Structural and Functional Characterization of the Human Decorin Gene Promoter

A HOMOPURINE-HOMOPYRIMIDINE S1 NUCLEASE-SENSITIVE REGION IS INVOLVED IN TRANSCRIPTIONAL CONTROL

(Received for publication, June 17, 1993, and in revised form, August 27, 1993)

Manoranjan Santra, Keith G. Danielson, and Renato V. Iozzo†

From the Department of Pathology and Cell Biology and the Jefferson Cancer Institute, Thomas Jefferson University, Philadelphia, Pennsylvania 19107

Decorin is a leucine-rich, chondroitin/dermatan sulfate proteoglycan which binds collagen and growth factors. We have recently completed the genomic organization of human decorin and discovered two alternatively spliced leader exons, designated exon Ia and Ib, in the 5'-untranslated region. Initial analysis of the sequences upstream to these two exons showed that promoter Ia contained only two GC boxes while promoter Ib contained a CAAT and two TATA boxes in close proximity to the transcription start site. To determine if these 5'-flanking sequences exhibited promoter activity, chimeric chloramphenicol acetyltransferase expression plasmids containing the promoter region of either exon Ia or Ib were transfected into HeLa and MG-63 osteosarcoma cells. The results showed that only the region flanking exon Ib was functional. In vitro transcription assay generated two transcripts of 92 and 82 base pairs (bp) indicating that both TATA boxes could be used. Using stepwise 5' deletion analysis we found that the minimal promoter region at -140 bp from the transcription start site, which contained only the CAAT and the two TATA boxes, exhibited strong promoter activity. When a larger construct containing an additional 800 bp of upstream region was tested, a significant increase in transcriptional activity was observed. Interestingly, this promoter region contained several putative binding sites for ubiquitous factors (AP1, AP5, and NF-kB) and for transforming growth factor-β and a 150-bp homopurine/homopyrimidine element with several mirror repeats. When contained in a supercoiled plasmid, this sequence exhibited sensitivity to endonuclease S1, an enzyme that preferentially digests single-stranded DNA. Precise S1 mapping, obtained by direct sequencing of nine distinct S1-generated clones, revealed that in all cases the borders of the sensitive sequence resided within the pur/pyr segment. We propose that this region of the promoter could adopt an intramolecular hairpin triplex structure in vivo and may play a role in the chromatin organization at the decorin gene locus. In addition, this region was able to up-regulate a minimal heterologous promoter in transient transfection assays. The results show that the structure of the decorin gene promoter is different from that of any other proteoglycan promoter characterized so far and indicate that the pur/pyr segment plays a role in the regulation of gene transcription.

The interest in decorin, a leucine-rich proteoglycan expressed by both vascular and avascular connective tissues, has recently increased by discoveries that this molecule displays diverse functions serving as a biological ligand for both matrix molecules and growth factors (1–3). Specifically, decorin binds collagen type I in vitro thereby inhibiting collagen fibrillogenesis (4, 5) and also binds TGF-β1 removing this potent cytokine from the immediate microenvironment thus neutralizing its biological activity (6). This binding property of decorin to TGF-β has been recently confirmed in an experimental animal model of glomerulonephritis in which infusion of recombinant decorin prevented the fibrotic degeneration of renal glomeruli (7). These biological attributes of decorin suggest that this molecule may also play a role in the development of cancer since constant remodelling of extracellular matrix is required for tumor growth and development (8). Indeed, we have discovered a marked up-regulation of decorin gene expression in the stroma of human colon cancer (9, 10) and have demonstrated that this phenomenon could be reproduced in an in vitro system utilizing mesenchymal cells and colon cancer cells (11, 12). Subsequently, we have shown that the aberrant expression of decorin in colon cancer is associated with a hypomethylation of the decorin gene (13, 14).

We have recently characterized the genomic structure of the human decorin gene and discovered two novel leader exons in the 5'-untranslated region of decorin (15). Over 13 kb of intervening sequence separate these two exons which are alternatively spliced to exon II, a feature also present in avian and bovine decorin genes (15). Because of the existence of two leader exons, we considered the possibility of a two-promoter system regulating cell or tissue-specific expression of decorin. To make possible a detailed molecular understanding of the regulation of decorin gene expression, we investigated the structural and functional activity of the two putative promoter regions (Ia and Ib). Using transient transfection assays and stepwise 5' deletion analysis, we found transcriptional activity only in promoter Ib. A number of positive and negative cisacting elements were identified, including a TGF-β negative element and a homopurine/homopyrimidine stretch in the distal promoter region that was sensitive to S1 nuclease. This region may adopt an intramolecular hairpin triplex in vivo and could play a role in the organization of the chromatin at the transcriptional start site, which contained only the CAAT and the two TATA boxes.
decorin gene locus. The results of this study identify sequences important for the transcriptional control of the decorin gene and offer the opportunity to investigate the abnormal expression of this gene during fibrosis and tumor growth.

