

Characterization of porcine dentin sialoprotein (DSP) and dentin sialophosphoprotein (DSPP) cDNA clones

Yasuo Yamakoshi¹, Jan C.-C. Hu¹,
Shengxi Liu², Chuhua Zhang²,
Shinichiro Oida³, Makoto Fukae³,
James P. Simmer¹

¹University of Michigan Dental Research Laboratory, Ann Arbor, MI, USA; ²University of Texas School of Dentistry, Health Science Center at San Antonio, Department of Physiology, TX, USA; ³Department of Biochemistry, School of Dental Medicine, Tsurumi University, Yokohama, Japan

Yamakoshi Y, Hu JC-C, Liu S, Zhang CH, Oida S, Fukae M, Simmer JP. Characterization of porcine dentin sialoprotein (DSP) and dentin sialophosphoprotein (DSPP) cDNA clones. *Eur J Oral Sci* 2003; 111: 60–67. © Eur J Oral Sci, 2003

Dentin sialophosphoprotein (DSPP) is a chimeric glycoprotein with dentin sialoprotein (DSP) on its *N*-terminus and dentin phosphoprotein (DPP) on its *C*-terminus. We have constructed and screened a unidirectional cDNA library derived from the pulp organ of developing pig teeth, and isolated cDNA clones encoding DSP-only, as well as two DSPP clones with alternative sequences in their 3' coding regions. The DSP-only transcript has an open reading frame of 386 codons, and is generated through the use of a polyadenylation signal within intron 4, immediately following the DSP coding region. The use of this polyadenylation signal deletes the DPP coding region and places a TGA translation termination signal as the fourth codon following the exon 4-encoded segment. The DSPP cDNAs contain open reading frames of 593 and 600 codons. Northern blots hybridized to radiolabeled DSP probes showed bands at 1.4, 2.5, 4.4, and 4.8 kb. Cloning and characterization of reverse transcriptase polymerase chain reaction products confirmed the existence of mRNA encoding pDSP₃₈₆, pDSPP₅₉₃, and pDSPP₆₀₀ *in vivo*, but also suggested that DNA sequence redundancies in the DSPP coding region make it prone to cloning artifacts.

James P. Simmer, University of Michigan Dental Research Laboratory, 1210 Eisenhower Place, Ann Arbor, MI 48108, USA

Telefax: +1-734-9759329
E-mail: jsimmer@umich.edu

Key words: dentin sialophosphoprotein (DSPP); dentin sialoprotein (DSP); dentin phosphoprotein (DPP); odontogenesis; porcine

Accepted for publication November 2002

Collagen and non-collagenous proteins are required for proper biomineralization of dentin. Collagen is the predominant protein in dentin and constitutes about 90% of the organic matrix (1). The major non-collagenous proteins of dentin are dentin sialoprotein (DSP) and dentin phosphoprotein (DPP) (2–5), which are encoded by the dentin sialophosphoprotein (DSPP) gene (6, 7). Non-collagenous proteins are believed to associate at specific sites on the collagen molecule to promote the nucleation and growth of hydroxyapatite crystals (8). This hypothesis is consistent with recent studies of inherited malformations of dentin. Defects in the human DSPP gene (9, 10) and a Gly559Cys mutation in the human COL1A1 gene (11) were associated with dentinogenesis imperfecta (DGI), while characterization of dentin affected by DGI showed a significant reduction of intra-fibrillar (within the gap in collagen fibrils) dentin mineralization (12). These studies clearly show the importance of dentin sialophosphoprotein in both normal and diseased dentin formation.

Much of what we know about DSPP comes from the studies performed on rats. Rat dentin sialoprotein (DSP), the *N*-terminal portion of DSPP, comprises about 5–8% (by weight) of the dentin non-collagenous proteins and has a molecular mass of 95 kDa on sodium dodecyl-sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) (13). The *N*-terminal sequence of rat DSP was determined

by Edman degradation (IPVQLV), and its molecular mass was determined by sedimentation equilibrium analysis to be 52.57 kDa, which included 29.6% carbohydrate (14). The DSP in rat dentin is derived from the *N*-terminus of DSPP, with the major component terminating with Tyr⁴²¹ and a minor component ending with His⁴⁰⁶ (15), which appear to be generated by proteolytic cleavage of DSPP. The calculated molecular mass (excluding post-translational modifications) of the 421 amino acid rat DSP is 43.12 kDa, leaving only 9.45 kDa (or 18% of its mass) to be contributed by post-translational modifications; this suggests that the molecular mass of rat DSP may be higher, or its percentage carbohydrate may be lower than published values.

Dentin phosphoprotein (DPP), the *C*-terminal portion of DSPP, is the most abundant non-collagenous protein in rat dentin (16), and is remarkable for its extremely acidic isoelectric point (pH 1.1) (17). Rat DPP seems to be composed of a group of related, stains-all positive proteins with different levels of phosphorylation, apparent molecular weights around 95 kDa, and an *N*-terminal sequence of DDPN (5, 18). The DDPN sequence is part of the deduced rat DSPP amino acid sequence, indicating that DPP is generated by proteolytic cleavage of DSPP after Gly⁴³⁰. Bovine DPP has the *N*-terminal sequence DSPNSSDESNGDSPNSSDESN-GDSNDSSS DSDSSDSSDSSDSDSDS (with every

serine in this sequence being phosphorylated) and migrates as a single band on SDS-PAGE at 156 kDa (19).

Better characterization of the DSPP protein and its proteolytic cleavage products is needed to understand the functions of DSPP, DSP, and DPP in normal and diseased dentin formation. It is possible that an additional layer of complexity at the protein level results from the expression of multiple mRNA transcripts from the DSPP gene. The rat DSPP gene (20) and a growing number of rat cDNAs derived from the rat DSPP gene have been cloned and characterized (6, 20–26). In this study we report the isolation and characterization of three novel pig DSPP-derived cDNAs from a pig pulp-derived λ cDNA library, and the use of Northern and reverse transcriptase polymerase chain reaction (RT-PCR) analyses to investigate the expression of DSP and DSPP in developing pig teeth.

