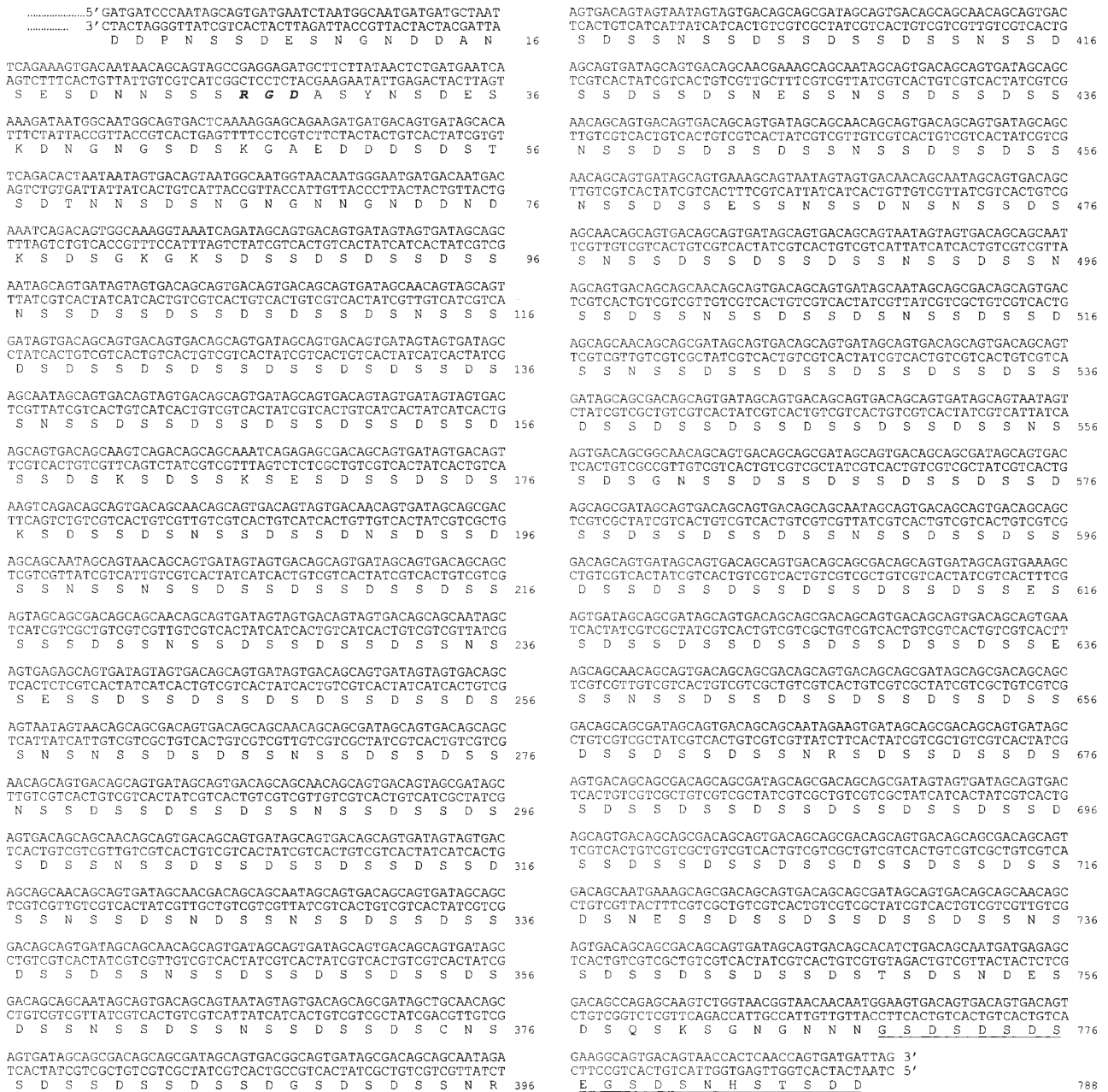
```
..........5'GATGATCCCAATAGCAGTGATGAATCTAATGGCAATGATGATGCTAAT
..........3'CTACTAGGGTTATCGTCACTACTTAGATTACCGTTACTACTACGATTA
          D  D  P  N  S  S  D  E  S  N  G  N  D  D  A  N   16

TCAGAAAGTGACAATAACAGCAGTGACCGAGGAGATGCTTCTTATAACTCTGATGAATCA
AGTCTTTCACTGTTATTGTCGTCATCGGCTCCTCTACGAAGAATATTGAGACTACTTAGT
 S  E  S  D  N  N  S  S  S  R  G  D  A  S  Y  N  S  D  E  S   36

AAAGATAATGGCAATGGCAGTGACTCAAAAGGAGCAGAAGATGATGACAGTGATAGCACA
TTTCTATTACCGTTACCGTCACTGAGTTTTCCTCGTCTTCTACTACTGTCACTATCGTGT
 K  D  N  G  N  G  S  D  S  K  G  A  E  D  D  S  D  S  T   56

TCAGACACTAATAATAGTGACAGTAATGGCAATGGTAACAATGGGAATGATGACAATGAC
AGTCTGTGATTATTATCACTGTCATTACCGTTACCATTGTTACCCTTACTACTGTTACTG
 S  D  T  N  N  S  D  S  N  G  N  G  N  N  G  N  D  D  N  D   76

AAATCAGACAGTGGCAAAGGTAAATCAGATAGCAGTGACAGTGATAGTAGTGATAGCAGC
TTTAGTCTGTCACCGTTTCCATTTAGTCTATCGTCACTGTCACTATCATCACTATCGTCG
 K  S  D  S  G  K  G  K  S  D  S  S  D  S  D  S  S  D  S  S   96

AATAGCAGTGATAGTAGTGACAGCAGTGACAGTGACAGCAGTGATAGCAACAGTAGCAGT
TTATCGTCACTATCATCACTGTCGTCACTGTCACTGTCGTCACTATCGTTGTCATCGTCA
 N  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D  S  N  S  S   116

GATAGTGACAGCAGTGACAGTGACAGCAGTGATAGCAGTGACAGTGATAGTAGTGATAGC
CTATCACTGTCGTCACTGTCACTGTCGTCACTATCGTCACTGTCACTATCATCACTATCG
 D  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D   136

AGCAATAGCAGTGACAGTAGTGACAGCAGTGATAGCAGTGACAGTAGTGATAGTAGTGAC
TCGTTATCGTCACTGTCATCACTGTCGTCACTATCGTCACTGTCATCACTATCATCACTG
 S  N  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D   156

AGCAGTGACAGCAAGTCAGACAGCAGCAAATCAGAGAGCGACGATAGTAGCGACAGT
TCGTCACTGTCGTTCAGTCTATCGTCGTTTAGTCTCGCTGTCGTCACTATCACTGTCA
 S  S  D  S  K  S  D  S  S  K  S  E  S  D  S  S  D  S  S   176

AAGTCAGACAGCAGTGACAGCAACAGCAGTGACAGTAGTGACAACAGTGATAGCAGCGAC
TTCAGTCTGTCGTCACTGTCGTTGTCGTCACTGTCATCACTGTTGTCACTATCGTCGCTG
 K  S  D  S  S  D  S  N  S  S  D  S  S  D  S  S  D   196