EXPERIMENTAL PROCEDURES

Materials—All the reagents were of molecular biology grade. Ribonucleotides \(\left[\alpha-\text{P}\right]\text{CTP} (3000 \text{ Ci/mmol})\), \(\left[\alpha-\text{P}\right]\text{CTP} (400 \text{ Ci/mmol})\), \(\alpha-\text{32S}\)-\text{dATP} \(\left(1000 \text{ Ci/mmol}\right)\), and \(\text{[14C]chloramphenicol} \left(100 \text{ Ci/mmol}\right)\) were obtained from Amer sham Corp. Restriction enzymes were purchased from Promega Corporation ( Madison, WI). PCR products were obtained from Perkin-Elmer Cetus.

Genomic Clones and DNA Sequencing—DEC-10 consisted of a 3.4-kb fragment containing the first exon (Ib), the 5′-flanking DNA, and the second exon (244 bp) of the human decorin gene (15). DEC-1 consisted of a 4.6-kb fragment of the human decorin gene containing an alternative first exon (Ia) and the 5′- and 3′-flanking DNA. DNA isolated from both sources was subcloned into pBluescript KS (Stratagene). Plasmids were sequenced by a modified diodeoxyribonucleotide chain termination method (16) using either polynucleotides primer T3 and T7 or synthetic oligonucleotide primers as described before (17, 18). In addition, automated sequencing of constructs was performed using the Applied Biosystems model 373A DNA sequencing system according to the manufacturer's instructions. Ambiguities were resolved by modifying sequencing reactions or electrophoretic conditions to enhance DNA sequences proximal or distal to the primers. Alignment of nucleotide sequences and comparison with EMBL and NCBI data bases were performed utilizing the programs contained in PC/GENE (Intelligenetics) including CD-ROM data base (release 7.1) and the programs contained in the GCG package of the Jefferson Cancer Institute.

Ribonuclease Protection Assay—Linearized plasmids (2–4 μg) were transcribed with 15 units of T7 or T3 RNA polymerase (Promega) for 1 h at 37 °C in a mixture containing transcription buffer (Promega), 10 mm dithiothreitol, 20 units of Rnase inhibitor (RNasin), 500 μm each of ATP, GTP, UTP, 12 μm CTP, and 50 μCi of \(\left[\alpha-\text{P}\right]\text{CTP}\). The DNA template was then removed by addition of RNase-free DNase (1 unit/μg DNA) and incubation at 37 °C for 15 min. The reaction volume was brought to 100 μl with 10 mm Tris-Cl, pH 7.6, 1 mm EDTA, and 10 μg/ml yeast RNA (2 μg/μl) was added, and the radiolabeled transcript was precipitated with ammonium acetate. Precipitation of riboprobe was repeated twice to ensure removal of unincorporated radiolabeled ribonucleotides. The pellet was washed once with 75% ethanol, 6.1 m sodium acetate (pH 7.0), air dried, and dissolved in hybridization buffer (90% deionized formamide, 10 mm sodium phosphate, pH 6.4, 400 mm NaCl, 1 mm EDTA) (19). Total RNA was isolated from cultured cells or frozen tissues using a guanidinium isothiocyanate method (19). Approximately 50 μg of RNA was precipitated with ammonium acetate, resuspended in 15 μl of hybridization buffer with 10% dextran sulfate, and 50 μl of yeast tRNA (2 μg/μl) was added. The samples were digested with S1 nuclease to a final concentration of 5 U per 5 μl and the reaction was terminated by adding EDTA to a final concentration of 5 mm and heating the reaction at 70 °C for 10 min to inactivate the S1 nuclease. The DNA was then subjected to phenol extraction and ethanol precipitation. Half of this DNA was digested with EcoRI, while the remaining was incubated with deoxynucleotides and Klenow polymerase to generate blunt ends (20). Following ligation the resulting plasmids were sequenced to determine the precise locations of the S1 nuclease cleavage sites.

