Genomic Characterization of Human *DSPG3*

Michelle Deere,[1,2] Jose L. Dieguez,[4–6] Sung-Joo Kim Yoon,[8] David Hewett-Emmett,[1,3] Albert de la Chapelle,[4,5,7] and Jacqueline T. Hecht[1,2]

[1]Graduate School of Biomedical Sciences, Program in Human and Molecular Genetics; [2]Department of Pediatrics and [3]Human Genetics Center, University of Texas, Health Science Center, Houston, Texas 77030 USA; [4]Department of Medical Genetics, University of Helsinki, Finland; [5]Folkhalsan Institute of Genetics, Helsinki, Finland; [6]Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Malaga, Spain; [7]Human Cancer Genetics Program, Comprehensive Cancer Center, Ohio State University, Columbus, Ohio 43210 USA; and [8]Research Laboratory of Molecular Genetics, Research Institutes of Medical Science, Catholic University Medical College, Seoul, Korea

DSPG3, the human homolog to chick PG-Lb, is a member of the small leucine-rich repeat proteoglycan (SLRP) family, including decorin, biglycan, fibromodulin, and lumican. In contrast to the tissue distribution of the other SLRPs, *DSPG3* is predominantly expressed in cartilage. In this study, we have determined that the human *DSPG3* gene is composed of seven exons: Exon 2 of *DSPG3* includes the start codon, exons 4–7 code for the leucine-rich repeats, exons 3 and 7 contain the potential glycosaminoglycan attachment sites, and exon 7 contains the potential N-glycosylation sites and the stop codon. We have identified two polymorphic variations, an insertion/deletion composed of 19 nucleotides in intron 1 and a tetranucleotide $(TATT)_n$ repeat in intron 5. Analysis of 1.6 kb of upstream promoter sequence of *DSPG3* reveals three TATA boxes, one of which is 20 nucleotides before the transcription start site. The transcription start site precedes the translation start site by 98 nucleotides. There are 14 potential binding sites for SOX9, a transcription factor present in cartilage, in the promoter, and in the first intron of *DSPG3*. We have examined the evolution of the SLRP gene family and found that gene products clustered together in the evolutionary tree are encoded by genes with similarities in genomic structure. Hence, it appears that the majority of the introns in the SLRP genes were inserted after the differentiation of the SLRP genes from an ancestral gene that was most likely composed of 2–3 exons.

[The sequence data described in this paper have been submitted to GenBank under accession nos. AF031658 and U63814.]

DSPG3 is the human homolog to chick PG-Lb, an extracellular matrix proteoglycan originally isolated from epiphyseal cartilage (Shinomura et al. 1983; Deere et al. 1996). DSPG3 (PG-Lb) is a member of the small leucine-rich repeat proteoglycan (SLRP) family, including decorin, biglycan, fibromodulin, and lumican (Krusius and Ruoslahti 1986; Fisher et al. 1989; Oldberg et al. 1989; Shinomura and Kimata 1992; Deere et al. 1996). The core proteins of the SLRPs are composed of 6–10 tandem repeats of 24 amino acid residues that are rich in leucine (for review, see Kobe and Deisenhofer 1994). The leucine-rich repeats (LRR) are preceded by four cysteines and followed by two cysteines that are presumed to form disulfide bonds on either side of the LRRs. Related LRR glycoproteins include prolargin (PRELP), osteoglycin (formerly known as osteoinductive factor), and osteomodulin (Madisen et al. 1990; Bengtsson et al. 1995; Grover et al. 1996; Ohno et al. 1996).

The SLRPs are clustered in a few chromosomal regions. *Decorin*, *lumican*, and *DSPG3* map to human chromosome 12q21–q22 (McBride et al. 1990; Danielson et al. 1993; Vetter et al. 1993; Chakravarti et al. 1995; Grover et al. 1995; Deere et al. 1996). *Fibromodulin* and *PRELP* are localized to human chromosome 1q32 (Sztrolovics et al. 1994; Grover et al. 1996). Currently, only one SLRP, *biglycan*, maps to human chromosome Xq28 (McBride et al. 1990; Fisher et al. 1991; Traupe et al. 1992).

The SLRPs have related genomic structures. The *fibromodulin*, *lumican*, and *PRELP* genes are composed of three exons, and the first intron, in each case, immediately precedes the start site of translation, whereas the second intron is in the last LRR (Antonsson et al. 1993; Grover et al. 1995, 1996). The *decorin* and *biglycan* genes are both composed of eight exons, with the positions of two introns corresponding with those identified in *fibromodulin*, *lumican*, and *PRELP* (Fisher et al. 1991; Danielson et al. 1993; Vetter et al. 1993). The other five introns are present in the LRRs at identical sites.

DSPG3, in contrast to the other SLRPs, is predominantly expressed in cartilage (Shinomura and Kimata 1992; Deere et al. 1996; Kurita et al. 1996). There are several important extracellular matrix proteins expressed primarily in cartilage: collagen types II, IX, X, and XI, aggrecan, and link protein (for reviews, see Heinegard and Oldberg 1989; Hall and Newman 1991).

[8]Corresponding author.
E-MAIL jhecht@ped1.med.uth.tmc.edu; FAX (713) 500-5689.

Promoter studies for these genes have identified regions that may be important for the cartilage-specific transcription of these genes, including a binding site for SOX9 in the first intron of type II collagen (Nishimura et al. 1989; Rhodes and Yamada 1995; Thomas et al. 1995; Krebsbach et al. 1996, Lefebvre et al. 1997). SOX9 is a HMG (high-mobility group) transcription factor. Mutations in *SOX9* cause campomelic dysplasia, a skeletal dysplasia, which suggests that SOX9 is important for the regulation of genes in normal cartilage development (Foster et al. 1994; Wagner et al. 1994).