AGCAGCAATAGCAGTAACAGCAGTGATAGTAGTGACAGCAGTGATAGCAGTGACAGCAGC
TCGTCGTTATCGTCATTGTCGTCACTATCATCACTGTCGTCACTATCGTCACTGTCGTCG
 S  S  N  S  S  D  S  S  N  S  S  D  S  S  D  S  S  D  S  S   216

AGTAGCAGCGACAGCAGCAACAGCAGTGATAGTAGTGACAGCAGTGATAGCAGCAATAGC
TCATCGTCGCTGTCGTCGTTGTCGTCACTATCATCACTGTCGTCACTATCGTCGTTATCG
 S  S  S  D  S  S  N  S  S  D  S  S  D  S  S  D  S  S  N  S   236

AGTGAGAGCAGTGATAGTAGTGACAGCAGTGATAGTGACAGCAGTGATAGTAGTGACAGC
TCACTCTCGTCACTATCATCACTGTCGTCACTATCACTGTCGTCACTATCATCACTGTCG
 S  E  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D  S   256

AGTAATAGTAACAGCAGCGACAGTGACAGCAGCAACAGCAGCGATAGCAGTGACAGCAGC
TCATTATCATTGTCGTCGCTGTCACTGTCGTCGTTGTCGTCGCTATCGTCACTGTCGTCG
 S  N  S  N  S  S  D  S  S  D  S  S  N  S  S  D  S  S  D   276

AACAGCAGTGACAGCAGTGATAGCAGTGACAGCAGCAACAGCAGTGACAGTAGCGATAGC
TTGTCGTCACTGTCGTCACTATCGTCACTGTCGTCGTTGTCGTCACTGTCATCGCTATCG
 N  S  S  D  S  S  D  S  S  D  S  S  N  S  S  D  S  S  D  S   296

AGTGACAGCAGCAACAGCAGTGACAGCAGTGATAGCAGTGACAGCAGTGATAGTAGTGAC
TCACTGTCGTCGTTGTCGTCACTGTCGTCACTATCGTCACTGTCGTCACTATCATCACTG
 S  D  S  S  N  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D   316

AGCAGCAACAGCAGTGATAGCAACGACAGCAGCAATAGCAGTGACAGCAGTGATAGCAGC
TCGTCGTTGTCGTCACTATCGTTGCTGTCGTCGTTATCGTCACTGTCGTCACTATCGTCG
 S  S  N  S  S  D  N  S  S  D  S  S  N  S  S  D  S  S  D   336

GACAGCAGTGATAGCAGCAACAGCAGTGATAGCAGTGATAGCAGTGACAGCAGTGATAGC
CTGTCGTCACTATCGTCGTTGTCGTCACTATCGTCACTATCGTCACTGTCGTCACTATCG
 D  S  S  D  S  S  N  S  S  D  S  S  D  S  S  D  S  S  D  S   356

GACAGCAGCAATAGCAGTGACAGCAGTAATAGTAGTGACAGCAGCGATAGCTGCAACAGC
CTGTCGTCGTTATCGTCACTGTCGTCATTATCATCACTGTCGTCGCTATCGACGTTGTCG
 D  S  S  N  S  S  D  S  S  N  S  S  D  S  S  D  S  C  N  S   376

AGTGATAGCAGCGACAGCAGCGATAGCAGTGACGGCAGTGATAGCGACAGCAGCAATAGA
TCACTATCGTCGCTGTCGTCGCTATCGTCACTGCCGTCACTATCGCTGTCGTCGTTATCT
 S  D  S  S  D  S  S  D  G  S  D  S  S  D  S  S  N  R   396
