The members of the syndecan family are temporally and spatially expressed heparan sulfate proteoglycans of various tissues, where they mediate extracellular influences on cell morphology and behavior. Functional characterization of the mouse syndecan-1 promoter was carried out in order to elucidate the mechanisms involved in the maintenance of the high transcription levels of syndecan-1 gene in various epithelia. For that 9.5 kilobase pairs of the upstream region of mouse syndecan-1 gene were cloned, sequenced, and used to prepare chimaeric constructs with a reporter gene followed by transient or stable transfections into NMuMG epithelial cells. In NMuMG cells, cultured either in the presence or absence of serum, the 2.5-kilobase pair promoter region resulted in the constitutive transcription activity, whereas in 3T3 cells the serum depletion decreased the promoter activity significantly. Deletion of the upstream sequences to \(-\)437 base pairs relative to the translation initiation site had little effect on this promoter activity. Further deletion to \(-\)365 base pairs removed three GT boxes and slightly increased the promoter activity, whereas the deletion of the next two GC boxes (to \(-\)326 base pair) reduced the promoter activity dramatically. All of the GC or GT box sequences bound the same set of Sp1-like nuclear proteins in gel shift assays. Nuclear protein binding was also demonstrated around both of the most intense transcription initiation sites. Mutation of these regions separately resulted in total loss of transcription initiation from the deleted site and decreased the promoter activity in relation to the intensity of the abolished start site. This indicates that the transcription initiation of the syndecan-1 gene is directed through initiator-like elements directly overlapping the start sites, as shown for several TATA-less housekeeping and growth regulated genes.

We assume that the constitutive high level gene expression in epithelial cells is achieved by the proximal promoter, which is controlled by members of Sp1 transcription factor family.

The syndecans are a family of integral-membrane proteoglycans. They take part in the regulation of cell morphology and behavior by conveying the extracellular information to cells. The syndecans share a common domain structure, first described to murine syndecan-1 (1). The extracellular domains of these proteins contain attachment sites for glycosaminoglycan side chains, which may be composed of either heparin or chondroitin sulfate (2). The intracellular domain is highly conserved between all four known members of the syndecan family (for review, see Refs. 3–5) and apparently contains signals for the proper localization of the molecule within polarized epithelial cells (6). Syndecans can simultaneously bind both structural proteins of the extracellular matrix and heparin-binding growth factors, such as basic fibroblast growth factor (7). Indeed, the presence of heparin or heparan sulfate seems to enhance the signal transduction by basic fibroblast growth factor (8, 9). Interestingly, however, forced expression of syndecan-1 in 3T3 cells down-regulates the growth response to basic fibroblast growth factor (10). Therefore, syndecan-like molecules may promote but also antagonize growth factor action (for review, see Ref. 11).

Each member of the syndecan family has a specific pattern of expression (12). Syndecan-1 expression is restricted mainly to epithelia in adults, but during embryonic development it is temporarily expressed at high levels in proliferating and condensing mesenchymes, e.g. in the development of teeth (13), limbs (14), kidneys (15), and lungs (16). Likewise, keratinocytes in healing wounds express enhanced levels of syndecan-1 (17).

The role of syndecan-1 in the control of cell growth and morphology is illustrated by its altered expression in clinical malignancies and experimental cell culture models of transformation. First, in steroid-regulated S115 mammary epithelial cells, testosterone-induced transformation is associated with the loss of syndecan-1 expression, while the non-transformed, epitheloid phenotype, together with organized actin cytoskeleton, and normal growth are restored in cells genetically engineered to express syndecan-1 in the presence of the hormone (18, 19). Second, decreased syndecan-1 expression is correlated with poor differentiation status of UV-induced skin tumors in mice (20) and tumor formation by transformed keratinocytes in nude mice (21). Third, syndecan-1 expression is restricted to myeloma tumors with a well-differentiated, i.e. less aggressive phenotype (22). Finally, patients with syndecan-1 positive squamous cell carcinomas have a more favorable overall and recurrence-free prognosis than patients with syndecan-1 negative carcinomas (23).

The unique developmental expression of syndecan-1 and its loss in several neoplasias prompted us to characterize the structure of the syndecan-1 gene and its transcriptional regulation. In a previous paper we reported the complete structure and nucleotide sequence of the murine syndecan-1 gene including also the first 1-kb upstream region (24). We have now...
sequenced a further 8.5-kb fragment of this upstream region, characterized the functional regions of the proximal promoter, mapped the protein-DNA interactions of the regions required for high level expression, and analyzed the remainder of the gene for putative enhancer or silencer elements. All this data suggest that the proximal promoter of the syndecan-1 gene is a major regulatory element for syndecan-1 expression in epithelial cells and is controlled by members of Sp1-transcription factor family.