RESULTS AND DISCUSSION

The discovery of two leader exons in the 5′-untranslated region of the human decorin gene (15) has raised the possibility that a two-promoter system and alternative splicing could be responsible for the presence of heterogeneous transcripts from a single gene. To identify active promoter sequences in the decorin gene, we analyzed the structural and functional activity of the DNA regions upstream and downstream of the two-phosphate-functional site. We provide evidence for promoter activity in the region 5′ to exon Ib only, identify novel structural motifs that may be involved in the decorin gene transcriptional control, and clarify some of the issues raised by recent studies on the cDNA cloning of the 5′ termini of the human decorin.

Complexity of Exon Ia Across Species—Previously, we characterized exon Ia as a 81-bp region which we amplified by anchored PCR from a human embryonic skin fibroblast cDNA library (15). Fig. 1 schematically illustrates the organization of human decorin gene and summarizes the size and homology of exon Ia in a variety of recently described cDNAs from human...
and non-human cells. It is notable that the decorin cDNA sequence from human embryonic skin fibroblasts originally described by Krusius and Ruoslahti (24) contained a highly divergent region at the 5' end. Specifically, the initial 25-bp sequence was not found in any other cDNAs, including cDNAs isolated from murine and avian species, nor was it found in the genomic clone DEC-1 which contained exon Ia. This sequence (Fig. 1B) contains a terminal hexamer (GAATTC) recognized by the restriction enzyme EcoRI and, thus, it may represent a truncated cDNA. Interestingly, the sequence recently reported by Fisher (25) for the human bone exon Ia revealed a longer exon extending to 169 bp. This sequence matched perfectly our genomic sequence in DEC-1 clone, with the exception of the initial 23 bases (Fig. 1). To assess whether there was another exon in the upstream region of the decorin gene, we rescreened all our genomic clones (which include ~10 kb of 5'-flanking region and >13.2 kb of 3'-flanking region as related to exon Ia) using inverse complementary oligonucleotide probes (23-mer) based on the sequence of either the 5' end of human skin fibroblast decorin cDNA (24) or the bone cells exon Ia (25). In either case, all the Southern blotting of our genomic clones failed to reveal any additional reactive band (not shown).

To determine whether the two divergent sequences of exon Ia discussed above represented novel 5' exons, we amplified a AZAP cDNA library from embryonic fibroblasts using anchored PCR (15). In this strategy, we took advantage of the fact the library contained the pBluescript plasmid between the phage arms and thus we used as sense primers the T3 or T7 sequences as antisense primers the 23-mer sequences described above. In the case of the sequence from embryonic lung fibroblasts (24), we had no specific amplification. In contrast, in the case of bone cells (25), several fragments were amplified (Fig. 2A). Specifically, two fragments of ~630 and ~940 bp were predominant (Fig. 2A, lane 3). Southern blotting using the 32P end-labeled 23-mer (the same primer used for amplification) showed a strong reactivity with the amplified bands (Fig. 2B, lanes 2 and 3), but not with 500 bp of decorin cDNA which was amplified at the same time (Fig. 2B, lane 4). To test whether the two major amplified bands were related to decorin (i.e. they could represent a novel 5' untranslated exon of decorin), we purified the two DNA fragments from the gel and used them as probes for Northern blotting of mRNA known to contain decorin. Interestingly, both bands hybridized with a 4.2-kb band, markedly different from the two typical transcripts of ~1.9 and ~1.6 kb found in decorin (15, 24), in both skin fibroblasts and rhabdomyosarcoma cells.

**Fig. 1. Schematic representation of decorin gene organization (A) and of exon Ia in various organisms (B).** A, the exons of the human decorin gene are indicated by roman numerals, with the empty and filled rectangles representing untranslated and translated regions, respectively. The introns are indicated by a continuous thin line and are not in scale. B, the light-shaded rectangles represent the regions of exon Ia which were either identical (for the three human species) or with high homology (85–87%, as indicated) to the human genomic sequence contained in clone DEC-1 (15). The darkly shaded rectangles represent the regions of divergent sequence with the corresponding number of nucleotides and sequence contained within brackets. The numbers on the left represent the bp number reported for each individual exon Ia in human embryonic lung fibroblasts (24), human embryonic skin fibroblasts (15), human bone cells (25), chicken cornea (26), mouse 3T3 cells, and rat uterus (27), respectively.

