THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 271, No. 21, Issue of May 24, pp. 12532–12541, 1996

Printed in U.S.A.

Functional Characterization of Mouse Syndecan-1 Promoter*

(Received for publication, January 3, 1996, and in revised form, February 13, 1996)

Tapani Vihinen‡§, Arto Määttä§, Panu Jaakkola§, Petri Auvinen§¶, and Markku Jalkanen§‖

From the §Turku Centre for Biotechnology, BioCity, P. O. Box 123, FIN-20521 Turku, Finland and the ‡Department of Medical Biochemistry, University of Turku, FIN-20520 Turku, Finland

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 and 3T3 fibroblastic 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[1] upstream region (24). We have now

* This work was financially supported by the Academy of Finland, the Technology Development Center of Finland (TEKES), the Farmos Science Foundation, the Cancer Union of Finland, and the Juselius Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) Z22532.

¶ Current address: EMBL, European Molecular Biology Laboratory, Mayerhofstrasse 1, Postfach 10.2209, 69012 Heidelberg, Germany.

‖ To whom correspondence should be addressed: Turku Centre for Biotechnology, Tykistökatu 6, BioCity, P. O. Box 123, FIN-20521 Turku, Finland. Tel.: 358-21-3338601; Fax: 358-21-3338000; E-mail: Markku.Jalkanen@btk.utu.fi.

[1] The abbreviations used are: kb, kilobase pair(s); bp, base pair(s);

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 *Xba*I/*Bam*HI (Xb2/5′) fragment and the two following *Xba*I fragments (Xb7 and Xb4) from the cosmid clone cmsyn211 (24) were subcloned into pBluescript KS M13(±) vectors (Stratagene). The subclone Xb7 (−5063 to −4376) was sequenced from double stranded template, but the subclones Xb2/5′ (−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 as described previously (25). DNA sequencing was performed by the dideoxy chain termination method (26) using the Sequenase™ kit (U. S. Biochemical Corp.) and [α-^{35}S]dATP (Amersham). Sequence data base comparisons were made with the Wisconsin package (Genetics Computer Group, Inc.).

Plasmid Constructs—Various promoter fragments ending to the *Xho*I restriction site at position −137 in the 5′-untranslated region of the exon I were cloned into the polylinker of pCAT-basic (Promega) or pCAT-basicZ, upstream from the chloramphenicol acetyltransferase (CAT) gene. The pCAT-basicZ vector was prepared to facilitate subcloning by adding the polylinker region *Sph*I/*Hin*dIII of pGEM7Zf(+) (Promega) into the polylinker region of pCAT-basic. Gene fragments *Hin*dIII/*Xho*I (−2528 to −137), *Stu*I/*Xho*I (−1023 to −137), and *Dra*I/*Xho*I (−830 to −137) were ligated into the polylinker of pCAT-basicZ, 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 *Xba*I/*Xba*I digestion. To generate the p-271CAT construct (−271 to −137), the *Pst*I 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 *Xho*I/*Xba*I 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 fragments share the same downstream primer (5′-dTTGCTCTAGACTTTGCTG-3′) located in the 5′-untranslated region of exon I. A *Xba*I 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 positioning of the oligonucleotide primers in the template, and to check the orientation of the insert. Inserts from pGEM-T vectors were excised by *Sph*I/*Xba*I digestion and recloned into the *Sph*I/*Xba*I 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 protein binding area.

PCR amplification was used to delete the TATA sequence and footprinted region B from the promoter. The TATA sequence TTTATTATAA was removed from the upstream primer (5′-dCTGCAGAGCCTTTGGGGGCGGAGCG-3′) and the downstream primer (5′-dGGCAGGCTGCAGGCGCACGCCAGCG-3′) was selected 332 bp downstream from it. In order to get the region B deletion the sequence AACTAG was left out from the downstream primer (5′-TCTGCAGTTGCAACCCACCCCCAGC-3′) and the upstream primer (5′-AGAGTGGGGTGGGCTTCGA) was selected 137 bp upstream from it. The following reaction conditions were used in PCR amplification. For the first two cycles: denaturation at 94 °C (2 min), annealing at 50 °C (1 min), and extension at 72 °C (3 min). In the following 33 cycles: denaturation at 94 °C (1 min), annealing at 55 °C (1 min), and extension at 72 °C (2 min). The PCR products were first cloned into the pGEM-T vector, after which the p(ΔTATA)CAT and p(ΔB)CAT transfection constructs were generated by replacing the fragments *Pst*I/*Xho*I (−271 to −137) and *Apa*I/*Pst*I (−340 to −271) of construct p-437CAT with the mutated fragment from the pGEM-T clones, respectively. The mutations were verified by sequencing.

To search for putative enhancer elements, the vector pSynPromCAT was first modified from the pCAT-promoter plasmid (Promega) by exchanging the SV40 promoter region upstream of the CAT reporter gene with the mouse syndecan-1 promoter. The SV40 promoter from the pCAT-promoter plasmid was first excised by *Bgl*II/*Stu*I digestion, and then the syndecan-1 promoter fragment *Bgl*II/*Xho*I (−1310 to −137 bp) was ligated in its place after blunt ending the *Xho*I site with Klenow polymerase. Gene fragments were cloned into the polylinker region located far upstream of the reporter gene. Four *Xba*I fragments named Xb4 (−9422 to −5064), Xb7 (−5063 to −4376), Xb5 (+5370 to +9436), and Xb3 (+9752 to +14982) were cloned into the *Xba*I site of the polylinker in a forward orientation, and were designated as pSynProm-Xb4CAT, pSynProm-Xb7CAT, pSynProm-Xb5CAT, and pSynProm-Xb3CAT, respectively. The gene fragments *Xba*I/*Sph*I (−4375 to −2394) and *Bam*HI/*Xba*I (+191 to +5369) were cloned into the polylinker of the vector in the forward orientation, and named as pSynProm-Xb2/5′CAT and pSynProm-Xb2/3′CAT, respectively.