MATERIALS AND METHODS

Isolation of the 5′-Region of the Gene and DNA Sequencing—To sequence the 5′-region of the mouse syndecan-1 gene to -9.4 kb, the XbaI/BamHI (Xb25) fragment and the two following XbaI fragments (Xb27, Xb32) from the third upstream clone cDNA1 (21) (24) were subcloned into the pBluescript KS M13(-) vectors (Stratagene). The subclone Xb7 (-5063 to -4376) was sequenced from double stranded template, but the subclones Xb25 ( -4375 to -387) and Xb4 (-9422 to -5064) were further subcloned into M13mp18 and M13mp19 vectors (Boehringer Mannheim) prior to sequencing. The translation start site was numbered as +1. Sequencing was performed by the dideoxy chain termination method (26) using the Sequenase™ kit (U.S. Biochemical Corp.) and [α-35S]dATP (Amersham). Sequence data base comparisons were made with the Wisconsin package (Genetics Computer Group, Inc.).

Plasmid Constructs—Various promoter fragments ending to the XhoI site (-271 to -137) in the 5′-untranslated region of the exon I were cloned into the polylinker of pCAT-basic (Promega) or pCAT-basiZ/II, upstream from the choloramphenicol acetyltransferase (CAT) gene. The pCAT-basicZ vector was prepared to facilitate subcloning by adding the polylinker region SpHil/HindII of pGEMZf(+)/(Promega) into the polylinker region of pCAT-basic. Gene fragments HindII/XhoI (-2528 to -137), Stul/XhoI (-1023 to -137), and DraI/XhoI (-830 to -137) were ligated into the polylinker of pCAT-basiZ/II, and the constructs were designated as p-2.5CAT, p-1.0CAT, and p-830CAT, respectively. The polylinker region between the promoter fragment and the CAT gene was deleted from these constructs by XbaI/Xhol digestion. To generate the p-271CAT construct (–271 to –137), the PstI fragment covering nucleotides from –271 to +54 bp was first ligated into pCAT-basic, after which the 3′-end of the fragment was deleted by Xhol/XbaI digestion.

The promoter fragments for constructs p-492CAT (–492 to –95), p-437CAT (–437 to –95), p-365CAT (–365 to –95), p-351CAT (–351 to –95), p-326CAT (–326 to –95), and p-289CAT (–289 to –95) were generated by polymerase chain reaction (PCR). All of these fragment sequences share the downstream primer 5′-dTTGCCCTAGACATTCTCGG-3′ located in the 5′-untranslated region of exon T. A XbaI site (underlined) was incorporated into this primer. The reaction conditions were chosen according to the manufacturer’s recommendations (Perkin Elmer). PCR products were first cloned into the pGEM-T vector (Promega) and then sequenced from both ends to ensure the correct position of the upstream and downstream template, and then the orientation of the insert. Insertions from pGM-T vectors were excised by SpHil/XbaI digestion and cloned into the SpHil/XbaI site of pCAT-basic. Finally, all of the PCR-made inserts of these CAT constructs were sequenced using the Sequenase™ kit and synthetic oligonucleotide primers. One A to G mutation (at position –265) was found in the construct p-289CAT. This mutated sequence was not located in a promoter region B from the promoter. The TATA sequence TTTAT-271 to 9422 of construct p-2.5CAT was found in the fragment was deleted by PCR amplification was used to delete the TATA sequence and foot-}

Northern Blot—RNA was isolated from NMU MG and 3T3 cells by the single-step method using guanidium thiocyanate-phenol-chloroform extraction (30). RNA samples were size separated by electrophoresis in 1% agarose-formaldehyde gel and transferred to Hybond-N membranes (Amersham). The filters were prehybridized, hybridized, and washed as described by the manufacturer. The random prime labeled (Promega) mouse syndecan-1 DNA (2.1 kbp) and the rat glyceraldehyde-3-phosphate dehydrogenase (31) were used as probes to quantitate the syndecan-1 mRNA. Hybridization signals were quantitated by MCID imaging analyzer (IMAGING Research Inc.).

Probes for DNA I Footprinting—Two DNA fragments containing the proximal promoter regions of the mouse syndecan-1 gene were used for footprinting analysis. To generate these two fragments, the 5′-end of the construct p-830CAT was labeled with 3′-end deoxyribonuclease (Smal) from the promoter and the enzymes in the gel, transferred to Hybond N membranes, and hybridized using the promoter and 3T3 cells were pooled together and the relative copy number of the CAT constructs was determined by Southern hybridization using the 563-bp long XbaI/NcoI fragment from the vector pCAT-basic (Promega) as a probe.

Northern Blot—RNA was isolated from NMU MG and 3T3 cells by the single-step method using guanidium thiocyanate-phenol-chloroform extraction (30). RNA samples were size separated by electrophoresis in 1% agarose-formaldehyde gel and transferred to Hybond-N membranes (Amersham). The filters were prehybridized, hybridized, and washed as described by the manufacturer. The random prime labeled (Promega) mouse syndecan-1 DNA (2.1 kbp) and the rat glyceraldehyde-3-phosphate dehydrogenase (31) were used as probes to quantitate the syndecan-1 mRNA. Hybridization signals were quantitated by MCID image analyzer (IMAGING Research Inc.).