```

```
AGTGACAGTAGTAATAGTAGTGACAGCAGCGATAGCAGTGACAGCAGCAACAGCAGTGAC
TCACTGTCATCATTATCATCACTGTCGTCGCTATCGTCACTGTCGTCGTTGTCGTCACTG
 S  D  S  S  N  S  S  D  S  S  D  S  S  D  S  S  N  S  S  D   416

AGCAGTGATAGCAGTGACAGCAACGAAAGCAGCAATAGCAGTGACAGCAGTGATAGCAGC
TCGTCACTATCGTCACTGTCGTTGCTTTCGTCGTTATCGTCACTGTCGTCACTATCGTCG
 S  S  D  S  S  D  S  N  E  S  S  N  S  S  D  S  S  D  S  S   436

AACAGCAGTGACAGTGACAGCAGTGATAGCAGCAACAGCAGTGACAGCAGTGATAGCAGC
TTGTCGTCACTGTCACTGTCGTCACTATCGTCGTTGTCGTCACTGTCGTCACTATCGTCG
 N  S  S  D  S  S  D  S  S  D  S  S  N  S  S  D  S  S  D  S   456

AACAGCAGTGATAGCAGTGAAAGCAGTAATAGTAGTGACAACAGCAATAGCAGTGACAGC
TTGTCGTCACTATCGTCACTTTCGTCATTATCATCACTGTTGTCGTTATCGTCACTGTCG
 N  S  S  D  S  S  D  S  S  D  S  S  N  S  S  D  S  S  N   476

AGCAACAGCAGTGACAGCAGTGATAGCAGTGACAGCAGTAATAGTAGTGACAGCAGCAAT
TCGTTGTCGTCACTGTCGTCACTATCGTCACTGTCGTCATTATCATCACTGTCGTCGTTA
 S  N  S  S  D  S  S  D  S  S  D  S  S  N  S  S  D  S  S  N   496

AGCAGTGACAGCAGCAACAGCAGTGACAGCAGTGATAGCAATAGCAGCGACAGCAGTGAC
TCGTCACTGTCGTCGTTGTCGTCACTGTCGTCACTATCGTTATCGTCGCTGTCGTCACTG
 S  S  D  S  S  D  S  S  D  S  S  D  S  N  S  S  D  S  S  D   516

AGCAGCAACAGCAGCGATAGCAGTGACAGCAGTGATAGCAGTGACAGCAGTGACAGCAGT
TCGTCGTTGTCGTCGCTATCGTCACTGTCGTCACTATCGTCACTGTCGTCACTGTCGTCA
 S  S  N  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D  S   536

GATAGCAGCGACAGCAGTGATAGCAGTGACAGCAGTGACAGCAGTGATAGCAGTAATAGT
CTATCGTCGCTGTCGTCACTATCGTCACTGTCGTCACTGTCGTCACTATCGTCATTATCA
 D  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D  S  S  N  S   556

AGTGACAGCGGCAACAGCAGTGACAGCAGCGATAGCAGTGACAGCAGCGATAGCAGTGAC
TCACTGTCGCCGTTGTCGTCACTGTCGTCGCTATCGTCACTGTCGTCGCTATCGTCACTG
 S  D  S  G  N  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D   576

AGCAGCGATAGCAGTGACAGCAGTGACAGCAGCAATAGCAGTGACAGCAGTGACAGCAGC
TCGTCGCTATCGTCACTGTCGTCACTGTCGTCGTTATCGTCACTGTCGTCACTGTCGTCG
 S  S  D  S  S  D  S  S  D  S  S  N  S  S  D  S  S  D  S  S   596