In this study we have delineated the genomic structure of human *DSPG3* and compared the exon/intron boundaries with those of *lumican* and *decorin*. We have also sequenced the promoter region of human *DSPG3* and have identified transcriptional elements that may be important for the cartilage-specific expression of *DSPG3*. In addition, using available protein sequence data, we have performed the most complete evolutionary analysis of the SLRP gene family.

RESULTS AND DISCUSSION

Genomic Structure of *DSPG3*

The genomic structure of human *DSPG3* is composed of seven exons and spans more than 12 kb (Table 1). This structure is conserved with murine *PG-Lb* (Iwata et al. 1998). Exon 1 consists of 5′-untranslated region, and the start codon is present in the second exon. Exons 4–7 encode the LRRs. Exons 3 and 7 contain the potential glycosaminoglycan attachment sites (codons 64, 96, and 320), and exon 7 includes the consensus N-glycosylation sites (codons 283 and 302), the stop codon, and 3′-untranslated region. The majority of the intron sizes are ~1 kb. Exceptions are intron 1 (2.2 kb) and intron 2, whose size as determined by Southern blot analysis is at least 5 kb (data not shown). All of the splice donor and acceptor sites follow the consensus GT–AG rule. Two intronic polymorphisms, with low heterozygosities, were identified: An insertion/deletion of 19 nucleotides (TTGAACATCTGGCAGCAAT, nucleotides 3747–3765) in intron 1 (heterozygosity = 0.21) and a tetranucleotide $(TATT)_n$ repeat in intron 5 (heterozygosity = 0.39) (Table 2) (GenBank accession nos. AF031658 and U63814, respectively).

Table 2. Allele Frequencies

Allele	Size (bp)	Frequency
a. For the insertion/deletion in intron 1 of the *DSPG3* gene		
A1	449	0.8
A2	430	0.2
b. For the tetranucleotide repeat in intron 5 of the *DSPG3* gene		
A1	279	0.04
A2	275	0.14
A3	271	0.69
A4	267	0.08
A5	263	0.02
A6	255	0.03

Promoter Sequence of *DSPG3*

The sequence of the promoter and first intron of *DSPG3* has been deposited in GenBank (accession no. AF031658). The start site of transcription is 98 nucleotides upstream of the start codon (mRNA sequence) as demonstrated by a ribonuclease protection assay (data not shown). The transcription start site is 20 nucleotide downstream of a TATA box (Fig. 1). There are several potential transcription factor binding sites present in the promoter sequence. The most noteable site, (A/T)(A/T)CAA(A/T)G, is the consensus DNA-binding site for HMG domain transcription factors, including SOX9 (Grosschedl et al. 1994; Sudbeck et al. 1996; Lefebvre et al. 1997). This site is present 4 times in the promoter region and 10 times in the first intron of *DSPG3*. Several of these sites are conserved with murine *PG-Lb* (Iwata et al. 1998). Mutations in *SOX9* cause campomelic dysplasia, a skeletal dysplasia associated with sex reversal and

Table 1. Genomic Structure of Human *DSPG3*

Exon no.	Exon size	Intron location	Intron size	Splice donor	Splice acceptor	AA site	Type
1	85	85/86	2368	AAG gtaag	tag GAA		
2	178	263/264	5000+	GAG gtaat	cag ATT	Glu/Ile	0
3	175	438/439	830+	AAG gtcag	tag ACT	Asp	I
4	159	597/598	819	TAA gtatg	tag GTG	Ser	I
5	203	800/801	825+	AAA gtaag	cag GAC	Lys/Asp	0
6	96	896/897	800+	CAG gtagg	tag AAT	Gln/Asn	0
7	671						

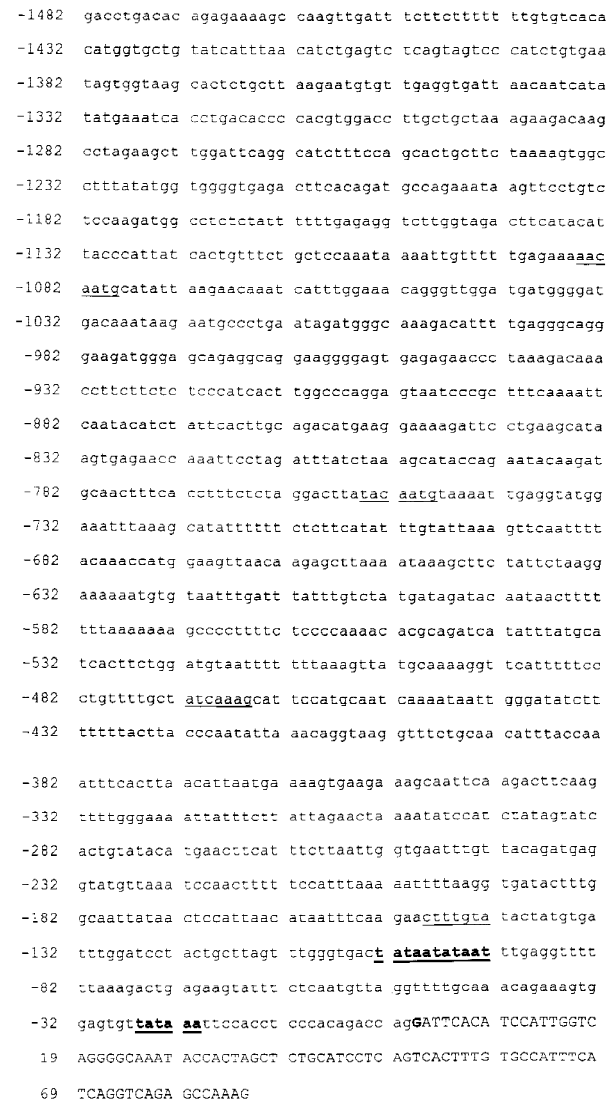

Figure 1 Sequence of the promoter of *DSPG3* with the potential SOX9-binding sites (underlined), TATA boxes (bold and underlined), and transcription start site (bold) noted. Exon 1 is in uppercase letters.

lethality (Foster et al. 1994; Wagner et al. 1994). These mutations suggest that SOX9 is an important transcription factor in cartilage, as well as testis development. *DSPG3* is primarily expressed in cartilage, and SOX9 may play an important role in the tissue-specific expression of this gene. Support for this conclusion comes from the cartilage-specific, type II collagen gene, *COL2A1*. *COL2A1* was recently shown to contain a SOX9-binding site in the first intron that was necessary for the correct tissue expression of the gene (Bell et al. 1997; Lefebvre et al. 1997). Further studies will be necessary to prove whether the potential SOX9-binding sites are necessary for *DSPG3* expression in cartilage.