**Fig. 2. Amplification of 5' decorin cDNA clones by anchored PCR (A), Southern blotting (B), and Northern blotting (C and D).** Panel A depicts an ethidium bromide-stained gel showing a PCR-amplified cDNA library from human embryonic skin fibroblasts using as sense primers the T3 (lane 2) or T7 (lane 3) promoter sequence and as antisense primer the divergent sequence (23-mer) from the 5' end of bone cell decorin cDNA (lanes 2 and 3). Notice the amplification of different size fragments in both cases, with two prominent bands of ~630 and ~940 bp (lane 3), respectively. As the control, a ~500-bp fragment of the 5' end of human decorin (lane 4) was amplified. Lane 1 depicts DNA molecular weight markers with their size in kb on the left. Panel B is a Southern blotting of the gel shown in panel A using as probe the 32P-end labeled 23-mer. Notice that the amplified bands from the cDNA library were strongly reactive (lanes 2 and 3), while the molecular weight markers (lane 1) or the decorin cDNA (lane 4) remained unlabeled. Panel C shows a Northern blotting of mRNA isolated from human fibroblasts (lane 1) or rhabdomyosarcoma cells (lane 2) hybridized under stringent conditions with the 630-bp fragment shown in panel A. Two of the promoters (lane 3) were strongly reactive (lanes 2 and 3). Notice the presence of a single hybridizing band of 4.2 kb, markedly different from the two typical transcripts of ~1.9 and ~1.6 kb found in decorin (15, 24), in both skin fibroblasts and rhabdomyosarcoma cells.
decorin showed no additional 5' clones. Second, PCR amplification of decorin from several cDNA libraries, using exon Ia-specific primers, always yielded fragments of the predicted length and always in frame to exon II (15). Third, as illustrated in Fig. 1, exon Ia sequences with a high degree of conservation (85–90%) were also observed in decorin cDNAs isolated from chicken cornea (26), rat uterus (27), and mouse 3T3 cells.2 Interestingly, in the murine sequence the homology with the human genomic clone extended for 207 bp, thus suggesting that the size of exon Ia may be significantly greater than 80 bp as originally described (15).

Ribonuclease Protection Assay for Exon Ia and Exon II—In order to establish the transcription start site for exon Ia, we performed a series of experiments using primer extension of various RNAs and RNase protection assays. Primer extension often gave products significantly shorter than predicted probably due to premature termination of the reverse transcriptase enzyme in areas of complex RNA secondary structures (22). In contrast, RNase protection assay of RNA from human prostate, a tissue enriched in decorin message (15), gave a protected band of 173 bp (Fig. 3, lane 2) based on the migration of the 32P-labeled DNA standard from φX174 virus. Because it is well established that RNA has a lower mobility than DNA of the same length when run on urea-polyacrylamide gels (22), the corrected size is likely to be ~5% smaller. Using this correction, then, the size of the protected exon Ia is 164 bp, in close agreement with the size reported for exon Ia in human bone cells (25).

Exon II was easily detected in a variety of tissues and cells expressing decorin. Specifically, exon II was detected in prostate tissue (not shown), embryonic lung fibroblasts (Fig. 3, lane 4), human adult kidney (Fig. 3, lane 5), but was not found in prostate cancer PC-3 cells (Fig. 3, lane 6), which do not express decorin,3 or when yeast tRNA was used instead of human RNA (Fig. 3, lane 7). The protected exon II (lanes 4 and 5) gave a value of 255–256 bp, slightly higher than the 244 bases, the actual value obtained by cloning and sequencing the gene in two separate laboratories (15,27). After correction for different migration as described above, the size of the protected exon II was 243 bp, remarkably close (within one nucleotide) to that determined by direct sequencing. Thus, based on these data, we feel that our estimates of the size of exon Ia are accurate and within experimental error.