GACAGCAGTGATAGCAGTGACAGCAGTGACAGCAGCGACAGCAGTGATAGCAGTGAAAGC
CTGTCGTCACTATCGTCACTGTCGTCACTGTCGTCGCTGTCGTCACTATCGTCACTTTCG
 D  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D  S  S  E  S   616

AGTGATAGCAGCGATAGCAGTGACAGCAGCGACAGCAGTGACAGTGACAGCAGCAATGAA
TCACTATCGTCGCTATCGTCACTGTCGTCGCTGTCGTCACTGTCGTCGTTGTCGTTACTT
 S  D  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D  S  S  E   636

AGCAGCAACAGCAGTGACAGCAGCGACAGCAGTGACAGCAGCGATAGCAGCGACAGCAGC
TCGTCGTTGTCGTCACTGTCGTCGCTGTCGTCACTGTCGTCGCTATCGTCGCTGTCGTCG
 S  S  N  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D  S  S   656

GACAGCAGCGATAGCAGTGACAGCAGCAATAGAAGTGATAGCAGCGACAGCAGTGATAGC
CTGTCGTCGCTATCGTCACTGTCGTCGTTATCTTCACTATCGTCGCTGTCGTCACTATCG
 D  S  S  D  S  S  D  S  S  N  R  S  D  S  S  D  S  S  D   676

AGTGACAGCAGCGACAGCAGCGATAGCAGCGACAGCAGCGATAGTAGTGATAGCAGTGAC
TCACTGTCGTCGCTGTCGTCGCTATCGTCGCTGTCGTCGCTATCATCACTATCGTCACTG
 S  D  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D   696

AGCAGTGACAGCAGCGACAGCAGTGACAGCAGCGACAGCAGTGACAGCAGCGACAGCAGT
TCGTCACTGTCGTCGCTGTCGTCACTGTCGTCGCTGTCGTCACTGTCGTCGCTGTCGTCA
 S  S  D  S  S  D  S  S  D  S  S  D  S  S  D  S  S  D  S   716

GACAGCAATGAAAGCAGCGACAGCAGTGACAGCAGCGATAGCAGTGACAGCAGCAACAGC
CTGTCGTTACTTTCGTCGCTGTCGTCACTGTCGTCGCTATCGTCACTGTCGTCGTTGTCG
 D  S  N  E  S  S  D  S  S  D  S  S  D  S  S  D  S  S  N  S   736

AGTGACAGCAGCGACAGCAGTGATAGCAGTGACAGCACATCTGACAGCAATGATGAGAGC
TCACTGTCGTCGCTGTCGTCACTATCGTCACTGTCGTGTAGACTGTCGTTACTACTCTCG
 S  D  S  S  D  S  S  D  S  S  D  S  T  S  D  S  N  D  E  S   756

GACAGCCAGAGCAAGTCTGGTAACGGTAACAACAATGGAAGTGACAGTGACAGTGACAGT
CTGTCGGTCGTCGTTCAGACCATTGCCATTGTTGTTACCTTCACTGTCACTGTCACTGTCA
 D  S  Q  S  K  S  G  N  G  N  N  N  G  S  D  S  D  S  D  S   776

GAAGGCAGTGACAGTAACCACTCAACCAGTGATGATTAG 3'
CTTCCGTCACTGTCATTGGTGAGTTGGTCACTACTAATC 5'
 E  G  S  D  S  N  H  S  T  S  D  D                      788
```