Probes for DNA I Footprinting—Two DNA fragments containing the proximal promoter regions of the mouse syndecan-1 gene were used for footprinting analysis. To generate these two fragments, the 5′-end of the construct p-830CAT was labeled with 3′-end deoxyribonuclease (Smal) from the promoter and the enzymes in the gel, transferred to Hybond N membranes, and hybridized using the promoter and 3T3 cells were pooled together and the relative copy number of the CAT constructs was determined by Southern hybridization using the 563-bp long XbaI/NcoI fragment from the vector pCAT-basic (Promega) as a probe.
and G from –415 to –392 bp (5’-dCTAGGAGGCGTTAGGAGGGG-
TGT-3’). The oligonucleotides were synthesized with an Applied Bio-
systems automatic DNA synthesizer. The sense strands were end-
labelled with [γ-32P]ATP (Amersham) by T4 polynucleotide kinase (Promega) and annealed to the complementary oligonucleotide. The double stranded probes were separated from single stranded material by polyacrylamide gel electrophoresis. The probes were eluted from the gel slices in 0.3 M sodium acetate, phenol/chloroform extracted, and ethanol precipitated.

Nuclear Extracts and DNase I Footprinting—Nuclear proteins were extracted from NMuMG and 3T3 cells as described by Ausubel and co-workers (27). DNase I footprinting was carried out to map the nu-
clear protein binding sites in the proximal promoter region. 50,000 cpm of the labeled DNA fragments (described above) were first incubated with 5–10 μg of crude nuclear extract and 2 μg of poly-(dl-dC) (Boeh-
ringer Mannheim) in a buffer containing 10 μg Tris (pH 8.0), 5 μm MgCl2, 1 μm CaCl2, 2 μm dithiothreitol, 50 μM bovine serum albu-
mion, and 100 μM KCl in a final volume of 50 μl. After 10 min incubation in room temperature, 0.2 or 2 μl of DNase I (Boehringer Mannheim) was added. In the control reaction 10 μl of 0.375 M NaCl was added instead of nuclear extract and 0.025 units of DNase I were used. The reactions were stopped by phenol/chloroform extraction after 2 min incubation at room temperature. A chemical G + A sequencing ladder (32) was run alongside the digestion products in a 6% sequencing gel to deduce the nucleotide sequence in areas protected from DNase I.

Gel Mobility Shift Assays—Nuclear extracts (usually 1–3 μl of either un-dialyzed or dialyzed extract corresponding to 5–15 μg of protein) were incubated for 10–20 min at room temperature with the labeled oligonucleotide probe (10,000 cpm; about 1 ng of DNA) in a buffer consisting of 20 μl HEPES pH 7.9, 100 μM KCl, 0.2 μM EDTA, and 20% glycerol. In all binding assays, 2 μg of poly-(dl-dC) (Boehringer Mannheim) was added as a nonspecific competitor. The complexes were resolved by electrophoresis in 4.5% nuclease resistant gels using 50 μm Tris, 400 μm glycine buffer (pH 8.5). Indicated amounts of unla-
beled specific and nonspecific competitors were used with the oligonu-
cleotide probes to distinguish sequence-specific interactions. Consensus binding site oligonucleotides for Sp1, AP-2, and NF-κB were purchased from Promega. Polydornal peptide antibody (Santa Cruz Biotechnologies) was used to demonstrate the presence of Sp1 protein in the shifted complexes, by formation of high molecular weight super-
shifted complexes. As control antibodies, rabbit IgG (Sigma) and a polydornal peptide antibody against human neurofibromin (anti-P111; kindly provided by Dr. Jorma Hermonen) were used. Antibodies were added to the binding reactions to a final concentration of 1 or 10 μg/ml 10 min before addition of the probe and the reactions were further incubated for 10 min at room temperature before loading into the gel.

Primer Extension Analysis—Primer extension was done using oligo-
nucleotide (5’-AGCTCTGAAAATCTCCGCAAGTCTACAGTC-3’) hy-
bridizing to the 5’-end of the CAT gene. The primer was end-labeled with [γ-32P]ATP (Amersham) by T4 polynucleotide kinase (Promega) after which it was hybridized to 50 μg of total RNA (30 μg of 
65°C in 0.15 M KCl, 10 μl Tris (pH 8.0), 1 μl EDTA). After hybrid-
ization the nucleic acids were ethanol precipitated and resuspended in 26 μl of water and 9 μl of 5 x Moloney murine leukemia virus reverse transcriptase buffer (Promega). 1.5 μl of 5 x DNA template mixture, 6.5 μl of 1 mg/ml actinomycin D (Sigma), and 2 μl (400 units) of Moloney murine leukemia virus reverse transcriptase (Promega) were added and the reactions incubated for 60 min at 42°C. After that 105 μl of RNase reaction mixture (100 μl/ml salmon sperm DNA, 20 μg/ml RNase A) was added and the incubation was continued for 15 min at 37°C. The reactions mixtures were phenol-chloroform extracted, ethanol precipi-
tated, and analyzed in 6% sequencing gel.