The 5'-Flanking Sequence to Exon Ia Does Not Show Functional Promoter Activity—Analysis of the putative promoter Ia, a 2.1-kb sequence 5' to exon Ia (15), showed a lack of a CAAT or TATA box as in most eukaryotic promoters. Only a few possible regulatory cis-acting elements were observed, including two GC/SP1 boxes and two direct repeats of 51 bp. The DEC1-CAT construct consistently showed no ability to direct heterologous transcription in transient transfection assays under conditions where the control plasmid driven by the SV40 promoter was highly active (see below). The DEC1-CAT construct was found to be inactive in a number of transfected cells including several human lines of embryonic and skin fibroblasts, MG-63 osteosarcoma cells, and HeLa cells (not shown). The transfection efficiency for each plasmid was tested by cotransfection of each cell with a plasmid coding for the β-galactosidase enzyme under the strong SV40 promoter. The data for each transfection experiment were thus normalized on the activity of the β-galactosidase enzyme. These results, thus, indicate that the 2.1-kb sequence proximal to exon Ia does not function as a decorin promoter. Although the possibility for the existence of an additional exon upstream to exon Ia cannot be totally excluded, the evidence discussed above suggests that no exons related to decorin have been discovered.

A more likely possibility is the presence of a distal promoter several kilobases from the actual transcription start site. This question could be better addressed once the structural and functional activities of the decorin promoter regions from different species are characterized.

The Structure of Promoter Ib: Several Regulatory Cis-acting Sequences—In contrast to the putative promoter Ia, the region 5' to exon Ib (Fig. 4), contained several regulatory cis-acting elements known to play a role in transcriptional activity. Computer-assisted analysis using the GCG FIND and PCGENE EUKPRM programs revealed a number of interesting features. Two TATA-like motifs (TATAAT and TTAAAA) and a
CAAT box were detected within the first 95 bp from the putative cap site (Fig. 4). By analogy with other well-characterized euukaryotic promoters (29), these TATA boxes may serve to fix the start site for decorin transcription (see below). Distal to this region, there were other cis-acting motifs originally described as SV40 enhancer sequences, AP5 and AP1, located at −525 and −813 bp, respectively (for a review, see Ref. 30). Two direct repeats of 18 and 19 bp were found upstream of the AP5 motif. Interestingly, both repeats contained sequences that conform to the consensus site for NF-κB (31). These sequences have been involved in interleukin-1-induced stimulation of transcription in a variety of genes (31). It is noteworthy that interleukin-1 stimulates the synthesis of chondroitin/dermatan sulfate proteoglycans in human synovial fibroblasts (32) and that decorin mRNA levels are increased 3-fold following interleukin-1 treatment of human skin fibroblasts (33).

A TGF-β-negative element was found at position −685. This motif conforms to the consensus sequence GnnTGGGnGa that has been found in the promoter region of transin/stromelysin, and its expression is mediated through the proto-oncogene c-Fos (34). This putative TGF-β-negative element could function as a transcriptional silencer and suppress decorin gene activity in TGF-β-sensitive cells. Although it was originally reported that TGF-β up-regulates dermatan/chondroitin sulfate proteoglycan (35), it is now becoming clear that decorin mRNA is either unaffected or markedly suppressed by TGF-β while biglycan and versican are more likely to be the gene products induced by the cytokine (36–38).

Finally, in the distal 5′ region of the promoter Ib there were two direct repeats of 18 bp which were contained in a 150-bp region composed of homopyrimidine residues (CT) in the coding strand. This region contained mirror repeats and, as demonstrated below, was sensitive to S1 nuclease digestion.

**Determination of the Transcription Start Site of Exon Ib by In Vitro Transcription Assay**—We have previously detected exon Ib in a number of cells and human tissues using either reverse-transcriptase PCR or Northern blotting with exon-specific probes (15). Indeed, using either technique, it appeared that exon Ib is the most widely distributed exon among tissues with an average expression that is about 10 times greater than that of exon Ib (15). In addition, exon Ib sequence has been independently discovered in human adult skin fibroblasts and in decorin cDNA from human (24) and bovine (39) bone cells. In the human bone cells, a much longer exon Ib has been reported (24) that follows exactly our genomic sequence of DEC-1 and extends for about 200 bp including the two TATA boxes. This long transcript, as recently discussed for the osteopontin gene (40), formally actually represent an incompletely processed nuclear transcript present in the RNA preparation that was used in the original cloning of the cDNA library (41).