Fig. 2. Human DPP nucleotide sequence and deduced amino acid sequence. The human DPP DNA sequence was principally derived from two cDNA clones which overlap 215 bp (Fig. 1). The open reading frame spans 2364 bp encoding 788 amino acids with 461 serine and 205 aspartic residues. These residues are mostly arranged as 151 DSS and 32 DS. The deduced DPP amino acid sequence has an N-terminal sequence DDPN. *RGD* at amino acid position 26, a putative cell attachment sequence, is bold and italicized. The conserved 19 C-terminal amino acids are underlined. (Accession number AF 09450).

(DDPNSSDESNG) (5) and to the human DPP N-terminal sequence, DDP (9).

Using primer C (Fig. 1), derived from the human 708 bp DNA sequence, and an oligo(dT)-adaptor primer to perform RT-PCR, we obtained a 2.4 kb PCR fragment (Fig. 1) which was subcloned and sequenced. These two clones overlapped by 215 bp. Additional PCR primers based upon the corrected human clones and the vector primers were synthesized to assist in sequencing the highly repetitive regions of DPP (Fig. 2).

The open reading frame of human DPP was found to span 2364 bp encoding 788 amino acids (Fig. 2). It was rich in serine (461 residues, 58.6%) and aspartic acid (205 residues, 26%). These residues were found to be arranged as 151 $(DSS)_n$ ($n = 1$–16), 32 (DS) and 37 (NSS) groups. The carboxyl-terminal 19 amino acids (Fig. 2) were

Table 1.

Amino acid composition of deduced and native human DPP

Amino acid	Human[a] Res No.	Human[a] %	Human[b,c] %	Human[b,d] %
Asp	205	26	36	41
Asn	68	9		
Ser	461	58	38	42
Glu	13	2	5	5
Gly	16	2	5	4.6
Lys	9	1.1	1.3	2.1
His	1	0.1	1.4	0.4
Tyr	1	0.1	0.44	0.4
Thr	4	0.5	2	1.1
Pro	1	0.1	5	0.6
Ala	3	0.4	0.68	1.3
Cys	1	0.1		0
Ile	0	0	0	0.3

[a] Amino acid composition derived from deduced protein sequence; %, amino acid composition represented in percent.
[b] Amino acid composition derived from native protein.
[c] TAKAGI & VEIS (15).
[d] MCCURDY et al. (28).

identical to those published earlier for rat (5, 14) and mouse (6) DPP. Subsequently, the complete open reading frame (Fig. 2) was cloned using an up-stream and a downstream primer specific to this human gene. Partial sequencing confirmed the molecular cloning of a human DPP. In aggregate, these data strongly support designation of this molecule as a human DPP.