RESULTS

Nucleotide Sequence of the 5’-Flanking Region of the Mouse Syndecan-1 Gene—The nucleotide sequence of the mouse syn-
decan-1 gene has previously been reported to –2.4 kb (24, 25). This 5’-flanking region contains a putative TATA box sequence and consensus binding sites for several transcription factors, such as Sp1, AP-2, MyoD, and NFκB. Three major transcrip-
tion initiation sites have been described to locate around the putative TATA box sequence (–250 bp) at positions –223, –249, and –280 bp (25). In addition, three minor sites have been suggested to be dispersed in a wide area upstream of the gene (–456, –506, –831 bp) (24). The translation start site has been numbered as +1 (25), and this numbering is also used in this article. We have now sequenced the upstream region of the gene to a Xbal site at –9.4 kb. A comparison of this upstream sequence with all other sequences in GenBank revealed the presence of several regions of high homology to repetitive se-
quences of other genes (between nucleotides –7762 and –2702).

Identification of the Mouse Syndecan-1 Promoter—In order to char-
acterize the regions regulating the transcription activity of the gene, a series of chimaeric constructs containing differ-
ent lengths of the 5’-flanking region of the mouse syndecan-1 gene were fused to the CAT reporter gene and tested in transient and stable transfection experiments. The 3’-ends of all inserts in these constructs extended to the 5’-untranslated region of the first exon 137 or 95 bp upstream from the ATG codon. Two different cell lines were used in these studies. The epithelial NMuMG cells expressed constantly high levels of syndecan-1 mRNA both in normal and serum-depleted culture conditions, whereas in the mesenchymal NIH3T3 cells the syn-
decan-1 mRNA levels decreased to one-third from the original, after the cells were cultured in serum-deprived conditions (Fig. 1A).

We first analyzed the longest proximal promoter construct p-2.5CAT (–2528 to –137) both in NMuMG and 3T3 cells cultured in the presence and absence of serum, using stably transfected cells. The serum depletion was done by culturing clones 2 days before harvest in 2% carboxymethyl-Sephadex eluted FCS. In both cell lines the CAT construct followed closely the expression of the endogenous syndecan-1 gene. In NMuMG cells the promoter activity was high in both culture conditions. The CAT activity of NMuMG cell extracts after serum depletion was even 65% higher than the activity of extracts from normal culture conditions (Fig. 1B). On the other hand, in the 3T3 cell clone the serum depletion decreased the CAT activity of the cell extracts to one-half of the activity seen in normal culture conditions (Fig. 1B). The suppression of the promoter activity is actually more dramatic, because the half-
life of the CAT enzyme is over 50 h. According to these results the 2.5-kb syndecan-1 promoter fragment seemed to be responsible for the basal transcription activity of the syndecan-1 gene both in NMuMG and 3T3 cells.

The more detailed characterization of the promoter region was done by transiently transfecting a series of promoter con-
structs into NMuMG cells. The deletion of the 5’-flanking re-

2 The nucleotide sequence of the mouse syndecan-1 gene in EMBL data bank (msyn-1gen; accession number Z22532) has now been sup-
plemented with the upstream sequence to –9.4 kb described in this article.
exhibited about 30% suppression of CAT activity in both cell lines (data not shown). These results confirmed the assumption that the constitutive expression of syndecan-1 in NMuMG and 3T3 cells is achieved by a highly active proximal promoter, which could be, however, negatively controlled by some upstream elements not discovered yet.

DNase I Footprinting Analysis of the Proximal Promoter—DNase I footprinting analysis was used to characterize the protein/DNA binding sites within the promoter. We focused on the promoter fragment of the construct p-437CAT, which contained all functional promoter sequences according to our transfection results. The nuclear extracts used in these studies were prepared from NMuMG cells. The shorter probe XB extended from the Xho site (−137) to the BamHI site (−388), and the longer probe XD to the DraI site (−830). When the probe XB was used in the DNase I footprinting analysis, a protected area covering the putative TATA sequence and neighboring 15 bp downstream from it was seen (FP-A in Fig. 3, A and C). To detect the protein binding sites upstream from the TATA sequence, the longer probe XD was used. DNase I footprinting analysis with this probe resulted in the protection of at least five distinct regions (FP-C, -D, -E, -F, and -G in Fig. 3B). Two of these footprints (E and F) were surrounded by DNase I hypersensitive sites, indicating that protein binding changed the structure of the DNA (Fig. 3B). Footprints D (−350 to −337) and E (−362 to −354) located at the GC box sequences (GGGGCCGGG), and footprints C (−296 to −284) and F (−383 to −371) covered the GT box sequences (GGGTTGGG) (Fig. 3C). GT boxes have been shown to bind transcription factor Sp1 and other members of the Sp1 multigene family (33). The longest footprint G (−434 to −397) also covered a GT box sequence and the sequence GAGCGTGG matching one nucleotide (tymidine) to the consensus DNA binding sequence of Sp1 (Fig. 3C). In some of our footprinting gels we also saw a short protected region about 20 base pairs upstream from the TATA sequence (FP-B in Fig. 3C), which extended from −281 to −271 (data not shown). According to these results it was assumed that the TATA sequence as well as all the GC and GT boxes bind nuclear proteins. In addition, the region covering one of the transcription initiation sites upstream from the TATA sequence footprinted weakly.