In spite of wide expression of exon Ib, both primer extension and RNase protection assays failed to provide consistent results. A likely explanation for this is that this region has the possibility of a hairpin loop structure similar to that observed in the c-myc gene (42). To circumvent this problem, we utilized an *in vitro* transcription assay using HeLa nuclear extracts and, as template, the construct containing promoter Ib and the entire exon Ib (cf. Fig. 4). The results, using commercially available purified nuclear extracts (Fig. 5, lane 1) or HeLa extracts purified by us as described before (43) (Fig. 5, lane 2) gave two transcripts of 98 and 86 bp, respectively. If one corrects for the overestimate due to the difference between RNA and DNA mobility as above, then two possible transcription start sites of 93 and 82 bp are detected for this decorin promoter. A closer analysis of the two transcripts revealed that the 92-bp transcript was the predominant one and was remarkably similar in size to exon Ib (81 bp) which we previously PCR amplified from a human fibroblast cDNA library (15). Of note, the minor transcript of 93 bp has been found in a cDNA isolated from human adult skin fibroblasts (4). These data are compatible with the notion that the two transcripts are driven by RNA polymerase and that the two TATA boxes are responsible for the presence of two separate transcription start sites. A detailed analysis of HeLa transcription factors that bind to TATA boxes have revealed a remarkable conservation in DNA motifs between human and yeast (44). The two TATA-like motifs located in the upstream region of the decorin gene are indeed highly conserved and active in a variety of euukaryotic and prokaryotic promoters (44). Significantly, mutational analysis of the consensus TATA box sequence (TATAAA) has shown that motifs identical to those found at −66 (TATAAT) and −50 (TTTAAA) of the decorin promoter maintain 21 and 54% of the transcriptional activity as compared to the wild type (44). This is in agreement with our finding that the 92-bp transcript is more actively transcribed than the 93-bp transcript (Fig. 5). It has also been demonstrated that the presence of a T or an A instead of a G at the seventh position of TATA elements enhances transcription (44). It is noteworthy that the TATA-like box of decorin at −66 contains a G as the seventh nucleotide, while the one at −50 contains an A, thus further indicating that the latter TATA element should exhibit more transcriptional activity. Finally, because the studies discussed above (44) have used HeLa...
A Homopurine/Homopyrimidine Region in Promoter Ib Is Sensitive to S1 Nuclease Digestion—The 5'-flanking region to exon Ib (Fig. 4) contained a unique 150-bp segment featuring pyrimidine deoxynucleotides (CT) in the coding strand and purine deoxynucleotides (AG) in the non-coding strand. In addition there were mirror repeats in the 3' end side of the homopyrimidine stretch (Fig. 4). Pur/Pyr sequences have recently attracted significant interest because they can adopt a novel hairpin triplex referred as the H form of DNA (45, 46). In this unusual structure, half of the pyrimidine strands form a normal Watson-Crick duplex with the corresponding homopurine strand, while the remaining (palindromic) homopurimidine stretch folds back and binds into the major groove of DNA via Hoogsteen base pairing. The peculiar feature of such a structure is that half of the homopurimidine strand remains unmatched and thus becomes accessible to cleavage by single strand-specific endonucleases such as S1 nuclease (46). We, therefore, inspected the 5'-flanking region to decorin Ib for its sensitivity to S1 endonuclease. In these experiments, we banded supercoiled plasmid DNA twice by CsCl density gradient ultracentrifugation to avoid any shearing or nicking of the DNA which would preclude any meaningful conclusions since nicked DNA is highly sensitive to S1 nuclease (22). A typical DNA preparation of this construct migrated as a single band of ~3.1 kb in a supercoiled conformation (Fig. 6, lane 2). When this DNA preparation was incubated with S1 nuclease at pH 4.5, it was rapidly converted into a linearized plasmid which now migrated as a ~6-kb fragment (Fig. 6, lane 4). To identify the S1-sensitive site, we sequentially digested this plasmid with the restriction endonucleases EcoRI and HindIII, which gave the predicted fragments of 2.9, 1.9, and 1.3 kb, respectively (Fig. 6, lane 3). If the S1-sensitive site were located in the homopyrimidine stretch (i.e. at the beginning of the insert as illustrated in the top panel of Fig. 6), one would expect that digestion with S1 nuclease followed by digestion with EcoRI should specifically cleave only in the 5' of the 1.3 kb band. The results supported this notion since S1/EcoRI treatment of the construct reduced the 1.3 kb band by about ~100 nucleotides (Fig. 6, lane 5, labeled by arrowhead). These results are consistent with the presence of S1-hypersensitive sites in supercoiled DNA containing the decorin promoter region. S1 sensitivity required supercoiled plasmid DNA since predigestion with either EcoRI or HindIII prevented S1 digestion (not shown).