According to the present data, this human DPP has 299 (788–489) more amino acids than mouse DPP (6), 548 (788–240) more than rat DPP as reported by RITCHIE & WANG (5), and 573 (788–215) more than the DPP region of rat DMP-3 (14). Like mouse (6) but unlike rat (5), human DPP contains an RGD (Arg-Gly-Asp) sequence, start-ing at position 26 (Fig. 2).

Hydropathy distribution analysis revealed that the protein is extremely hydrophilic (data not shown). The net charge of the core protein (before phosphorylation) was calculated to be –206 with an isoelectric point of 2.65. The molecular weight for the core protein is 76,906 Da. The deduced amino acid percentages obtained from our cDNA and those obtained from native proteins by conven-tional amino acid analysis (10, 15) are compared in Table 1. Our human DPP preparation as reported (9) was likely contaminated by collagen. Increased purity of such preparations has been associated with a relative increase in aspartic acid and serine residues, concomitant with a decrease in glycine and proline (Chang, unpublished data). These observations and the possibility of multiple DPP molecules (1) could account for the apparent discrepancies between the deduced amino acid

composition data and those obtained from extracted DPP (Table 1).

DPP is rich in phosphoserine (1), and serine residues are likely to be phosphorylated post-translationally by protein kinase(s). Therefore, we examined the potential phosphorylation sites of our deduced human DPP sequence and found consensus sequences for casein kinase I [Asp/Glu/pSer-X-X-Ser/Thr (16–18)] and casein kinase II [Ser/Thr-X-X-Asp/Glu/pSer (16, 19–21)]. Overall, 53% of the serine residues could potentially be phosphorylated by casein kinases I or II. With this number of phosphoserines, DPP could carry an additional charge of –395 from phosphate groups alone [-1.6/Ser(P) or Thr(P)]. Therefore, DPP theoretically could carry an overall net charge of –601 at physiological pH. In addition, using Asn-X-Ser/Thr as consensus N-glycosylation sites (22, 23), we found 47 potential sites which, if glycoslyated, might further increase the net charge and the molecular mass.

It is widely accepted that DPP has an essential role in the regulation of dentin mineralization by virtue of its amino acid composition and repetitive arrangement of aspartic acid (D) and serine (S) residues (1, 3, 24). The presence of a similar composition and arrangement of D and S residues in human DPP (Fig. 2) strongly supports a putative role for this protein in human dentin mineraliza-tion. In addition, the presence of a cell-binding RGD sequence near the N-terminus in mouse (6) and human deduced DPP (Fig. 2) sequences suggests that human DPP may attach cells expres-sing the appropriate integrins. Since DPP is

expressed in pre-ameloblasts during odontogenesis (25), this might possibly be related to the cellular differentiation processes occurring at this stage. In contrast, however, no RGD sequence was found in the deduced rat DPP sequence (5, 8), and experiments are clearly needed to investigate the function of this RGD sequence in human and mouse DPP.

Preliminary data (K. Gu, unpublished) imply that the human DPP mRNA reported here is a 3′ extension of an open reading frame encoding for human DSP and the DSP-PP linker region, and suggest that in humans, like rats and mice, a single gene encodes DSP and DPP. Additionally, DPP isoforms (1, 26, 27) as well as multiple DPP transcripts have been identified in rat (1, 3, 8). Our preliminary data also imply the presence of multiple human DPP molecules (K. Gu, unpublished).

Current experiments are focusing on sequencing the human DSP region, isolating and characterizing DSP from human dentin, and determining the patterns of expression of these molecules during tooth formation and reparative dentinogenesis.