Material and methods

All experimental procedures involving the use of animals were reviewed and approved by the Institutional Animal Care and Use Program at the University of Texas Health Science Center at San Antonio.

RNA isolation from pig tissues

Developing teeth were dissected from 6-month-old commercial pigs. Enamel organ epithelia and pulp tissue were removed and the hollow pulp chamber, lined with residual odontoblasts, was extracted with RNAwiz RNA Isolation Reagent (Ambion, Austin, TX, USA), extracted with chloroform, precipitated with isopropanol, treated with DNase I (Invitrogen, Carlsbad, CA, USA), and then quantified by UV spectroscopy at 260 nm.

Isolation of pig DSP and DSPP cDNA from a λ cDNA library

A unidirectional cDNA library was constructed from RNA extracted from the pulp organ of developing pig teeth (Stratagene, La Jolla, CA, USA), and screened using a DSP-specific cDNA probe. To obtain the cDNA probe, degenerate primers were synthesized to regions of protein conservation in an alignment of mouse, rat, and human DSP amino acid sequences. The 5' primer (5'-GCATGGG CCATTCRGTTCCTCA-3') was made to the amino acid sequence AWAIPVPQ and the 3' primer (5'-TGTAAYC-CATCATGACCCYTATGTTTC-3') was made to the amino acid sequence ETYGHGDIH. These primers amplified, by RT-PCR, a 332-bp segment of the pig DSP coding region, which was cloned into the TA cloning vector (Invitrogen), and designated as clone 138. A pig DSP-specific cDNA probe was generated by random priming of a gel-purified *EcoRI* restriction fragment from clone 138 and used to screen the cDNA library. This screening resulted in the isolation of clone #29, encoding only the DSP portion of DSPP. A second screening was performed using a DSP-specific cDNA probe generated by random priming of a gel-purified 600 bp *EcoRI* restriction fragment from 5' end of clone #29, which resulted in the isolation of a single DSPP cDNA (clone #17). A third screening was performed using a probe made by random priming a gel-purified 1.2 kb PCR

product that had been amplified using a DPP-specific primer pair (5'-CATTCCAGTTCCTCAAATCAAG-3', and 5'-GCTTCATAAACTAACCTTATCTT-3') and clone #17 as template. This resulted in the isolation of DSPP clones #4 and #12.

Northern blot analysis

A radiolabeled cDNA probe was synthesized by random priming a gel-purified 1.45 kb *KpnI/SacI* restriction fragment containing the entire DSP cDNA insert of clone #29. DNA-free total RNA (15 μ g) was fractionated by agarose-formaldehyde gel electrophoresis, transferred to a nylon membrane by upper capillary transfer and hybridized to the radiolabeled probe using the ultrahyb, ultrasensitive hybridization kit and protocol (Ambion). The hybridized membrane was autoradiographed overnight using Kodak (Rochester, NY, USA) BioMax MR-1 film (Sigma, St Louis, MO, USA).

Reverse transcription-polymerase chain reaction

Reverse transcription was performed using SuperScript First-Strand Synthesis System (Invitrogen) followed by PCR using *Taq* DNA Polymerase (Invitrogen) and assorted pig DSPP-specific primer pairs. A negative control for each primer pair included an amplification of the RNA sample (without reverse transcription), which checked for template contamination of the primers as well as DNA contamination of the RNA preparation. Selected primer pairs were used to amplify clone #29 (pDSPP3F, 5'-TTCTGCACTC CCATGTTAG-3', and pDSP4F, 5'-CCGTAGCCAAAG GATAGAGG-3', paired with pDSP29R, 5'-GGGAAGG AAGGGGAGAATTT-3'), clone 12 (pDSPP4F, 5'-AG TGATGGCAATGGTGACAA-3', paired with pDSPP12R, 5'-AAGGGAGCCTCAGAATTTCC-3'), and clones #4 and #17 (pDSPP4F paired with pDSPP17R, 5'-TTGCAA TATTGGTCTCTTGCAT-3').

Sequence analyses

Database searches were performed at the National Center for Biotechnology Information website (<http://www.ncbi.nlm.nih.gov/blast/blast.cgi?Jform=1>) using BLAST (27). The pig DSP (AF332578) nucleotide and derived amino acid sequences were aligned with those of human DSP (AF163151), mouse DSP (U67916), and rat DSP (NM_012790) using CLUSTALW 1.8 (28), available at the Baylor College of Medicine website (<http://dot.imgen.bcm.tmc.edu:9331/multialign/multialign.html>). The molecular weights and isoelectric points were predicted using the Compute pI/Mw tool (http://c.expasy.org/tools/pi_tool.html) on the Expasy molecular biology server (<http://expasy.cbr.nrc.ca/tools/scnpsit1.html>) at the Swiss Institute of Bioinformatics. Searches for potential phosphorylation, and O-linked and N-linked glycosylation sites were performed at the Center for Biological Sequence Analysis at the Technical University of Denmark Website (<http://www.cbs.dtu.dk/>) using the NetNGlyc 1.0, NetOGlyc 2.0 (29), and NetPhos 2.0 (30) prediction servers.

Results

A consistent shorthand for designating the various transcripts and proteins expressed from the DSPP gene is

currently lacking. In this report, we merge two systems that are in use. The first letter, in lower case, indicates the organism (r, rat; m, mouse, p, pig; h, human). Adding DSP indicates only the DSP domain is present; DSPP and DSP-PP indicate that both the DSP and DPP domains are present. The number of amino acids, in subscript, follows these designations. Thus, the designation rDSPP₉₇₀/rDSP-PP₅₂₃ refers to the rat DSPP protein having 970 amino acids, inclusive of a DPP domain of 523 amino acids, extending from the DDPN sequence to the C-terminus of the protein.