Evolutionary Analysis of the SLRP Gene Family

In this study we have compared the genomic structures of the human SLRP genes and related proteins to determine whether there were indications of shared placement of the introns (Fig. 2). Analysis of the newly determined genomic structure of *DSPG3* demonstrates that the first intron is present in the 5′-untranslated region, whereas the last (sixth) intron is present in the last LRR, a pattern similar to the genomic structures of the other SLRPs. However, the placement of introns 2–5 in the LRR region of *DSPG3* shows no correspondence with introns 2–6 in *decorin/biglycan* (which are identical with each other) that are also located in the region encoding the LRRs (Fig. 2). Also, the last (seventh) LRR of *DSPG3* aligns with the seventh LRR of *decorin* and *biglycan* and not the last (tenth) LRR. Therefore, the genomic structures of *DSPG3* and *decorin/biglycan* appear to have evolved independently but with a shared tendency for the introns to occur in the LRR-encoding region. The other major group of SLRP genes (*fibromodulin*, *lumican*, and *PRELP*) have three exons. As with *DSPG3*, the first intron occurs upstream of the start site, whereas the second occurs in the final LRR. It should be noted that the final LRR of DSPG3 is in repeat seven (LRRs 8–10 are absent or deleted) and the exact position of the final intron differs between *biglycan/decorin* and *fibromodulin/lumican/PRELP*. Also, the final intron in *fibromodulin* is shifted 1 bp relative to that of *lumican* and *PRELP*. Therefore, it seems that the ancestral SLRP had one intron upstream of the start codon, and that most, if not all, of the additional introns were introduced separately in at least three lineages: *DSPG3*; *biglycan/decorin*; *fibromodulin/lumican/PRELP*. This observation is in accord with the introns late hypothesis (Stolzfus et al. 1997).

To analyze further the evolutionary history of the SLRP gene family, protein sequences from SLRPs and related *LRR* genes were aligned. Because the alignment was uncertain in the amino- and carboxy-terminal regions, only the LRR-containing region bounded by the four amino-terminal and two carboxy-terminal cysteines was used in the analysis (see Fig. 2) (The full alignment used is available on request.). Table 3 summarizes data on the SLRPs and related genes used in this analysis. neighbor-joining trees (Saitou and Nei 1987) were built with both *p*-distances (no correction for multiple mutations) and poisson-corrected distance matrices, each calculated by pairwise or complete deletions (Kumar et al. 1993). Because all four trees were very similar, only that using the *p*-distance with complete deletion of sites in which one or more sequence has a deletion is shown in Figure 3. The tree shows that the SLRP genes with similar genomic structures group together. *DSPG3* (*PG-Lb*) and *osteoglycin* appear to have