Gel Mobility Shift Analysis of the Proximal Promoter—To characterize the complexes seen in footprinting analysis the following assays were carried out with synthetic oligonucleotides as probes. Double-stranded oligonucleotides were synthesized and named according to the sequences protected by bound proteins in the DNase I footprinting analysis (Fig. 3). GC box sequences were located within probes D (5′-dACCCAGGGCGGGCCCCGAGGGGTGG), and E (5′-dCCTGCAAGGGCGGGGGGCAACCCGAGGGGTGG), and a GT box sequence within probe F (5′-dDGATCCTGGGCGGGGGCGTGGC). Nuclear extracts from either NMuMG or 3T3 cells produced the same set of retarded complexes with all of these probes (Fig. 4). At least three separate bands could be seen, of which the two slowest migrating bands were equally intense, whereas the fastest migrating band was weaker. To confirm that the bound complexes were the same with all probes competition experiments were carried out. When oligonucleotide E was used as a probe, the unlabeled double-stranded oligonucleotides D, E, and F were able to compete similarly, in a dose-dependent fashion, with the interaction between the probe and nuclear proteins (Fig. 5). On the other hand, oligonucleotide B (covering footprinted region B), which did not contain any GC or GT boxes, did not show any competition in this experiment (Fig. 5).

Because the proximal promoter included several binding se-
quences for the transcription factors Sp1, AP-2, and NFκB, we also carried out competition assays with their double stranded consensus binding site oligonucleotides. Oligonucleotide E was used as a probe in these experiments. The binding of all complexes were clearly inhibited by the Sp1 binding site oligonucleotide (ATTGATCGGGGCGGGCGAGC), but not by the AP-2 (GATCGAAGCTGACCGCCGCCCGT) or NFκB (AGTTCGATCGGGGCGGGGCGAGC) binding site oligonucleotides (Fig. 6A). These results indicated that the nuclear protein binding to the promoter regions D, E, and F share similar binding specificity through Sp1-like binding domains. This was supported by the fact that the retarded complex generated by purified Sp1 protein co-migrated in the mobility shift assays with the uppermost complex produced by the nuclear extract (Fig. 6B). Moreover, in all supershift experiments using the polyclonal anti-Sp1 antiserum with probes D, E, and F, and nuclear extract the binding of the slowest migrating complex was lost and a supershifted complex was produced (Fig. 7). The intensities of the two other complexes were partly reduced (Fig. 7). As a negative control a polyclonal antiserum against a peptide of human neurofibromin (anti-P111) was used. These experiments confirmed that Sp1 was the nuclear factor bound to the probe in the slowest migrating complex. The two faster migrating complexes may represent other members of the Sp1 multigene family, which share some immunological homology with Sp1 (33). Hence, we conclude that the GC and GT box sequences of the syndecan-1 promoter bind Sp1, and possibly some other Sp1 multigene family members. Interestingly, there were no differences in the binding of nuclear proteins from NMuMG or 3T3 cell extracts to these regions.

Footprinted region C and the 5' region of region G were not studied in gel shift experiments. They both included GT box sequences and we thus assumed that their binding properties were the same as footprinted region F. On the contrary, the 3' part of footprinted region G showed binding both to Sp1-like nuclear proteins and to two unknown cell type-specific nuclear proteins. The double stranded oligonucleotide probe G (5'-GAGGCCGCTAGCTTGTGAAAATGC-3') covering the 3' part of footprinted region G included a possible Sp1 binding sequence (GAGGCCGCTG) mismatching only one nucleotide from the consensus binding sequence for Sp1 (G/GAGGCCT/TA/GAGG/T). In a gel shift assay using probe G and nuclear extract from NMuMG cells, at least five specifically retarded complexes were obtained (Fig. 8, lanes 2 and 3). The proteins in the three slowest migrating complexes (I, II, and III) appeared to be the same Sp1-like proteins as seen with oligonucleotide probes D, E, and F as described above. These complexes co-migrated with the nuclear protein complexes produced with probe D (Fig. 8, lane 10), and their formation was specifically competed with the Sp1 consensus binding site oligonucleotide or oligonucleotide D (Fig. 8, lanes 4 and 6). As with oligonucleotide probes D, E, and F, the slowest migrating complex (I) contained transcription factor Sp1, as in supershift experiments using the polyclonal anti-Sp1 antiserum it was replaced by a supershifted complex (Fig. 8, lane 1). The rabbit IgG used as a negative control in this experiment did not result in a supershift complex (Fig. 8, lane 8). The intensity of complex I was relatively low as compared with those seen with probes of similar specific activity including complete GC or GT box sequences. This might be a result of the mismatch in the Sp1 binding sequence. The protein complexes IV and V were not competed with Sp1 binding oligonucleotides, indicating that they bound to different parts of the probe. Interestingly, nuclear extract from 3T3 cells only produced protein complexes I, II, and III when incubated with probe G (Fig. 8, lane 9). Thus, the nuclear factors producing complexes IV and V were present only in NMuMG cell extract. It was concluded that the 3' region of footprinted region G binds the same set of Sp1-like proteins as footprinted regions D, E, and F, probably through the unusual GC box sequence. In addition, two unknown epithelial cell-specific nuclear protein complexes bound to this footprinted region.