To establish more precisely the sites of S1 sensitivity in the decorin promoter, we briefly digested the construct with S1 nuclease and the linearized plasmids were recovered from the gel. The ends were filled with Klenow polymerase and blunt end-ligated. A number of clones were isolated and sequenced with a reverse primer 3' to the homopyrimidine stretch. The results showed that in all nine clones isolated (Fig. 7) there were deletions varying in size between 27 and 95 bp, which were located exclusively within the homopurine/homopyrimidine region.

Significance of the Homopurine/Homopyrimidine Region in the Decorin Promoter—It is now well established that triplex (H-forms) of DNA develop by the unpairing and folding back of the pyrimidine strand, which forms Hoogsteen hydrogen bonds in parallel orientation to the opposite half of the purine strand (46). This creates a triplex stem of CGC and TAT triplets, a loop of pyrimidine residues, and a single stretch of half of the purine mirror repeat (45, 46). In general, low pH, negative supercoil and increasing length of the pur/pyr segment act independently to stabilize intramolecular triplex formation. Because cytosine residues are protonated at low (<5.0) pH, it has been proposed that these residues are involved in Hoogsteen pairing for which protonation is required (45). Interestingly,
pur/pyr structures frequently occur upstream from eukaryotic genes and in recombination sites. Similar repetitive structures have been found in a number of proto-oncogenes including the epidermal growth factor receptor (47, 48), the c-Ets-2 (49), and c-Ki-ras (50, 51). In addition, extracellular matrix proteins such as chicken collagen a2(I) (52, 53) and a2(VI) (54), bovine osteonectin (55), and bone sialoprotein (56) have similar structures in the 5'-flanking region of their promoters.

What could be the role of such repetitive elements near the promoter region of genes? Although not completely established, current evidence indicates that pur/pyr stretches are directly cis-active motifs that bind specific protein moieties. For example, in Drosophila a DNA-binding protein with an affinity for the GAGA consensus sequence similar to that present in the noncoding strand of the decorin pur/pyr segment regulates the transcriptional activity of the Ultrabithorax gene (57). It is a relatively abundant sequence-specific DNA-binding protein and has been shown to bind to a number of promoter regions in Drosophila (58). This protein recognizes a sequence C/AGAGAGAGC which differs from the GAGA factor described by Young and collaborators (55, 56, 59) that recognizes GGGA or GGA repeats. In addition, another GAGA-binding protein has been recently characterized in Drosophila that binds to purine-rich regions in a homeobox gene designated Krüppel (60) and shows homology to the osteonectin factor (56). Finally, several proteins from Drosophila nuclei bind to regions of alternating C and T residues present in the promoters of the heat shock genes hsp70 and hsp26 and the histone genes his3 and hsi3 (61).

**Reporter Gene Analysis of the Decorin Promoter Ib and 5' Deletion Constructs**—To prove that the putative promoter region upstream of exon Ib does in fact exhibit transcriptional activity, we performed reporter gene analysis using various decorin promoter-CAT fusion plasmids. Fig. 8 diagrams the decorin promoter-CAT fusion plasmids used in these experiments. Reporter gene analyses were performed using several controls and at least in two types of cells, HeLa (primarily) and MG-63 osteosarcoma cells. In each transfection experiment we used a CAT fusion plasmid driven by the SV40 promoter and enhancer (pSV2-CAT) as a positive control and a promoterless

---

**Fig. 6.** S1 endonuclease sensitivity of the homopurine/homopyrimidine region of the decorin promoter. The top panel is a graphic representation of the construct containing the 1.3-kb exon Ib and 5'-flanking region (indicated by DEC) fused to the 5' end of CAT reporter gene (indicated by CAT). EcoRI, HindIII, and S1 nuclease sensitive sites are indicated by E, H, and S1, respectively. The figure represents a 0.8% agarose gel stained with ethidium bromide of supercoiled plasmid DNA treated with different enzymes. Undigested supercoiled (lane 6), DNA digested with EcoRI and HindIII (lane 3), DNA digested with S1 nuclease (lane 4), DNA digested with S1 nuclease and EcoRI (lane 5). Lanes 1 and 6 represent DNA size markers, while the numbers on the right represent their respective size in kilobases. The white arrowheads mark the transformation of the 1.3-kb insert (lane 3) into a ~1.1 kb (lane 5) upon cleavage with S1 nuclease of the homopurine/homopyrimidine region.