Using a radiolabeled probe to identify transcripts generated from the DSPP gene, four cDNA clones were isolated from a porcine pulp λ cDNA library (#29, #12, #17 and #4). The four cDNA clones shared identical DSP coding sequences, indicating that they were all derived from a single gene. Although the porcine DSPP gene has not been cloned and characterized, we assume that it shares the same intron/exon structure that is common to the human (Acc No. AF163151), mouse (Acc. No. AF135799, AJ002141), and rat (Acc. No. RNO278306) DSPP genes (20, 24, 31, 32). The shortest porcine cDNA clone (Acc. No. AF332578) was 1344 nucleotides long (excluding the polyA tail), and encoded only the DSP domain (pDSP₃₈₆) (Fig. 1A). The pig DSP cDNA extended 83 nucleotides into exon 1, which is non-coding. The coding region starts in exon 2 at the most 5'-ATG in the cDNA, which imitates an open reading frame of 386 codons. There is no possibility that a longer cDNA clone could extend the reading frame to a more 5'-ATG codon, as there is an in-frame TGA translation stop codon starting 21 bases (7 codons) prior to the assigned translation initiation site.

The first 15 amino acids at the start of the deduced pig DSP amino acid sequence (MKIIYFCIWAIAWA) contain a positively charged 'n-region' followed by a hydrophobic 'h-region', which constitute a signal peptide (33). The next five amino acids (IPVPQ) are identical to the Edman sequence of rat DSP (14) and represent the N-terminus of pig DSPP. The DSP protein secreted has 371 amino acids, a predicted molecular mass of 38.74 kDa, and an isoelectric point of 4.35. The protein is expected to be larger because of two post-translational modifications. There are seven Asn residues in pig DSP (amino acids 67, 97, 107, 166, 185, 191, 206, and 345) that are in an appropriate context (NX (S/T), where X is any amino acid except proline) for N-linked glycosylation, but only two (Asn⁹⁷ and Asn¹⁶⁶) are 'very likely' to be glycosylated by computer prediction, while Asn⁹⁷ is in the only glycosylation context that is conserved among all of the known (rat, mouse, human, and pig) DSP sequences (Fig. 1B). There are two potential O-linked glycosylation sites at Thr²¹⁴ and Thr²¹⁵. A potential integrin-binding site (RGD) is found in the pig DSP sequence, but is not conserved in the rodent and human sequences.

The structure of the 3' end of the pig DSP cDNA reveals how this short transcript was generated by the DSPP gene. Nucleotide alignment of the pig DSP cDNA with the mouse and human DSPP genes reveals that the 3' end of the pig DSP cDNA is homologous to the

beginning of intron 4 in the human DSPP gene (Fig. 2). Translation terminates at the fourth codon following the end of exon 4. The 3' untranslated region (UTR) is short, but contains a distinctive polyadenylation/cleavage signal (AATAAA) that ends 15 nucleotides prior to the poly (A+) tail (Fig. 1A). Therefore, the pig DSP₃₈₆ mRNA transcript is generated by transcription termination in intron 4, followed by the splicing out of introns 1, 2 and 3. The pig and human DSPP genes both contain an in-frame stop codon followed by a transcription termination signal in intron 4, while the mouse and rat DSPP genes do not. As a consequence, the mechanism that yields the DSP-only cDNA in pig could theoretically also yield a DSP only transcript in humans, but not in rodents.

To test the hypothesis that the pDSP₃₈₆ transcript is not a cloning artifact, but in fact corresponds to an mRNA naturally present in odontogenic tissues, we isolated RNA from cells still lining the pulp chamber of developing pig teeth following removal of the dental pulp. Previous studies have shown that odontoblasts typically stay attached to dentin when the pulp is removed (34). RNA isolated from the tissue clinging to the sides of the pulp chamber was used for Northern blot and RT-PCR analyses. The Northern blot analysis showed bands at 4.8, 4.4, 2.6, and 1.4 kb when hybridized to the radiolabeled pDSP₃₈₆ cDNA probe (Fig. 3A). The smallest band, at 1.4 kb, corresponds in size to the pDSP₃₈₆ cDNA. This band is prominent, which suggests that the pDSP₃₈₆ transcript may be an important source of DSP protein *in vivo*. A RT-PCR using a primer pair that anneals to exon 3 and the 3' non-coding region in what amounts to intron 4 in larger DSPP transcripts sequences found on the pDSP₃₈₆ clone showed three bands (Fig. 3B). The upper and lower bands were cloned and characterized by DNA sequencing. The larger band, at 925 bp was identical to the sequence from pDSP₃₈₆. The smallest band was 500 bp and was identical to the pDSP₃₈₆ clone with a single deletion of 425 bp. The borders of the deleted segment do not obey the GU...AG rule for splice junctions, suggesting that the band is an artifact generated during the RT-PCR procedure. Experimental controls rule out the possibility that the PCR amplification products were amplified from contaminating cDNA or genomic DNA, as no amplification product was observed when reverse transcriptase was omitted from the RT step, and the primer annealing sites spanned intron 3 so that genomic template would have generated a larger amplification product.

The second round of library screening yielded a comparatively short DSPP clone (#17, Acc. No AY161863). The cDNA insert is 2533 bp, excluding the polyA tail. The clone contains a polyadenylation signal (AATAAA) 22 nucleotides upstream of the polyA tail. The deduced amino acid sequence is 600 amino acids (pDSPP₆₀₀/pDSP-PP₁₂₈). The putative porcine DPP domain begins with Asp⁴⁷³, at the start of the conserved DDPN sequence. The equivalent starts of the rat and human DPP domains are Asp⁴⁴⁸ in rat (rDSPP₉₇₀/rDSP-PP₅₂₃), and Asp⁴⁶³ in human (hDSPP1253/hDSP-PP₇₉₁).