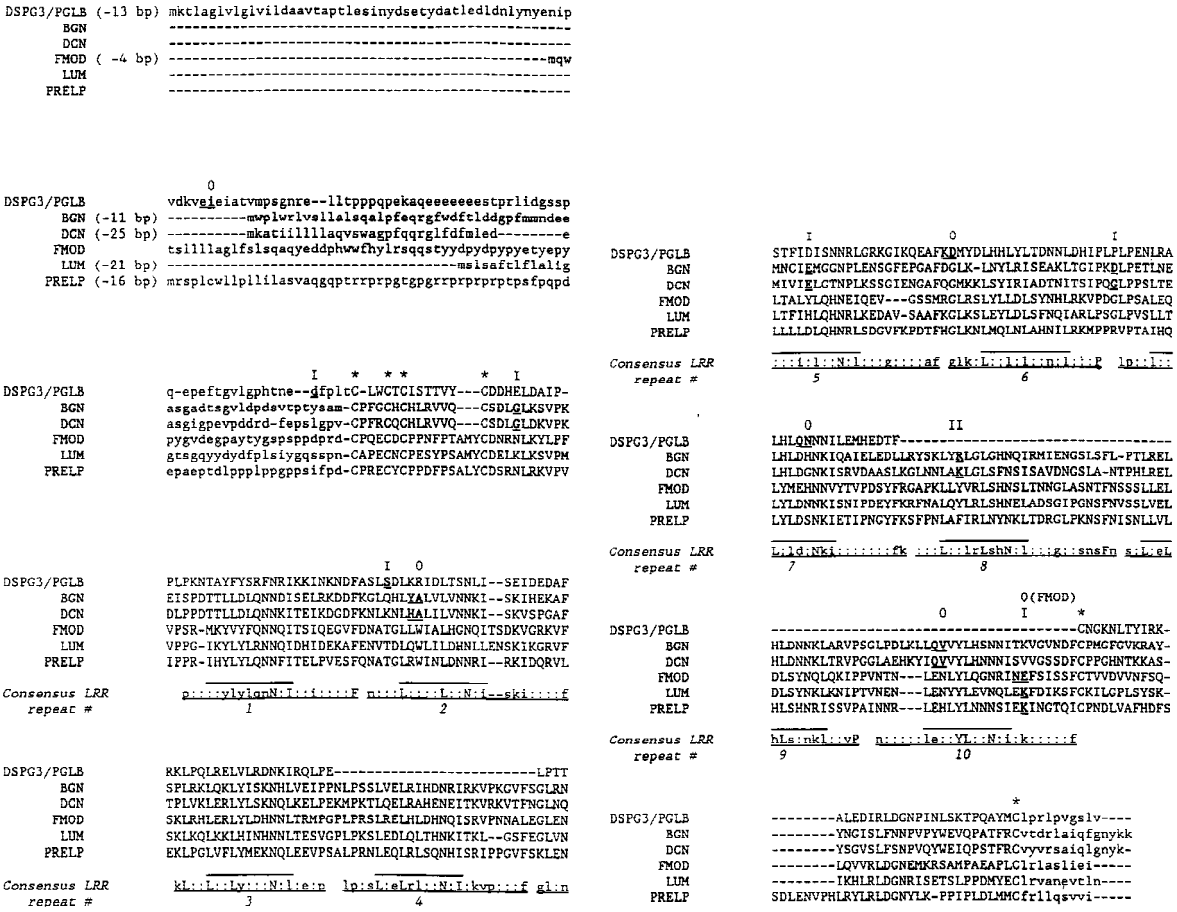

Figure 2 Protein sequence alignment of human DSPG3/PGLB, biglycan (BGN), decorin (DCN), lumican (LUM), fibromodulin (FMOD), and prolargin (PRELP). Amino acids whose codons are interrupted by an intron are in bold and underlined; where an intron occurs between two codons, both encoded amino acids are underlined. Intron phase is shown at *top* (0) Intron between codons; (I) intron after first base of codon; (II) intron after second base of codon. The first intron of each gene is upstream of the start codon; distance upstream is shown (e.g., −13 bp for DSPG3/PGLB). The LRRs are numbered and indicated by lines at *bottom*; consensus sequence of each LRR is also shown. (Uppercase letters) Amino acids completely conserved; (lowercase letters) >50% conserved in the six sequences; (*) the conserved cysteines flanking the LRRs. The central portion of the protein alignment that was used in the evolutionary analysis is in uppercase letters.

evolved separately from *biglycan/decorin* and *lumican/fibromodulin/PRELP/osteomodulin*. Our data correlates with previous dendrograms that have been constructed (Bengtsson et al. 1995; Iozzo 1997, 1998; Sommarin et al. 1998). However, this is the first study to include sequence data from multiple species allowing us to analyze the conservation of these genes as a family.

Interestingly, in all four trees (Fig. 3; data not shown), the chicken *decorin* gene is more closely related to the human and bovine *decorin* genes than is the murine *decorin* gene. Because the bootstrap support for this unexpected finding is highly significant (97%, Fig. 3), it indicates that there was probably a duplication of an ancestral *decorin* gene prior to the bird/reptile/mammal divergence and that the cloned mouse gene represents one paralog, whereas the chicken, bovine, and human genes represent the other paralog. Somatic cell mapping of *decorin* demonstrated two signals on chromosome 12, also indicating that there may potentially be another member of the SLRPs present at that chromosomal region that is highly homologous to *decorin* (McBride et al. 1990). This may prove to be the human ortholog of the cloned mouse *decorin* gene.

In this study we have identified and sequenced the intron/exon borders of human *DSPG3* and determined that the gene is composed of seven exons. Two intronic polymorphisms were identified and characterized. We have also cloned and sequenced the promoter and first intron of *DSPG3*. Several putative transcription factor-binding sites, including the potential SOX9-binding sites, were identified. Further analysis of the transcriptional elements present in *DSPG3* will be necessary to determine the mechanisms involved in

Table 3. SLRP and Related Gene Sequences Used in this Study

Gene name	Gene symbol	Animal	GenBank acc. no.	Reference
DSPG3/PG-Lb	*DSPG3/PG-Lb*	human	U59111	Deere et al. (1996)
PG-Lb/epiphycan	*PG-Lb*	bovine	U77127	Johnson et al. (1997)
PG-Lb	*PG-Lb*	murine	D78274	Kurita et al. (1996)
		chicken	D10485	Shinomura et al. (1992)
Osteoglycin/ osteoinductive factor	*OGN/OIF*	human	B35272	Madisen et al. (1990)
		bovine	M37974	Madisen et al. (1990)
		murine	D31951	Ujita et al. (1995)
Biglycan	*BGN*	human	J04599	Fisher et al. (1989)
		bovine	S82652	Xu et al. (1995)
		murine	L20276	Rau et al. (1994)
Decorin	*DCN*	human	L01125–L01131	Vetter et al. (1993)
			M98262–M98263	Danielson et al. (1993)
		bovine	S06280	Day et al. (1987)
		murine	A55454	Scholzen et al. (1994)
		chicken	X63797	Li et al. (1992)
Lumican	*LUM*	human	U21128	Chakravarti et al. (1995)
		bovine	L11063	Funderburgh et al. (1993)
		murine	S79461	Funderburgh et al. (1995)
		chicken	M80584	Blochberger et al. (1992)
Fibromodulin	*FMOD*	human	X72913	Antonsson et al. (1993)
		bovine	X16485	Oldberg et al. (1989)
		murine	X94998	Saamanen et al. (1996)
		chicken	U34977	Nurminskaya and Birk (1996)
Prolargin	*PRELP*	human	U29089	Bengtsson et al. (1995)
Osteomodulin	*OMD*	human	AB000114	Ohno et al. (1996)

the specific regulation and expression of DSPG3 in cartilage. We have compared the genomic structures of *DSPG3* and other members of the SLRP gene family, and have shown that the introns within the LRRs must have arisen separately in *DSPG3* and *decorin/biglycan*. Our evolutionary analysis of the SLRP gene family confirms this hypothesis. SLRP genes with similar gene structures were more closely related to each other than they were to the other SLRP genes. It appears that the ancestral SLRP gene was composed of two (or possibly three) exons and that additional introns were inserted in *DSPG3*, *decorin/bigycan*, and (probably) *fibromodulin/ lumican/PRELP*. In addition, there appears to have been intron slippage in the intron upstream of the start codon and in the second fibromodulin intron. It will be interesting to see how additional data (genomic structure, chromosomal location) on the *osteoglycin* and *osteomodulin* genes help shape the evolutionary scheme proposed, and whether a second *decorin* gene is present.