Footprinted region B, which covered one of the transcription initiation sites, also bound a nuclear protein complex. This was demonstrated in gel shift assays using the double stranded oligonucleotide probe B (5'-GAGGGGCTGAGCTTTTCTGGAAGGGGCTG-3') and nuclear extracts from NMuMG and 3T3 cells. A single specifically retarded complex was produced when nuclear extract from NMuMG cells was used (Fig. 9, lanes 2 and 7), whereas, that from 3T3 cells produced only a barely visible band (Fig. 9, lane 3). Formation of this complex was not competed by oligonucleotides D, E, and F (Fig. 9, lanes 4–6), which demonstrated that the binding property of this protein complex was different from the Sp1-like proteins described above.
Deletion of the Transcription Initiation Sites at -249 and -280 — It has previously been assumed that the transcription initiation of the syndecan-1 gene would be directed both by the TATA sequence as in tissue-specific genes, and by GC boxes as is the case in several housekeeping genes (24, 25). Hinkes and co-workers (25) have previously located by primer extension and S1 mapping three transcription initiation sites at positions -223, -249, and -280. Interestingly, the most intense initiation site -249 was located on the putative TATA box sequence. Nuclear protein binding was also demonstrated around the second abundant initiation site (-280) in our footprinting and gel shift assays (Fig. 3). To study the transcription initiation of the gene, the TATA sequence (TTTTATTATAA) and footprinted region B (CTAGTT), both covering a transcription initiation site, were separately deleted from the promoter construct p-437CAT. In transiently transfected NMuMG cells the TATA sequence deletion reduced the promoter activity about 50% (construct p(TATA)CAT in Fig. 10A) and the FP-B deletion about 30% (construct p(AB)CAT in Fig. 10A), as compared to the wild type construct. Primer extension analysis was done to see whether the corresponding transcription initiation sites were lost along the deletions. The RNAs used in primer extension reactions were isolated from NMuMG cells stably transfected by constructs p-437CAT, p(TATA)CAT, and p(AB)CAT.

Fig. 4. Gel mobility shift analysis of footprinted regions D, E, and F including GC or GT box sequences. End-labeled double stranded oligonucleotide probes covering footprinted regions D (lanes 1-3), E (lanes 4-6), and F (lanes 7-9) were incubated without (lanes 1, 4, and 7) or with (lanes 2, 5, and 8) nuclear extract from epithelial NMuMG or from mesenchymal 3T3 cells (lanes 3, 6, and 9). The shifted protein complexes are indicated by horizontal lines to the left of the panel. Fig. 5. Gel mobility shift/competition analysis of the footprinted regions including GC/GT boxes. The end-labeled double stranded oligonucleotide covering footprinted region E was used as a probe in all of the assays. In lane 1 the probe only was loaded. The nuclear extracts from NMuMG (lanes 2–10) or 3T3 cells (lanes 11–15) were first incubated for 10 min with the unlabeled double stranded competitor oligonucleotides F (lanes 3, 4, and 12), E (lanes 5, 6, and 13), D (lanes 7, 8, and 14), and B (lanes 9, 10, and 15) after which the probe was added. 20- (lanes 3, 5, 7, and 9) and 100-fold (lanes 4, 6, 8, 10, and 12-15) molar excess of the competitor was used. In lanes 2 and 11 no competitors were used. The specifically shifted protein complexes are indicated by horizontal lines to the left of the panel.
transcription initiation sites used from the wild type construct p-437CAT co-located with start sites −249 and −280 of the endogenous gene, previously determined by Hinkes and co-workers (25) (lane 1 in Fig. 10B). In addition, some extra bands were located between these two sites. They probably resulted from the unspecific termination of reverse transcriptase, because they were not seen in all of the primer extensions or in any RNase protection assays (data not shown). The deletion of the TATA sequence abolished totally the usage of the start site −249, and shortened the RNA products initiated from start site −280 exactly the length of the deleted fragment (Fig. 10B, lane 2). On the other hand, the deletion of region B eliminated only
The usage of initiation site −280 (Fig. 10B, lane 3). Thus the transcription initiation seemed to be sequence-specific and directed by the site itself.

**DISCUSSION**

In vivo the expression of syndecan-1 gene is constitutive in several epithelial cells. In addition, during wound healing and embryonic development the expression appears to be strongly inducible (for reviews, see Refs. 3 and 34). We have previously reported the genomic organization and nucleotide sequence of the mouse syndecan-1 gene (24). In this work we characterized the upstream gene regions responsible for its transcription regulation in cell culture conditions. We have focused mainly to the identification of the promoter elements involved in the constitutive gene expression. Using CAT assays from transiently and stably transfected cells we mapped a highly active proximal promoter region, which binds Sp1 and probably other members of the Sp family. Moreover, we found that transcription initiation was directed by initiator-like elements as in TATA-less promoters. No enhancer elements were found in the proximal promoter region, which binds Sp1 and probably other proteins in DNase I footprinting assays. GC boxes are typical inhibitory region contained two GT box sequences and the positive region contained three GT box sequences, which were all protected by nuclear proteins in DNase I footprinting assays. GC boxes are typical of promoters and enhancers, including those of the β-globin gene (38), tyrosine aminotransferase gene (39, 40), tryptophan oxygenase gene (41, 42), and interleukin 2 gene (43). Both of the two GC boxes and one of the GT box sequences were studied by gel mobility shift assays. Interestingly, they all bound a similar set of three protein complexes from nuclear extracts of several other promoters and enhancers, including those of the β-globin gene (38), tyrosine aminotransferase gene (39, 40), tryptophan oxygenase gene (41, 42), and interleukin 2 gene (43). Both of the two GC boxes and one of the GT box sequences were studied by gel mobility shift assays.