A

```

GGCCACTCTCAGTCTTCAAAGAGAAAGATAAGAAAATCCTGGATTTTCAAAGTCCCTTTGAAGGCTTTTAAGAAC 75
AGCCGCTGATTAATATTTCTAAAGAAAATGAAGATAATATATATATTTTGCATTTGGGGCAATAGCATGGGCCATT 150
          M K I I I Y F C I W A I A W A / I
CCAGTTCCTCAAATCAAGCCATTGGAGAGGCATGCCATTGACAAAAGTGTGAACCTTAAATCTTCTAGCAAATCG 225
P V P Q I K P L E R H A I D K S V N L N L L A K S 41
AAAGCGCCAGTACAGGATGAGTTAAATGCCAATGATACCACAAAGGAAGTGGCATCCCCATGCATGACCATGAC 300
K A P V Q D E L N A N D T T K G S G I P M H D H D 66
ATAGGAAGGCAGCAGGATACCAAAGATGGTTACAAAGGAGAGAGAAATGGCTCTGAGTGGGCAGACGTGGGAGGG 375
I G R Q Q D T K D G Y K G E R N G S E W A D V G G 91
AACAGTTCCTTGCACGCTCCCATGTTAGCAAACAAGAGAGAACACGGAGGATCCGAATGGGGACGCAGGACAA 450
N S S S A R P M L A N K E E N T E D P N G D A G Q 116
CCAGAACCATATGGCCACGATGGGCTCCATGGACGAGGGGACAGCAGCCCCGCAAATGGCTCAGAGGGCAAGTC 525
P E P Y G H D G L H G R G D T S S P A N G L R G Q V 141
AGCATTCTTGACAACACTGGCACAGCAAATGGGAGCCATGTTAATGGGGTACTGATAAGAATTCAAAAAACGAG 600
S I L D N T G T A N G S H V N G V T D K N S K N E 166
GATGTTGGAAATGCAAGTCAGAGTGAAGATGCCACTGTGGTCCCGGAAGATAGATATCAAGTAGCTGGAAGCAAT 675
D V G N A S Q R I E N A T V V P E D R Y Q V A G A N 191
AACAGTATAGGTCATGAAGATGAAATAAATGGGAATTTCTGTAGAAATGGGGAGATGTAAGTGAACAACACCT 750
N S I G H E D E I N G N F C R N G G D V S E T T P 216
CCAGGAGAAGGTGAGATAAATGGGAATGAAGAGACAGGGGTAACATCTGGGGAAAGTGGAGCTGGCAATAGAGAA 825
P G E G E I N G N E A T G V T S G G S G A G N R E 241
TTTGCTGGCCTGGATAATCTGATGGCAGTCCCAGTGGGAATGGAGCAGATGAAGAGGAAGACAAGGGCTCTGGG 900
F A G L D N S D G S P S G N G A D E E E D K G S G 266
GACGATGAAGGTGAAGAACGGGGAATGGAGAGGACTGCTGATACCAGCAAGGCCAGGAGAATCCGAGTGCAT 975
D D E G E E T G N G E R T A D T S K G Q E N P S H 291
GGAGAGGAGGAGCGGAGGAGGAGGACGACACCATAGCTTAGTCAAAATTCATAGTAGTGAAGATGAGGGC 1050
G E E A E E D D D H S L G Q N S I S S E D E G 316
CCTGGGCACAAAGAAGCTGCCACTGACGAGAGACAACCTCCAAGAGTGAAGGAAATCTGACAAATATT 1125
P G H K E A A H A T D G D N T S K S E E D S D N I 341
CCAGGAGAAAGCCGTAGCCAAAGGATAGAGGACCCAGAAACCAATCAGAGAGAAACCAAAGCTGTGGCAAAT 1200
P G E S R S Q R I E D T Q K P N Q R E T K A V A N 366
GGAGTCACTGCAATATCAGAGCCCTTGCTATTGGGAAGAGCCAAGATAAGGTTAGTTTATGAAGCTGATTTCTT 1275
G V T A I S E P L A I G K S Q D K V S L * 386
TCAAAGCGGTTTAAATCTCCCTTCCCTGGTGGTAGCATGAAAATAAAATCATGACAAACAGTCAAAAA 1350

```

B

```

pDSP --MKIIIFYFCIWAIAWA IPVPQIKPLERHAIDKSVNLNLLAKSKAPVQDELNANDTTKGSIPMHDDHIGRQDQTKDGY 62
hDSP --MKIIIFYFCIWAIAWA IPVPQSKPLERH-VEKSNMLHLLARSNVSVQDELNANSGTIKESGLVHHEGDRGRQENTQDGH 62
mDSP MKMKIIIFYCIWAIAWA IPVPQLVPLERDIVENSVAVPLLTHPGTAAQNELINSSTSNNSNDSPDGSEIGEQLVSEdGY 62
rDSP MKTKIIIFYCIWAIAWA IPVPQLVPLERDIVEKSDVPPLAHPGTAAQNELHINNAT--NDDSPKGSSELGRQVHNSNGY 60

pDSP KGERNGSEWADVGGNSSARPMLANKEENT-EDPNGDAGQPEPYGHDLHGRGDSSPANGLRGQVSI LDNTGTANGSHVN141
hDSP KGENGSKWAEVGGKSFSTYSTLANEENI-EGWNGDTGKAETVYHDIHGKEENITANGIQQVSIIDNAGATNRENTN141
mDSP KRDNGESIHHVGGKDFPTQPILVNEQGNTEAEEHNDI----ETYGHdGVHARGENSTANGIRSQVGIENAEBA-ESSVH137
rDSP ERDRNGSESIAVGGKSSPTQPILANAQNSAKEREDV----ETYGHdGIHAGGENSTANGIRGQVGIENAEBAEKESKVH136

pDSP GVTDKNSKNEDVGNASQSENATVVPEDRYQVAGSNNSIGHEDIENGNFCRNGGDVSETTPPGEGEINGNEETGVT-SGGS220
hDSP GNTDKNTQNGDVGDAHNEDVAVVQEDGPQVAGSNSTDNEDIENSRCRNEGNTSEITPQINSKRNGTKEAEVTPGTGE221
mDSP GQAGQNTKSGGASDVSNQGDATLVQENEPPEASIKNSTNHEAGIH----GSGVATHETTPQREGLSENQGTVEVTPSIGE213
rDSP GQPQDQTKTGLASDTSQNGDATLVQENEPQVAGSKNSTNHEVGT----GSGVAAQETTPQREGESESNQCAEVTPSIGE212

pDSP GAGNREFAGLDNSDGSPPSGNGADEEEDKSGDDEGEETNGERTADTSKQENPSHGEAEEDDDHSLGQNSISSEDE300
hDSP DA-----GLDNSDGSPPSGNGADEEEDKSGDDEGEETNGERTADTSKQENPSHGEAEEDDDHSLGQNSISSEDE300
mDSP DA-----GLDNTDGSPPSGNGVEEEDDTGSGDGEAGADGRESHDGTGQGGQSHGNTDHRGQSSVSTEDDSDKE---284
rDSP GA-----GLDNTDGSPPSGNGIEEEDDTGSGDGVGADAGDRESHDGTGEGHGGQSSGNNDRGQSSVSTEDDSDKE---284

pDSP PGGHKEAAHATDGNSTKSEEDSDNIPGESRSQRIEDTQKPNQRETKAVANGVTAISEPLAIGKSQDKVSL 371
hDSP PEGKEDPHNEVDGDKTSKSEENSAGIPEDNGSQRIEDTQKLNHRESKRVENRITKESETHAVGKSQDK 360
mDSP QEGFPNGHN---GDNSS--EEN--GVEEGDSTQATQDKERLSPKDRDAEGGIIISQSEACPSSGKSQDQ 345
rDSP QEGSPNRRG---GDNTSSSEET--GIEEGDGTQTTQDNQNLSP-----TEGGIIISQAEACPSGQSQDQ 341