METHODS

Identification of Cosmids Containing *DSPG3*

cDNA template was amplified with primers hepn3/hepn2 (Deere et al. 1996). This PCR product was random prime labeled and used to hybridize a dot blot of chromosome 12-specific cosmids following standard procedures. Cosmids 167H5, 24C10, 231B8, 133C5, 204F1 196B7, 61B9, and 207C11 were positive for the *DSPG3* probe.

Identification and Sequencing of the Intron/Exon Borders of Human *DSPG3*

cDNA primers were used to amplify genomic DNA from cosmid 167H5 to identify the locations of introns in the gene. The primer sets used were hepn3/hepn15, hepn1/239861, bepn4/hepn6, and hepn5/hepn8 (Deere et al. 1996; Table 4). Intron/exon borders were sequenced by a series of primers (Table 4) by direct sequencing of cosmid 167H5 using an ABI automated sequencer. Sequencing of each region was performed at least twice in two separate laboratories. The resulting sequence was analyzed by the GCG database system (Genetics Computer Group 1994).

Analysis of Polymorphic Repeats

The 19-bp insertion/deletion in intron one was amplified from 120 unrelated caucasian individuals by PCR primers (forward, 5′-TCTTCACCTATAAAATGGTATGACA-3′; and reverse, 5′-TCTTCATTTTTCAAGCTTTCC-3′) following standard conditions (Sambrook et al. 1989). The PCR products were analyzed on 6% acrylamide gels.

PCR primers (forward, 5′-TTTGCTGTCATTGACTACC-3′; and reverse, 5′-GCGAAACCATGTCTCTAC-3′) were designed to amplify the tetranucleotide repeat $(TATT)_n$ in intron 5 of *DSPG3* with a predicted PCR product size of 275 bp. Fifty-six unrelated individuals were amplified following standard procedures (Sambrook et al. 1989). The samples were analyzed on

6% denaturing polyacrylamide gels and silver-stained by the GelCode System (Pierce).

Sequencing Promoter Region of Human *DSPG3*

Cosmid 167H5 does not contain the promoter region of *DSPG3*. Therefore, cosmid 207C11 was used to sequence the promoter region. Cosmid 207C11 was subcloned into the *Eco*RI site in pBlueScript SK(+). Clones were then sequenced with a series of primers (Table 4) on an ABI automated sequencer. The promoter sequence was verified by amplification and sequencing of the promoter region from genomic DNA in a separate laboratory. The sequence was analyzed by the GAP program from the GCG database system (Genetics Computer Group 1994).

Ribonuclease Protection Assay

Primers, hepn46 (5′-GAATTTGTTACAGATGAGG-3′) and hepn47 (5′-GCAAGTATAAAAACTTACCT-3′), were used to amplify the first exon and 313 bp of the promoter region. The PCR product was cloned into the pGEM-T vector (Promega), and the clone was digested with *Nco*I to linearize the DNA for the probe. The probe was transcribed with SP6 polymerase and gel purified. The ribonuclease protection assays were performed with the RPA II kit from Ambion, Inc.

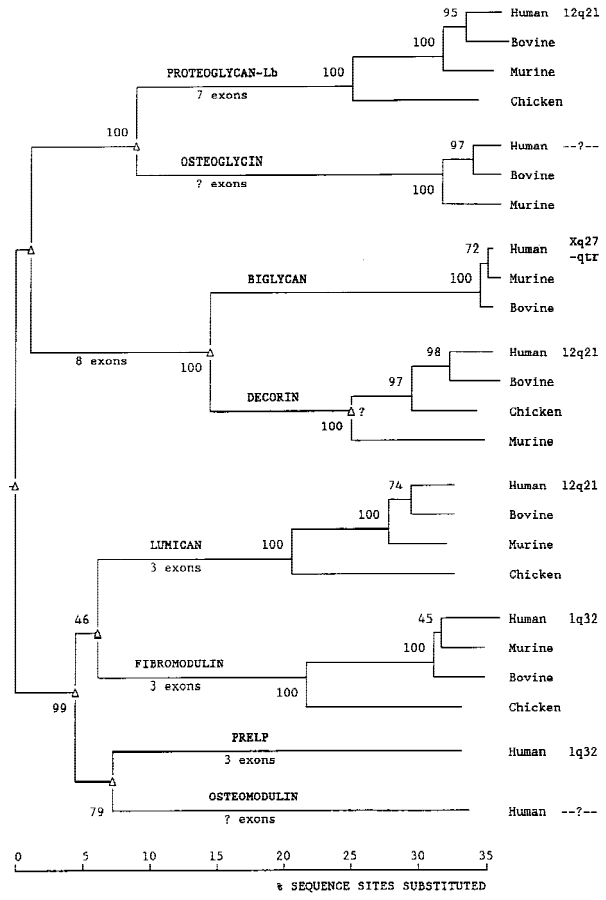

Figure 3 Phylogenetic tree of the SLRP gene family and related genes. (△) Gene duplication. Numbers on branches (at *left* of node) represent the percent bootstrap replicates supporting that node. Chromosomal location of the human genes is shown. Branches are to scale and represent percent amino acid sequence sites substituted. No correction for undetected multiple replacements at a site was made in the tree shown. Root of the tree is arbitrarily placed.