Previously described features of the promoter included a TATA-like sequence 250 bp upstream from the translation start site (counted as +1) and several Sp1 and AP-2 transcription factor binding sites upstream of it (from −284 to −430) (24). Five E box sequences were found (from −549 to −1612) and a long TAATAA repeat (from −917 to −879), possibly binding sites for Antennapedia homeobox transcription factor (25). Three major transcription initiation sites were located around the putative TATA sequence (25). The genomic organization of chicken syndecan-4 gene resembles that of mouse syndecan-1 (35), but at the moment no information is available on the structure or function of the promoters of the other members of the syndecan gene family.

As described previously by Kim and co-workers (12) syndecan-1 is expressed in most of the cultured epithelial and mesenchymal cells. We used epithelial NMuMG and mesenchymal 3T3 cells, which both expressed high levels of syndecan-1 mRNA in normal culture conditions. However, when these cells are cultured in serum depleted conditions, the syndecan-1 mRNA levels in 3T3 cells, but not in NMuMG cells, decreased dramatically (Fig. 1). By preparing polyclonal cell lines expressing the promoter-CAT constructs we were able to demonstrate that the 2.5-kb long promoter fragment was able to mimic the regulation of the endogenous gene.

To find out the functional elements of the promoter, the detailed characterization was done by transient transfections with NMuMG cells. The deletion of most of the 5′-flanking sequence (from −2.5 kb to −437 bp) had only a minimal effect on promoter activity. At least five E box sequences, which are possible targets for helix-loop-helix transcription factors, such as members of myc and MyoD gene families (36, 37), as well as the long TAATAA repeat were located in the deleted promoter area. They are, thus, most likely not involved in the constitutive epithelial expression of syndecan-1 gene. Further deletions of the promoter sequences first revealed some negative regulatory element(s) (−437 to −365). Downstream from the short region from −365 to −326 was shown to be critical for the function of the promoter, because the deletion of it effectively reduced the reporter gene expression to a minimum. The inhibitory region contained two GT box sequences and the positive region contained three GT box sequences, which were all protected by nuclear proteins in DNase I footprinting assays. GC boxes are typical of promoters and enhancers, including those of the β-globin gene (38), tyrosine aminotransferase gene (39, 40), tryptophan oxygenase gene (41, 42), and interleukin 2 gene (43). Both of the two GC boxes and one of the GT box sequences were studied by gel mobility shift assays. Interestingly, they all bound a similar set of three protein complexes from nuclear extracts of several other promoters and enhancers, including those of the β-globin gene (38), tyrosine aminotransferase gene (39, 40), tryptophan oxygenase gene (41, 42), and interleukin 2 gene (43). Both of the two GC boxes and one of the GT box sequences were studied by gel mobility shift assays.
For example, mutation of the Sp1 site from the keratinocyte-in hormone/growth factor-mediated transcription regulation could also exert a negative effect. Recently Sp1 has been shown into a general initiation complex and it is assumed that they of Sp1 in transcription regulation has been further expanded activity of Sp1 is also regulated by phosphorylation as demonstrated in a classical TATA box should have resulted in a loss of a downstream initiation site. In addition, the mutations decreased the transcription activity of the promoter in accordance with the preference of the abolished start site. This data indicated that the transcription initiation from both of these major start sites is independently regulated by elements directly overlapping the initiation sites. The sequences around the sites exhibited high homology with the consensus sequence of the initiator element (PyPyANPyPyPy), where the A is the start site (46). It has, furthermore, been shown that in TATA-less promoters the direct protein-protein interactions between the Sp1 and TFIIID complex are essential for the assembly of preinitiation complex (47). Consistent with this are the several Sp1 binding sequences upstream the transcription initiation sites in syndecan-1 gene.

The Sp1-like transcription factors seem to have an essential role in the regulation of the transcription activity of the syndecan-1 gene. The significance of Sp1 for the transcriptional activity has been studied in detail, for example, in the SV40 promoter (48-50), in the hamster dihydrofolate reductase promoter (51-53), and in the rat transforming growth factor-α promoter (54). In all of these promoters the Sp1 elements are required for efficient transcription. Although Sp1 was first thought to be a ubiquitous transcription factor chiefly regulating housekeeping genes, recent data indicates that its expression level varies severalfold in different cells and tissues, especially during development (55). Interestingly, the binding activity of Sp1 is also regulated by phosphorylation as demonstrated in terminal differentiation of liver (56). The versatility of Sp1 in transcription regulation has been further expanded by the demonstration that Sp1 also functions through a class of co-activators (57). The co-activators connect trans-activators into a general initiation complex and it is assumed that they could also exert a negative effect. Recently Sp1 has been shown to have an essential role in cell type-specific gene regulation or in hormone/growth factor-mediated transcription regulation. For example, mutation of the Sp1 site from the keratinocyte-specific rabbit K3 keratin promoter resulted in a 50% loss of promoter activity (58), transforming growth factor-β was shown to stimulate α2(I) collagen gene expression by increasing the affinity of an Sp1 containing protein complex for the promoter (59), and the induction of cathepsin D gene expression by estrogen was found to be mediated by an estrogen receptor-Sp1 complex (60). In addition, Li and co-workers (61) have demonstrated that the cellular transcription factor Sp1 and the bovine papillomavirus type 1 enhancer protein E2 synergistically activate transcription from the viral promoter. By electron microscopy, the DNA was shown to make a loop between the enhancer element and the Sp1 complex, which suggests that Sp1 by physically interacting with E2 protein brings the enhancer protein into an appropriate position to influence transcription.