```

Fig. 1. Porcine dentin sialoprotein (DSP) cDNA and derived protein sequence and alignment to known homologs. The number of the last nucleotide or amino acid in each row is provided on the right. Part A shows the pDSP₃₈₆ nucleotide and deduced amino acid sequences. A slash marks the end of the signal sequence. The translation initiation and termination codons are in bold type. The polyadenylation signal is double underlined. The primer annealing sites for pDSP3F (in exon 3) and pDSP29R (in intron 4) are underlined. Part B shows the pDSP₃₈₆ amino acid sequence aligned to its human (31), mouse (7) and rat (21) homologs. The eight potential N-linked glycosylation sites in pDSP₃₈₆ are underlined.

The third round of library screening yielded a clone (#4) with an identical coding region to pDSPP₆₀₀ isolated in the previous round of screening, and a second novel pig DSPP cDNA (#12, Acc. No. AY161862), 1962 bp long (excluding the polyA tail), with a polyadenylation signal (AATAAA) 20 nucleotides upstream of the polyA tail. The deduced amino acid sequence is 593 amino acids (pDSPP₅₉₃/pDSP-PP₁₂₁). The two pig DSPP deduced amino acid sequences

(pDSPP₆₀₀ and pDSPP₅₉₃) are identical for their first 567 amino acids and then diverge. The two clones have different 3'-non-coding regions and have different polyadenylation signals, suggesting that they are derived from pDSPP transcripts that terminated using different polyadenylation/cleavage sites, and were then differentially spliced at their 3' ends. The two pig DSPP clones appear to represent low abundance transcripts that make up part of a relatively minor

```

pDSP  A I S E P L A I G K S Q D K      V S L *
pDSP  GCAATATCAGAGCCCTTCTGCTATTGGGAAGAGCCAAGATAAG // GTTAGTTTATGAAGCTGA
hDSP  AAAGAATCAGAGACACATGCTGTGGGAAGAGCCAAGATAAG // GTTAGTTTGTAAAGCTGA
mDSP  AGCCAGTCAGAAGCATGTCTTCTGGGAAGAGCCAAGATCAG // GTAAGTTTAGAGGGCGGC
rDSP  AGCCAGGCAGAAGCATGTCTTCTGGGCAGAGCCAAAATCAG // AACTTACATCCTTCCTTC

pDSP  TTTCTTTCAAAGCGGTTTAAATTTCTCCCTTCCCTTCCCTTGGTGGTAGCATGAAAAATAAATCATGACAACAGTC
hDSP  TTTCTTTCAATGGCAGTTTAAATTTCTCCCTTCCCTTCCCTTGGTGGTAGCATGAAAAATAAATCATGACAACAGTC
mDSP  GACTTCCATTCTTCCCTCCATCTGTGATGGCTGTACCAAATAACTCCAGACAAACACGAGAGATAAAACCCCAA
rDSP  CAACTCTGGTGGCTGTAGCAAAAATCCAGACAAACACAAGATAAAACCCCAAACAGGCATAAAAGTCCACACTAA

```

Fig. 2. Alignment of the 3' end of the pDSP386 cDNA with the human and mouse dentin sialophosphoprotein (DSPP) genomic sequences. The 3' ends of the exon 4 segments are indicated by a double slash. In-frame translation stop codons are underlined, as are the polyadenylation signals. Note that the rodent sequences do not contain the necessary polyadenylation signal to generate a short dentin sialoprotein (DSP) transcript by this mechanism.