Table 4. Sequencing Primers for the Introns and Promoter of DSPG3

Primer	Intron/ promoter	Sequence (5′–3′)
hepn1	intron 4	CCGCTTATTTCTATTCCCGCTTTA
hepn2	intron 3	GCGGGAATAGAAATAAGCGGTGGT
hepn4	intron 1	GGTGGCATCATAGGTTTCTG
hepn6	intron 6	GCTTGTGGAGTTTTGCTGAG
hepn7	intron 1	TAGAGTTGGGGCAGTCACAG
hepn9	intron 2	CATACCTGTTGATAAAGTTG
hepn15	intron 2	GAGAAGAGCCATCAATCAGC
hepn16	intron 2	ATTCCTCCTCCTCCTCCTCT
hepn19	intron 2	CTCATGTGTTTTCAATATTA
hepn21	intron 2	ATTGAAAACACATGAGAAAT
hepn23	introns 2 and 3	AAGCAGGATGGTCAAACT
hepn25	intron 6	GTATGATCTCCATCATCTGT
hepn26	intron 5	TGAAGGGCTCGTAGATTTTC
hepn29	intron 5	TTTGCTGTCATTGACTACC
hepn30	intron 5	GCGAAACCATGTCTCTAC
hepn33	intron 1	GATGACGGTGATGATGACTG
hepn35	intron 1	ATTCCCTTGTATGCCTGTGG
hepn36	intron 1	CTCTATTTCAGTTGCCTTTG
hepn37	intron 1	TCACACAGGATAAACTAAGC
hepn40	intron 1	CCTTATCATCTTCAACTTCA
hepn42	intron 1	GCATTTTGCATCACTCCT
hepn43	intron 1	GTCAATAACAAAAATAAACCAA
hepn45	promoter	CTGGTCTGTGGGAGGTGGAA
hepn47	promoter	GCAAGTATAAAAACTTACCT
hepn48	promoter	TTTTTCCTTAGAATAGAAGC
hepn49	promoter	AACAGGGAAAAATGAACCTT
hepn50	promoter	TTTGAAAGCGGGATTACT
bepn3	intron 5	AGGCAGCTCCCAGAA
bepn4	intron 4	GTCACGCAGGACAAGC

Evolutionary Analysis of the SLRP Gene Family

Published protein sequences of the human, bovine, murine, and chicken SLRP genes and related proteins were collected and an alignment made of the region between the first and last cysteines flanking the leucine-rich repeats by the LINEUP, PILEUP, and PRETTY programs from the GCG database system (Genetics Computer Group 1994). The proteins included DSPG3 (PG-Lb), osteoglycin, biglycan, decorin, lumican, fibromodulin, PRELP, and osteomodulin (Table 3). Alignments were verified with TBLASTN output of the DSPG3 protein sequence for the GenBank database (Genetics Computer Group 1994). Phylogenetic trees were built with the Molecular Evolutionary Genetics Analysis program, version 1.01 (MEGA) (Kumar et al. 1993). Four different neighbor-joining trees were built from *p*-distance and poisson-corrected distance matrices with both complete and pairwise deletions with 1000 bootstraps (Saitou and Nei 1987; Kumar et al. 1993).

ACKNOWLEDGMENTS

We thank Dr. Raju Kucherlapati for providing the chromo-

some 12-specific cosmids. This research was supported by a Schissler Foundation Fellowship to M.D., Shriners Hospital for Children grants 15,955 and 15,957 to J.T.H., and grant support from The Academy of Finland and The Ulla Hjelt Fund to A.C. and J.L.D. The Department of Biomathematics at M.D. Anderson Cancer Center is supported by grant NCI CA-16672.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Antonsson, P., D. Heinegard, and A. Oldberg. 1993. Structure and deduced amino acid sequence of the human *fibromodulin* gene. *Biochim. Biophys. Acta* **1174:** 204–206.

Bell, D.M., K.K.H. Leung, S.C. Wheatley, L.J. Ng, S. Zhou, K.W. Ling, M.H. Sham, P. Koopman, P.P.L. Tam, and K.S.E. Cheah. 1997. SOX9 directly regulates the *type-II collagen* gene. *Nat. Genet.* **16:** 174–178.

Bengtsson, E., P.J. Neame, D. Heinegard, and Y. Sommarin. 1995. The primary structure of a leucine-rich repeat protein, *PRELP*, found in connective tissues. *J. Biol. Chem.* **270:** 25639–25644.

Blochberger, T.C., J.P. Vergnes, J. Hempel, and J.R. Hassell. 1992. cDNA to chick *lumican* (corneal keratan sulfate proteoglycan) reveals homology to the small interstitial proteoglycan gene family and expression in muscle and intestine. *J. Biol. Chem.* **267:** 347–352.

Chakravarti, S., R.L. Stallings, N. SundarRaj, P.K. Cornuet, and J.R. Hassell. 1995. Primary structure of human *lumican* (keratan sulfate proteoglycan) and localization of the gene (LUM) to chromosome 12q21.3–q22. *Genomics* **27:** 481–488.

Danielson, K.G., A. Fazzio, I. Cohen, L.A. Cannizzaro, I. Eichstetter, and R.V. Iozzo. 1993. The human decorin gene: Intron-exon organization, discovery of two alternatively spliced exons in the 5′ untranslated region, and mapping of the gene to chromosome 12q23. *Genomics* **15:** 146–160.

Day, A.A., C.I. McQuillan, J.D. Termine, and M.R. Young. 1987. Molecular cloning and sequence analysis of the cDNA for *small proteoglycan II* of bovine bone. *J. Biochem.* **248:** 801–805.

Deere, M., J. Johnson, S. Garza, W.R. Harrison, S.J. Yoon, F.F.B. Elder, R. Kucherlapati, M. Hook, and J.T. Hecht. 1996. Characterization of human *DSPG3*, a small dermatan sulfate proteoglycan. *Genomics* **38:** 399–404.

Fisher, L.W., J.D. Termine, and M.F. Young. 1989. Deduced protein sequence of bone small proteoglycan I (biglycan) shows homology with proteoglycan II (decorin) and several nonconnective tissue proteins in a variety of species. *J. Biol. Chem.* **264:** 4571–4576.