Structural organization and functional analyses of several matrix proteoglycan promoters have been recently reported. The published promoter sequences include those of mouse aggrecan (62), human versican (63), human decorin (64, 65), human biglycan (66), human perlecan (67), and mouse and human seryllys in (68, 69). The promoter structures of the extracellular proteoglycans perlecan and biglycan mostly resemble that of the mouse syndecan-1 gene. The promoter region and 5′-end of the perlecan gene were located in a CpG island. Also, the 5′-flanking region of the biglycan gene was GC-rich. In both genes no canonical TATA or CAAT boxes were found, but several Sp1 transcription factor binding sites were located within the first 200 bp of the promoter. In the perlecan gene five transcription initiation sites were dispersed around the GC boxes, but in the biglycan gene only a single transcription initiation site was found. Also in the mouse aggrecan gene the transcription was initiated from four separate sites and no TATA sequence was found, but otherwise the promoter structure of the extracellular proteoglycan aggrecan, versican, and decorin did not share remarkable homology with mouse syndecan-1 promoter. The functional analyses of these promoters is currently limited, not allowing conclusions of the cell specificity of these promoters. The promoter region of an intracellular proteoglycan seryllys has been functionally analyzed (70). The cell-specific regulatory element was present in the 250-bp long promoter fragment, however, no homology to mouse syndecan-1 promoter was found.

In summary, we have identified functional regions of the mouse syndecan-1 promoter, which are responsible for the constitutive gene expression in epithelial cells. The transcription initiation sites behaved like initiator elements of TATA-less promoters. The upstream region contained several functional GC and GT boxes, which bound members of the Sp1 gene family. This work will provide a basis for further work where we aim at the characterization of syndecan-1 gene suppression during malignant transformation and formation of carcinomas.

Acknowledgments—We thank Drs. Ulrika Pursiainen and J. Jaan Palgi for participating in the sequencing and transfection works. The skilled technical assistance of Taima Laaksonen, Taina Kalevo-Mattila, and Susanna Pykkäri are gratefully acknowledged. We thank Drs. Sue Edwards and Pia Vihinen for critical reading of the manuscript and Dr. Jorma Hermonen for the anti-neurofibromin serum.

REFERENCES
1. Saunders, S., Jalkanen, M., O’Farrel, S., and Bernfield, M. (1989) J. Cell Biol. 108, 1547–1556
2. Rapraeger, A., Jalkanen, M., Endo, E., Kodà, J., and Bernfield, M. (1985) J. Biol. Chem. 260, 11046–11052
3. Bernfield, M., Kokenyesi, R., Kato, M., Hinkes, M. T., Spring, J., Galló, R. L., and Loss, E. J. (1992) Annu. Rev. Cell. Biol. 8, 365–393
4. Jalkanen, M., Eliot, K., and Rapraeger, A. (1993) Trends Glycosci. Glycotex. 5, 107–120
5. Eliot, K., and Jalkanen, M. (1994) J. Cell Biol. 110, 2975–2982
6. Miettinen, H., Edwards, S., and Jalkanen, M. (1994) Mol. Biol. Cell 5, 1325–1339
7. Salmivirta, M., Heino, J., and Jalkanen, M. (1992) J. Biol. Chem. 267, 17606–17610
8. Hayon, A., Klagesmuller, M., Esko, J. D., Leder, P., and Ornitz, D. M. (1991) Cell 64, 841–848
9. Rapraeger, A. C., Krufka, A., and Olwin, B. B. (1991) Science 252, 1705–1708
10. Mali, M., Eliot, K., Miettinen, H. M., and Jalkanen, M. (1993) J. Biol. Chem. 268, 24215–24222
11. Schlesinger, J., L. X., and Lennom, M. (1995) Cell 83, 357–360
12. Kim, C. W., Goldberger, O. A., Galló, R. L., and Bernfield, M. (1994) Mol. Biol. Cell 5, 797–805
13. Vainio, S., Jalkanen, M., Vahtokari, A., Sahblin, C., Mali, M., Bernfield, M., and Theise, N. (1991) Dev. Biol. 147, 322–333
14. Sollors, M., Reiter, R. S., Jensen, K. L., Kato, M., and Bernfield, M. (1990) Dev. Biol. 140, 63–92
15. Vainio, S., Jalkanen, M., Bernfield, M., and Saxén, L. (1992) Dev. Biol. 152, 221–232