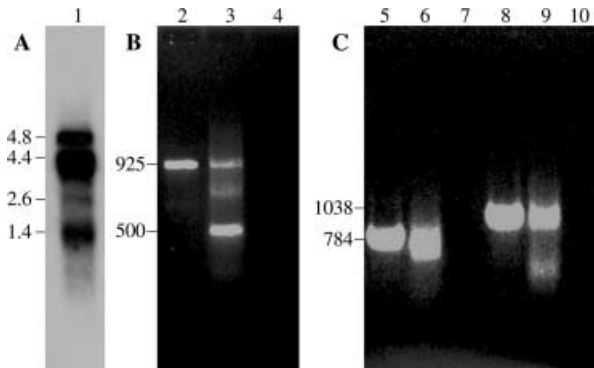


Fig. 3. Northern blot and reverse transcriptase polymerase chain reaction (RT-PCR) analyses of mRNA isolated from odontoblasts lining the pulp chamber of developing teeth after removal of the pulp tissue. Lane 1: autoradiograph of a Northern blot hybridized to a radiolabeled dentin sialoprotein (DSP) probe; the sizes of the four bands identified (in kb) are shown on the left. Lanes 2–4: PCR products using oligonucleotides pDSP3F and pDSP29R using pDSP₃₈₆ cDNA as template (lane 2), using odontoblast mRNA and reverse transcriptase in the RT reaction (lane 3), and using odontoblast mRNA with no reverse transcriptase in the RT reaction (lane 4). The 925 bp product was identical to the pDSP₃₈₆ cDNA sequence, while the 500 bp sequence in lane 3 showed a deletion of 425 bp that was not bracketed by appropriate splice junctions and was presumably generated during reverse transcription. Lanes 5–7: PCR products using oligonucleotides pDSP4F and pDSP17R using pDSP₆₀₀ cDNA as template (lane 5), using odontoblast mRNA and reverse transcriptase in the RT reaction (lane 6), and using odontoblast mRNA with no reverse transcriptase in the RT reaction (lane 7). The 784 bp product was identical to that of the pDSP₆₀₀ cDNA sequence. Lanes 8–10: PCR products using oligonucleotides pDSP4F and pDSP12R using pDSP₅₉₃ cDNA as template (lane 8), using odontoblast mRNA and reverse transcriptase in the RT reaction (lane 9), and using odontoblast mRNA with no reverse transcriptase in the RT reaction (lane 10). The 1038 bp product was identical to that of the pDSP₅₉₃ cDNA sequence.

band at 2.5 kb on the Northern blot (Fig. 3A). The RT-PCR reactions using oligonucleotides annealing to the DSP coding region paired with oligonucleotides annealing to the unique 3'-non-coding regions of cDNAs encoding pDSP₆₀₀ and pDSP₅₉₃ gave products identical in size and sequence to the pDSP cDNA clones (Fig. 3C). The major pig DSPP transcripts, that measure 4.4 kb and 4.8 kb on the Northern blot have not yet been cloned.

Discussion

The pig has long been an important animal model for the characterization of enamel proteins (35–39), but less has been done to characterize proteins in pig dentin. As dentin and enamel formation are intimately related, characterization of both dentin and enamel matrix proteins should lead to a better understanding of tooth formation in general and the dentino-enamel junction in particular, where both dentin and enamel proteins are most closely associated. In this study we have isolated three novel cDNA clones derived from the pig DSPP gene. The pDSP₃₈₆ was generated through the use of a polyadenylation signal in intron 4. Its length matches that of a prominent DSP band on Northern blots and has been successfully amplified by RT-PCR from pig odontogenic tissues. These findings provide strong evidence that the pDSP₃₈₆ transcript is present *in vivo*. Two novel DSPP cDNA clones (pDSP₆₀₀ and pDSP₅₉₃) were also characterized that are smaller than the major DSPP RNA messages observed by Northern blot analyses. These clones were apparently generated by alternative splicing, but this cannot be confirmed at this time because of the absence of pig DSPP genomic sequences. These clones are not likely to be cloning artifacts as they were regenerated in RT-PCR analyses, and two independently isolated pDSP₆₀₀ clones were characterized.

The pig DSPP-derived cDNA clones show significant differences from DSPP-derived clones characterized from other organisms. In mouse, large (4.4 kb) and small (1.4 kb) DSPP transcripts are observed on Northern blots, but only the 4.4 kb transcript has been cloned; this encodes a protein with 932 amino acids (mDSP₉₃₂, Acc. Nos. NM_010080 and AF135799) (7). In humans, the deduced DSPP sequence has 1253 amino acids (hDSP₁₂₅₃/hDSP-PP₇₉₁, Acc# NM_014208) (40). In rat, multiple DSPP-derived cDNA clones have been cloned and characterized (20,22–26,41). The structures of cDNAs derived from the rat DSPP gene are shown in Fig. 4.

The largest rat DSPP cDNA characterized (4457 bp) encodes a DSPP protein with 970 amino acids, 523 of which are in the DPP portion of the protein (rDSP₉₇₀/rDSP-PP₅₂₃, Acc. No. AJ403971 20). The C-terminal portion of rDSP₉₇₀ corresponds to the deduced DSPP amino acid sequence originally thought to be a product of a second gene, and designated dentin matrix protein-2 (DMP2 23). The rat DSPP genomic sequence (Acc. No.

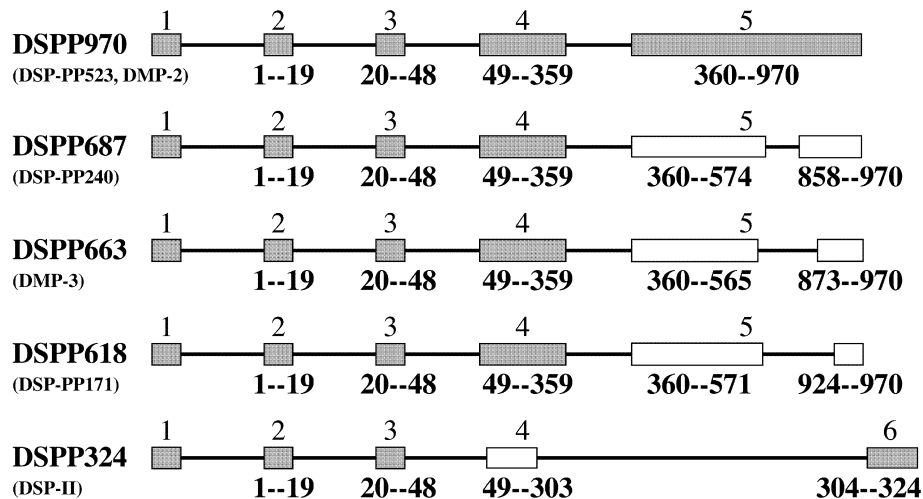


Fig. 4. Splicing diagram of the rat DSPP-derived cDNA clones. The dentin sialophosphoprotein (DSPP) designations for each cDNA clone, showing the number of amino acids in each derived protein sequence are shown on the left of each diagram. Other aliases for the clones are provided below these designations in smaller type size. Exons are indicated as rectangles, with the number of each exon shown above, and range of codons in the exon shown below each rectangle. Complete exons are indicated by shaded rectangles, interrupted exons are indicated by open rectangles.

AF251219) (20), shows that the mRNA encoding rDSPP₉₇₀ forms by splicing together five exons, four of which are coding. All of the DPP sequence is encoded in the last exon (exon 5). The rDSPP₉₇₀ is analogous to the mDSPP₉₃₂, and hDSPP₁₂₅₃.