Fisher, L.W., A.M. Heegaard, U. Vetter, W. Vogel, W. Just, J.D. Termine, and M.F. Young. 1991. Human *biglycan* gene: Putative promoter, intron-exon junctions, and chromosomal localization. *J. Biol. Chem.* **266:** 14371–14377.

Foster, J.W., M.A. Dominguez-Steglich, S. Guioli, C. Kwok, P.A. Weller, M. Stevanovic, J. Weissenbach, S. Mansour, I.D. Young, P.N. Goodfellow, J.D. Brook, and A.J. Schafer. 1994. Campomelic dysplasia and autosomal sex reversal caused by mutations in an *SRY*-related gene. *Nature* **372:** 525–530.

Funderburgh, J.L., M.L. Funderburgh, S.J. Brown, J.-P. Vergnes, J.R. Hassell, M.M. Mann, and G.W. Conrad. 1993. Sequence and structural implications of a bovine corneal keratan sulfate proteoglycan core protein: Protein 37B represents bovine lumican and proteins 37A and 25 are unique. *J. Biol. Chem.* **268:** 11874–11880.

Funderburgh, J.L., M.L. Funderburgh, N.D. Hevelone, M.E. Stech, M.J. Justice, C.-Y. Liu, W.W.-Y. Kao, and G.W. Conrad. 1995. Sequence, molecular properties, and chromosomal mapping of mouse *lumican*. *Invest. Opthalmol. Vis. Sci.* **36:** 2296–2303.

Genetics Computer Group. 1994. Program manual for the Wisconsin package, version 8. University of Wisconsin, Madison, WI 53711.

Grosschedl, R., K. Giese, and J. Pagel. 1994. HMG domain proteins: Architechtural elements in the assembly of nucleoprotein structures. *Trends Genet.* **10:** 94–100.

Grover, J., X.-N. Chen, J.R. Korenberg, and P.J. Roughley. 1995. The human *lumican* gene: Organization, chromosomal location, and expression in articular cartilage. *J. Biol. Chem.* **270:** 21942–21949.

Grover, J., X.-N. Chen, J.R. Korenberg, A.D. Recklies, and P.J. Roughley. 1996. The gene organization, chromosomal location, and expression of a 55-kD matrix protein (*PRELP*) of human articular cartilage. *Genomics* **38:** 109–117.

Hall, B. and S. Newman. 1991. *Cartilage: Molecular aspects*. CRC Press, Boca Raton, FL.

Heinegard, D. and O. Oldberg. 1989. Structure and biology of cartilage and bone matrix noncollagenous macromolecules. *FASEB J.* **3:** 2042–2051.

Iozzo, R.V. 1997. The family of small leucine-rich proteoglycans: Key regulators of matrix assembly and cellular growth. *Crit. Rev. Biochem. Mol. Biol.* **32:** 141–174.

———. 1998. Matrix proteoglycans: From molecular design to cellular function. *Annu. Rev. Biochem.* **67:** 609–652.

Iwata, Y., T. Shinomura, K. Kurita, M. Zako, and K. Kimata. 1998. The gene structure and organization of mouse *PG-Lb*, a small chondroitin/dermatan sulfate proteoglycan. *Biochem. J.* **331:** 959–964.

Johnson, H.J., L. Rosenberg, H.U. Choi, S. Garza, M. Hook, and P.J. Neame. 1997. Characterization of epiphycan, a small proteoglycan with a leucine-rich repeat core protein. *J. Biol. Chem.* **272:** 18709–18717.

Kobe, B. and J. Deisenhofer. 1994. The leucine-rich repeat: A versatile binding motif. *Trends Biochem. Sci.* **19:** 415–421.

Krebsbach, P.H., K. Nakata, S.M. Bernier, O. Hatano, T. Miyashita, C.S. Rhodes, and Y. Yamada. 1996. Identification of a minimum enhancer sequence for the Type II collagen gene reveals several core sequence motifs in common with the *link protein* gene. *J. Biol. Chem.* **271:** 4298–4303.

Krusius, T. and E. Ruoslahti. 1986. Primary structure of an extracellular matrix proteoglycan core protein deduced from cloned cDNA. *Proc. Natl. Acad. Sci.* **83:** 7683–7687.

Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: Molecular evolutionary genetics analysis, version 1.01. Pennsylvania State University, University Park, PA.

Kurita, K., T. Shinomura, M. Ujita, M. Zako, D. Kida, H. Iwata, and K. Kimata. 1996. Occurence of *PG-Lb*, a leucine-rich small chondroitin/dermatan sulfate proteoglycan in mammalian epiphyseal cartilage-molecular cloning and sequence analysis of the mouse cDNA. *Biochem. J.* **318:** 909–914.

Lefebvre, V., G. Zhou, K. Mukhopadhyay, C.N. Smith, A. Zhang, H. Eberspaecher, X. Zhou, S. Sinha, S.N. Maity, and B. de Crombrugghe. 1996. An 18-bp sequence in the mouse *proα1(II) collagen* gene is sufficient for expression in cartilage and binds nuclear proteins that are selectively expressed in chondrocytes. *Mol. Cell. Biol.* **16:** 4512–4523.

Lefebvre, V., W. Huang, V.R. Harley, P.N. Goodfellow, and B. de Crombrugghe. 1997. SOX9 is a potent activator of the chondrocyte-specific enhancer of the *proα1(II) collagen* gene. *Mol. Cell. Biol.* **17:** 2336–2346.

Li, W., J.P. Vergnes, P.K. Cornuet, and J.R. Hassell. 1992. cDNA clone to chick *corneal chondroitin/dermatan sulfate proteoglycan* reveals identity to *decorin*. *Arch. Biochem. Biophys.* **296:** 190–197.

Madisen, L., M. Neubauer, G. Plowman, D. Rosen, P. Segarini, J. Dasch, A. Thompson, J. Ziman, H. Bentz, and A.F. Purchio. 1990. Molecular cloning of a novel bone-forming compound: *Osteoinductive factor. DNA Cell Biol.* **9:** 303–309.