References

1. BUTLER WT. Dentin matrix proteins. *Eur J Oral Sci* 1998; **106**: 204–210.
2. BUTLER WT, BHOWN M, DIMUZIO MT, LINDE A. Noncollagenous 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. RITCHIE HH, WANG L-H. A mammalian bicistronic transcript encoding two dentin-specific proteins. *Biochem Biophys Res Commun* 1997; **231**: 425–428.
5. RITCHIE HH, WANG L-H. Sequence determination of an extremely acidic rat dentin phosphoprotein. *J Biol Chem* 1996; **271**: 21,695–21,698.
6. 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.
7. 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.
8. 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.
9. CHANG SR, CHIEGO D, CLARKSON BH. Characterization and identification of a human dentin phosphophoryn. *Calcif Tissue Int* 1996; **59**: 149–153.
10. MCCURDY SP, CLARKSON BH, FEAGIN FF. Comparison of phosphoprotein isolated from mature and immature human tooth roots. *Arch Oral Biol* 1992; **37**: 1057–1065.
11. MACDOUGALL M. Refined mapping of the human dentin sialophosphoprotein (DSPP) gene within the critical dentinogenesis imperfecta type II and dentin dysplasia type II loci. *Eur J Oral Sci* 1998; **106**: 227–323.
12. TAKAGI Y, VEIS A. Matrix protein differences between human normal and dentinogenesis imperfecta dentin, VEIS A. The chemistry and biology of mineralized connective tissue, Vol. II. New York: Elsevier North-Holland, 1981; 223–243.
13. 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.
14. 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**: 32,869–32,873.
15. TAKAGI Y, VEIS A. Isolation of phosphophoryn from human dentin organic matrix. *Calcif Tissue Int* 1984; **36**: 259–265.
16. PINNA L. Casein kinase 2: an "eminence grise" in cellular regulation. *Biochem Biopys Acta* 1990; **1054**: 267–284.
17. FLOTOW H, GRAVES P, WANG A, FIOL C, ROESKE R, ROACH P. Phosphate groups as substrate determinants for casein kinase I action. *J Biol Chem* 1990; **265**: 14,264–14,269.
18. FLOTOW H, ROACH P. Role of acidic residues as substrates determinants for casein kinase I. *J Biol Chem* 1991; **266**: 3724–3727.
19. MEGGIO F, MARCHIORI F, BORIN G, CHESSA G, PINNA L. Synthetic peptides including acidic clusters as substrates and inhibitors of rat liver casein kinase TS (type-2). *J Biol Chem* 1984; **259**: 14,576–14,579.
20. KUENZEL E, MULLIGAN J, SUMMERCORN J, KREBS E. Substrate specificity determinants for casein kinase II as deduced from studies with synthetic peptides. *J Biol Chem* 1987; **262**: 9136–9140.
21. MEGGIO F, PERICH J, MEYER H, HOFFMANN-POSORSKE E, LENNON D, JOHNS R, PINNA L. Synthetic fragments of beta-casein as model substrates for liver and mammary gland casein kinases. *Eur J Biochem* 1989; **186**: 459–464.
22. MARSHALL RD. Glycoproteins. *Ann Rev Biochem* 1972; **41**: 673–702.
23. BAUSE E. Structural requirements on N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem J* 1983; **209**: 331–336.
24. 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.
25. BÈGUE-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.
26. LINDE A, BHOWN M, BUTLER WT. Noncollagenous proteins of dentin. A re-examination of proteins from rat incisor dentin utilizing techniques to avoid artifacts. *J Biol Chem* 1980; **255**: 5931–5942.
27. BUTLER WT, BHOWN M, DIMUZIO MT, COTHRAN WC, LINDE A. Multiple forms of rat dentin phosphoproteins. *Arch Biochem Biophys* 1983; **225**: 178–186.
28. MCCURDY SP, CLARKSON BH, SPEIRS RL, FEAGIN FF. Phosphoprotein extraction from the dentine/cementum complex of human tooth roots. *Arch Oral Biol* 1990; **35**: 347–357.