Three shorter rat DSPP sequences exhibit unique deletions of DPP coding region from the middle of exon 5. Merging the original rat DSP (Acc. No. RNU02074) (21), and DPP (Acc. No. RNU63111) (6), cDNA sequences (using the rat DSPP genomic sequence to provide the linking-region sequence) produces a cDNA encoding a protein of 687 amino acids with a DPP region of 240 amino acids (rDSPP₆₈₇/rDSP-PP₂₄₀). The second rat DSPP sequence that displayed a deletion in the DPP region was designated dentin matrix protein-3 (DMP3) (24). This DSPP sequence has 663 amino acids, with a DPP region of 216 amino acids (rDSPP₆₆₃/rDSP-PP₂₁₆). The third rat DSPP sequence with a deletion in exon 5 encoded a DSPP protein with 618 amino acids (rDSPP₆₁₈/rDSP-PP₁₇₁) (25).

A rat cDNA with no DPP sequence at all, designated DSP-II, was reported (26). This cDNA (rDSP₃₂₄) shows a deletion that starts in exon 3, includes all of exons 4 and 5, and ends with a sequence sometimes found at the end of the 3' non-coding region of DSPP cDNA. Several factors tend to weaken the significance of this finding. The deleted segment does not have appropriate splice junctions at its ends, suggesting that it may be a cloning artifact. In contrast, RNase protection assay appeared to confirm the presence of the rDSP₃₂₄ transcript (26). Even if the rDSP₃₂₄ is real, it is likely to account for only a minor fraction of the DSP protein in rat dentin. Northern blots hybridized with a radiolabeled DSP probe did not detect the 1050 nucleotide rDSP₃₂₄ transcript (21). The exon 6-encoded sequence, which represents a distinctive 21 amino acid polypeptide on the DSP C-terminus, was not detected in a study that

characterized C-terminal sequences of DSP isolated from rat teeth (15).

With the exception of the rDSPP₉₇₀, all of the rat DSPP cDNAs are assumed to be generated by alternative RNA splicing. Contrary to this assumption, alignment of the cDNA sequences to the rat DSPP gene shows that *none* of the deleted segments have appropriate splice junctions on the ends of their proposed introns (GU...AG). The GU...AG rule is considered to be inviolate (42,43), and it is especially rare when both ends of the putative intron have non-canonical sequences (44). When the alignment of cDNA and genomic sequences identifies a putative intron sequence that lacks appropriate splice junctions, the clone is considered to be the result of imperfect cDNA construction, usually a reverse transcription artifact (45). We have characterized several different PCR amplification products that show deletions in the pig DSPP coding region, which are almost certainly artifacts. With the exception of rDSPP₉₇₀, we suggest that the four shorter cDNAs derived from the rat DSPP gene (rDSPP₆₈₇/rDSP-PP₂₄₀, rDSPP₆₆₃/rDSP-PP₂₁₆, rDSPP₆₁₈/rDSP-PP₁₇₁, and rDSP₃₂₄) could be cloning artifacts, as no known RNA splicing mechanism could theoretically have generated them from what is reported to be the rat DSPP gene sequence. Another possibility is that they represent low-abundance transcripts generated by rare splicing events that do not obey the GU...AG rule.

Dentin sialophosphoprotein is now known to be critical for healthy tooth (and possibly bone) formation. Dentin sialoprotein, in pig and probably in human, can be expressed alone, or joined to dentin phosphoprotein. It can be separated from the DSPP protein by proteolytic cleavage. At present, no mRNA transcripts have been cloned that carry only the DPP coding region, and the higher abundance of DPP relative to DSP in rat dentin remains unexplained.

Multiple cDNA clones have now been characterized that show deletions of various lengths from the DPP coding region. The validity and significance of these cDNAs remains to be demonstrated, as the mechanism by which they could have been generated from the DSPP gene is unclear.

Acknowledgements – This investigation was supported in part by USPHS Research Grant R01-DE12769, and Program Project DE13237 from the National Institute of Dental and Craniofacial Research, NIH.