McBride, O.W., L.W. Fisher, and M.F. Young. 1990. Localization of *PGI* (*biglycan*, *BGN*) and *PGII* (*decorin*, *DCN*, *PG-40*) genes on human chromosomes Xq13-qter and 12q, respectively. *Genomics* **6:** 219–225.

Nishimura, I., Y. Muragaki, and B.R. Olsen. 1989. Tissue-specific

forms of *type IX collagen-proteoglycan* arise from the use of two widely separated promoters. *J. Biol. Chem.* **264:** 20033–20041.
Nurminskaya, M.V. and D.E. Birk. 1996. Differential expression of *fibromodulin* mRNA associated with tendon fibril growth: Isolation and characterization of a chicken *fibromodulin* cDNA. *Biochem. J.* **317:** 785–789.
Ohno, I., J. Hashimoto, K. Takaoka, T. Ochi, K. Okubo, and K. Matsubara. 1996. The cloning of a cDNA for novel gene expressed in human osteoblast. DDBJ/EMBL/GenBank, accession no. AB000114.
Oldberg, A., P. Antonnson, K. Lindblom, and D. Heinegard. 1989. A collagen binding 59 kD protein (fibromodulin) is structurally related to the small interstitial proteoglycans PG S1 and PG S2 (decorin). *EMBO J.* **8:** 2601–2604.
Rau, W., W. Just, U. Vetter, and W. Vogel. 1994. A dinucleotide repeat in the mouse *biglycan* gene (EST) on the X chromosome. *Mamm. Genome* **5:** 395–396.
Rhodes, C. and Y. Yamada. 1995. Characterization of a glucocorticoid responsive element and identification of an AT-rich element that regulate the *link protein* gene. *Nucleic Acids Res.* **23:** 2305–2313.
Saamanen, A.M.K. 1996. *M. musculus* mRNA for *fibromodulin*. DDBJ/EMBL/GenBank accession no. X94998.
Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4:** 406–425.
Sambrook, J., F. Fritsch, and T. Maniatis. 1989. Molecular cloning: A laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Scholzen, T., M. Solursh, S. Suzuki, R. Reiter, J.L. Morgan, A.M. Buchberg, L.D. Siracusa, and R.V. Iozzo. 1994. The murine *decorin*. Complete cDNA cloning, genomic organization, chromosomal assignment, and expression during organogenesis and tissue differentiation. *J. Biol. Chem.* **269:** 28270–28281.
Shinomura, T. and K. Kimata. 1992. *Proteoglycan-Lb*, a small dermatan sulfate proteoglycan expressed in embryonic chick epiphyseal cartilage, is structurally related to *osteoinductive* factor. *J. Biol. Chem.* **267:** 1265–1270.
Shinomura, T., K. Kimata, Y. Oike, A. Noro, N. Hirose, K. Tanabe, and S. Suzuki. 1983. The occurrence of three different proteoglycan species in chick embryo cartilage: Isolation and characterization of a second proteoglycan (PG-Lb) and its precursor form. *J. Biol. Chem.* **258:** 9314–9322.
Sommarin, Y., M. Wendel, Z. Shen, U. Hellman, and D. Heinegard. 1998. Osteoadherin, a cell-binding keratan sulfate proteoglycan in bone, belongs to the family of leucine-rich repeat proteins of the extracellular matrix. *J. Biol. Chem.* **273:** 16723–16729.
Stolzfus, A., J.M. Logsden Jr., J.D. Palmer, and W.F. Doolittle. 1997. Intron "sliding" and the diversity of intron positions. *Proc. Natl. Acad. Sci.* **94:** 10739–10744.
Sudbeck, P., M.L. Schmitz, P.A. Baeuerle, and G. Sherer. 1996. Sex reversal by loss of the C-terminal transactivation domain of human SOX9. *Nat. Genet.* **13:** 230–232.
Sztrolovics, R., X.-N. Chen, J. Grover, P.J. Roughley, and J.R. Korenberg. 1994. Localization of the human *fibromodulin* gene (*FMOD*) to chromosome 1q32 and completion of the cDNA sequence. *Genomics* **23:** 715–717.
Thomas, J.T., W.A. Sweetman, C.J. Cresswell, G.A. Wallis, M.E. Grant, and R.P. Boot-Handford. 1995. Sequence comparison of three mammalian *type-X collagen* promoters and preliminary functional analysis of the human promoter. *Gene* **160:** 291–296.
Traupe, H., A.M.W. van den Ouweland, B.A. van Oost, W. Vogel, U. Vetter, S.T. Warren, M. Rocchi, M.G. Darlison, and H.H. Ropers. 1992. Fine mapping of the human *biglycan* (*BGN*) gene within the Xq28 region employing a hybrid cell panel. *Genomics* **13:** 481–483.
Ujita, M., T. Shinomura, and K. Kimata. 1995. Molecular cloning of the mouse *osteoglycin*-encoding gene. *Gene* **158:** 237–240.
Vetter, U., W. Vogel, W. Just, M.F. Young, and L.W. Fisher. 1993. Human decorin gene: Intron-exon junctions and chromosomal location. *Genomics* **15:** 161–168.
Wagner, T., J. Wirth, J. Meyer, B. Zabel, M. Held, J. Zimmer, J. Pasantes, F.D. Bricarelli, J. Keutel, E. Hustert, U. Wolf, N. Tommerup, W. Schempp, and G. Scherer. 1994. Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene *SOX9*. *Cell* **79:** 1111–1120.
Xu, J.H., B. Radhakrishnamurthy, S.R. Srinivasan, and G.S. Berenson. 1995. Primary structure of bovine aorta biglycan core protein deduced from cloned cDNA. *Biochem. Mol. Biol. Int.* **37:** 263–272.

Received July 10, 1998; accepted in revised form March 15, 1999.