**Fig. 7.** Fine mapping of the S1 nuclease-sensitive sites within the homopurine/homopyrimidine region of the decorin promoter. The plasmid DNA containing the decorin promoter Ib inserted into the PUC-CAT vector (top panel) was linearized by a brief treatment with S1 nuclease (cf. Fig. 6, lane 4) and recovered from the gel. The protruding termini were filled with nucleotide triphosphates using the Klenow polymerase. Following blunt end ligation of the plasmids, a number of clones were isolated and sequenced. The sequence data showed that only the homopurine/homopyrimidine region of the promoter was sensitive to S1 nuclease. The top sequence is the homopurine/homopyrimidine region of the decorin promoter in the original supercoiled plasmid with the arrows indicating the cleavage sites. The dotted lines indicate the portions deleted from the homopurine/homopyrimidine region after S1 nuclease digestion.

**Fig. 8.** 5' deletion constructs used for transient transfection assays. Schematic representation of the stepwise 5' deletion constructs used to test the functional activity of the decorin promoter Ib in transient transfection assays. Numbers on the left indicate the 5' end of the promoter fragment relative to the transcription start site (+1). Each of the constructs contains a unique 5' terminus of the promoter (empty bar) and is fused to the CAT gene at position +35 relative to the transcription start site in the noncoding exon Ib sequence (shaded bar). The regulatory motifs of the decorin promoter are schematically represented at the top by various symbols whose key is provided at the bottom left. See the text for additional details.
CAT plasmid as a negative control. In addition, the cells were cotransfected with a plasmid containing the 3′-galactosidase gene driven by the same strong viral promoter (pSV40-βGAL), and the transfection efficiency for each reaction was normalized for β-galactosidase activity. The relative activity of eight constructs of five independent experiments is summarized in Fig. 9. Although strong promoter activity was detected with the -140 bp construct, the smallest construct containing the two TATA boxes and the CAAT box (cf. Fig. 4), the activity was increased when the distal promoter region between -879 and -983 was included. Interestingly, this region contained the entire homopyrimidine segment, thus indicating that this pur/pyr region has enhancer activity. A reproducible, although moderate, decrease in functional activity of the promoter was observed when the construct including the TGF-β-negative element (the region extending to -735, Fig. 9) was used. These data suggest that this region may represent a negative cis-active element as described in stromelysin (34) and provide a molecular correlation of previous studies showing that decorin expression and corroborates previous observations on modula-
tion of decorin gene expression by growth hormones and cytoto-
kines.

Conclusions—In this paper we have characterized the structural and functional activity of human decorin gene which encodes a proteoglycan involved in the regulation of matrix assembly and cytokine binding. The results have shown that the organization of human decorin promoter is unusual and differs significantly from that of any other proteoglycan gene promoter so far described including the serglycin (64), biglycan (65), syndecan (66), versican (38 and perlecain) (67) genes, respectively. We have demonstrated that, although decorin contains two alternatively spliced leader exons with flanking putative promoter regions, only one of these regions upstream to exon 1b exhibits strong transcriptional activity. The elucidation of key regulatory elements in the active decorin promoter provides new insights into the understanding of the control of decorin gene expression and corroborates previous observations on modulation of decorin gene expression by growth hormones and cytokines.

Acknowledgments—We thank Drs. L. A. Beavan, K. L. Dreher, L. Fisher, and J. F. Woessner for sharing unpublished information and Drs. B. Calabretta and A. Sala for kindly providing plasmid constructs.

REFERENCES

1. Gallagher, J. T. (1989) Curr. Opin. Cell Biol. 1, 1201–1218
2. Ruoslhati, E., and Yamaguchi, Y. (1991) Cell 64, 867–869
3. Kjellen, L., and Lindahl, U. (1991) Annu. Rev. Biochem. 60, 443–475
4. Vogel, K. G., Paulsson, M., and Heinegård, D. (1984) Biochem. J. 223, 587–597
5. Vogel, K. G., and Trotter, J. A. (1987) Collagen Relat. Res. 7, 105–114
6. Yamaguchi, Y., Mann, D. M., and Ruoslhati, E. (1990) Nature 346, 281–284
7. Border, W. A., Noble, N. A., Yamamoto, T., Harper, J. A., Yamaguchi, Y., Pier-
schbacher, M. D., and Ruoslhati, E. (1992) Nature 356, 361–364
8. M. F. Naso and R. V. Iozzo, manuscript in preparation.