References

1. LINDE A. Dentin matrix proteins: composition and possible functions in. *Anat Rec* 1989; **224**: 154–166.
2. VEIS A, PERRY A. The phosphoprotein of the dentin matrix. *Biochemistry* 1967; **6**: 2409–2416.
3. LEAVER AG, TRIFFITT JT, HOLBROOK IB. Newer knowledge of non-collagenous protein in dentin and cortical bone matrix. *Clin Orthop*, 1975: 269–292.
4. DIMUZIO MT, VEIS A. Phosphophoryns – major noncollagenous proteins of rat incisor dentin. *Calcif Tissue Res* 1978; **25**: 169–178.
5. 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.
6. RITCHIE HH, WANG LH. Sequence determination of an extremely acidic rat dentin phosphoprotein. *J Biol Chem* 1996; **271**: 21695–21698.
7. MACDOUGALL M, SIMMONS D, LUAN X, NYDEGGER J, FENG J, GU TT. Dentin phosphoprotein and dentin sialoprotein are cleavage products expressed from a single transcript coded by a gene on human chromosome 4. Dentin phosphoprotein DNA sequence determination. *J Biol Chem* 1997; **272**: 835–842.
8. TRAUB W, JODAIKIN A, ARAD T, VEIS A, SABSAY B. Dentin phosphophoryn binding to collagen fibrils. *Matrix* 1992; **12**: 197–201.
9. ZHANG X, ZHAO J, LI C, GAO S, QIU C, LIU P, WU G, QIANG B, LO WH, SHEN Y. DSPP mutation in dentinogenesis imperfecta Shields type II. *Nature Genet* 2001; **27**: 151–152.
10. XIAO S, YU C, CHOU X, YUAN W, WANG Y, BU L, FU G, QIAN M, YANG J, SHI Y, HU L, HAN B, WANG Z, HUANG W, LIU J, CHEN Z, ZHAO G, KONG X. Dentinogenesis imperfecta 1 with or without progressive hearing loss is associated with distinct mutations in DSPP. [erratum appears in *Nat Genet* 2001 March; **27**(3): 345]. *Nature Genet* 2001; **27**: 201–204.
11. PALLOS D, HART PS, CORTELLI JR, VIAN S, WRIGHT JT, KORRKO J, BRUNONI D, HART TC. Novel COL1A1 mutation (G599C) associated with mild osteogenesis imperfecta and dentinogenesis imperfecta. *Arch Oral Biol* 2001; **46**: 459–470.
12. KINNEY JH, POPLE JA, DRIESSEN CH, BREUNIG TM, MARSHALL GW, MARSHALL SJ. Intrafibrillar mineral may be absent in dentinogenesis imperfecta type II (DI-II). *J Dent Res* 2001; **80**: 1555–1559.
13. 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. *Coll Relat Res* 1981; **1**: 187–199.
14. BUTLER WT, BHOWN M, BRUNN JC, D'SOUZA RN, FARACH-CARSON MC, HAPPONEN R-P, SCHROHENLOHER RE, SEYER JM, SOMERMAN MJ, FOSTER RA, TOMANA M, VAN DIJK S. Isolation, characterization and immunolocalization of a 53-kDa Dentin Sialoprotein (DSP). *Matrix* 1992; **12**: 343–351.
15. QIN C, COOK RG, ORKISZEWSKI RS, BUTLER WT. Identification and characterization of the carboxyl-terminal region of rat dentin sialoprotein. *J Biol Chem* 2001; **276**: 904–909.
16. BUTLER WT, RITCHIE H. The nature and functional significance of dentin extracellular matrix proteins. *Int J Dev Biol* 1995; **39**: 343–351.
17. JONSSON M, FREDRIKSSON S, JONTELL M, LINDE A. Isoelectric focusing of the phosphoprotein of rat-incisor dentin in ampholine and acid pH gradients. Evidence for carrier ampholyte-protein complexes. *J Chromatogr* 1978; **157**: 234–242.
18. BUTLER WT, BHOWN M, DIMUZIO MT, COTHRAN WC, LINDE A. Multiple forms of rat dentin phosphoproteins. *Arch Biochem Biophys* 1983; **225**: 178–186.
19. HUQ NL, CROSS KJ, TALBO GH, RILEY PF, LOGANATHAN A, CROSSLEY MA, PERICH JW, REYNOLDS EC. N-terminal sequence analysis of bovine dentin phosphophoryn after conversion of phosphoserine to S-propylcysteine residues. *J Dent Res* 2000; **79**: 1914–1919.
20. RITCHIE HH, WANG LH, KNUDTSON K. A novel rat 523 amino acid phosphophoryn: nucleotide sequence and genomic organization. *Biochim Biophys Acta* 2001; **1520**: 212–222.
21. 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.
22. 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.
23. 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.
24. GEORGE A, SRINIVASAN RSR, LIU K, VEIS A. Rat dentin matrix protein 3 is a compound protein of rat dentin sialoprotein and phosphophoryn. *Connect Tissue Res* 1999; **40**: 49–57.
25. RITCHIE HH, WANG L. The presence of multiple rat DSP-PP transcripts. *Biochim Biophys Acta* 2000; **1493**: 27–32.
26. RITCHIE HH, LI X. A novel rat dentin mRNA coding only for dentin sialoprotein. *Eur J Oral Sci* 2001; **109**: 342–347.
27. ALTSCHUL S, MADDEN T, SCHÄFFER A, ZHANG J, ZHANG Z, MILLER W, LIPMAN D. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* 1997; **25**: 3389–3402.
28. HIGGINS DG, THOMPSON JD, GIBSON TJ. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol* 1996; **266**: 383–402.
29. HANSEN JE, LUND O, TOLSTRUP N, GOOLEY AA, WILLIAMS KL, BRUNAK S. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate J* 1998; **15**: 115–130.
30. BLOM N, GAMMELTOFT S, BRUNAK S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 1999; **294**: 1351–1362.
31. 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.
32. 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 (Dsp) gene, which codes for both dentin sialoprotein and dentin phosphoprotein. *J Biol Chem* 1998; **273**: 9457–9464.
33. VON HEINE G. Signal sequences. The limits of variation. *J Mol Biol* 1985; **184**: 99–105.
34. OIDA S, NAGANO T, YAMAKOSHI Y, ANDO H, YAMADA M, FUKAE M. Amelogenin gene expression in porcine odontoblasts. *J Dent Res* 2002; **81**: 103–108.
35. FUKAE M, IJIRI H, TANABE T, SHIMIZU M. Partial amino acid sequences of two proteins in developing porcine enamel. *J Dent Res* 1979; **58**: 1000–1001.
36. FUKAE M, TANABE T, IJIRI H, SHIMIZU M. Studies on porcine enamel proteins: a possible original enamel protein. *Tsurumi U Dent J* 1980; **6**: 87–94.
37. YAMAKOSHI Y, TANABE T, FUKAE M, SHIMIZU M. Porcine amelogenins. *Calcif Tissue Int* 1994; **54**: 69–75.

38. FUKAE M, TANABE T, MURAKAMI C, DOHI N, UCHIDA T, SHIMIZU M. Primary structure of porcine 89 kDa enamel. *Adv Dent Res* 1996; **10**: 111–118.
39. SIMMER JP, FUKAE M, TANABE T, YAMAKOSHI Y, UCHIDA T, XUE J, MARGOLIS HC, SHIMIZU M, DEHART BC, HU CC, BARTLETT JD. Purification, characterization, and cloning of enamel matrix serine proteinase 1. *J Dent Res* 1998; **77**: 377–386.
40. 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.
41. RITCHIE H, WANG LH. A mammalian bicistronic transcript encoding two dentin-specific proteins. *Biochem Biophys Res Comms* 1997; **231**: 425–428.
42. SHAPIRO MB, SENAPATHY P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl Acids Res* 1987; **15**: 7155–7174.
43. WATAKABE A, TANAKA K, SHIMURA Y. The role of exon sequences in splice site selection. *Genes Dev* 1993; **7**: 407–418.
44. BURSET M, SELEDTSOV IA, SOLOVYEV VV. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucl Acids Res* 2000; **28**: 4364–4375.
45. MODREK B, LEE C. A genomic view of alternative splicing. *Nature Genet* 2002; **30**: 13–19.