Cell Culture and Transfection Experiments—All plasmids used in transfection experiments were purified by two CsCl gradient centrifugations (27). Normal mouse mammary gland epithelial NMuMG cells and mouse embryo mesenchymal NIH3T3 cells were routinely cultured in bicarbonate-buffered Dulbecco's modified Eagle's medium (Life Technologies, Inc.) supplemented with 10% FCS (Life Technologies, Inc.). Serum depletion was done by culturing cells in 2% carboxymethyl-Sephadex eluted FCS (28). The cells were plated at equal density on six-well plates (Falcon) 1 day before transfection. Plasmid DNA (10 μg) was transfected into cells using the calcium phosphate procedure (29). In transient transfection experiments, a β-galactosidase expressing plasmid (pSV-β-galactosidase, Promega) was co-transfected with the CAT construct to monitor transfection efficiencies. Three or four parallel transfections were used in every assay. Cells were harvested 2 days after transfection and CAT and β-galactosidase activities measured as described previously (24).

To make stably transfected cell clones, CAT constructs were co-transfected with pBGS (a gift from Bruce Granger, University of Montana), a plasmid containing an SV40 promoter, and a neomycin-resistance cassette. Two days after transfection the cells were plated on 10-cm dishes (Falcon) and the selection with Geneticin (G418; 750 μg/ml, 98% pure; Sigma) was started. The G418 concentration was based on titering the selection efficacy with wild type cells. The surviving cell clones were pooled together and the relative copy number of the CAT constructs was determined by Southern hybridization using the 563-bp long *Xba*I/*Nco*I fragment from the vector pCAT-basic (Promega) as a probe.

Northern Blot—RNA was isolated from NMuMG 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 cDNA PM-4 (1) 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 DNase I Footprinting—Two DNA fragments containing the proximal promoter regions of the mouse syndecan-1 gene were used in DNase I footprinting assays. To generate the ^{32}P-labeled 251-bp fragment XB (−388 to −137), the construct p-830CAT was first cleaved with *Xho*I, end-labeled with [α-^{32}P]dCTP and [α-^{32}P]dGTP using Klenow DNA polymerase (Promega), and then digested with *Bam*HI. The same method was used to create the longer labeled 693-bp fragment XD (−830 to −137), which was generated from the construct p-1.0CAT by *Xho*I and *Dra*I digestion. Both of the probes were purified after agarose gel electrophoresis by electroelution with a Gel Eluter (Hoefer Scientific Instruments).

Probes for DNA/Protein Binding Assays—Five oligonucleotide probes were selected on the basis of the DNase I footprinting experiments. Oligonucleotide pair B extended from −289 to −264 bp (5′-dGGGCGGCTAGTTTTGCAACTGCAGAG-3′), D from −351 to −324 bp (5′-dACCCAGGGGGCGGGGCCCGAGGGGTGGA-3′), E from −369 to −343 bp (5′-dGCTCCAAGGGGCGGGGCAACCCAGGGG-3′), F from −389 to −363 bp (5′-dGGATCCCTGGGGGGTGGGGCGCTCCAA-3′),

NMuMG, normal murine mammary gland; CAT, chloramphenicol acetyltransferase; PCR, polymerase chain reaction; FCS, fetal calf serum.

and G from −415 to −392 bp (5′-dCCTAGGAGGCGTGGAAGGGGGTGT-3′). The oligonucleotides were synthesized with an Applied Biosystems automatic DNA synthesizer. The sense strands were end-labeled with [γ-^{32}P]dATP (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 nuclear protein binding sites in the proximal promoter region. 50,000 cpm of the labeled DNA fragments (described above) were first incubated with 5–10 μl of crude nuclear extract and 2 μg of poly-(dI·dC) (Boehringer Mannheim) in a buffer containing 10 mM Tris (pH 8.0), 5 mM $MgCl_2$, 1 mM $CaCl_2$, 2 mM dithiothreitol, 50 μg/ml bovine serum albumin, and 100 mM KCl in a final volume of 50 μl. After 10 min incubation in room temperature, 0.2 or 2 units 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 undialyzed 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 mM HEPES pH 7.9, 100 mM KCl, 0.2 mM EDTA, and 20% glycerol. In all binding assays, 2 μg of poly-(dI·dC) (Boehringer Mannheim) was added as a nonspecific competitor. The complexes were resolved by electrophoresis in 4.5% native polyacrylamide gels using 50 mM Tris, 400 mM glycine buffer (pH 8.5). Indicated amounts of unlabeled specific and nonspecific competitors were used with the oligonucleotide probes to distinguish sequence-specific interactions. Consensus binding site oligonucleotides for Sp1, AP-2, and NFκB were purchased from Promega. Polyclonal rabbit anti-Sp1 peptide antibody (Santa Cruz Biotechnologies) was used to demonstrate the presence of Sp1 protein in the shifted complexes, by formation of high molecular weight supershifted complexes. As control antibodies, rabbit IgG (Sigma) and a polyclonal 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 oligonucleotide (5′-AGCTCCTGAAAATCTCGCCAAGCTCAGATC-3′) hybridizing to the 5′-end of the CAT gene. The primer was end-labeled with [γ-^{32}P]dATP (Amersham) using T4 polynucleotide kinase (Promega), after which it was hybridized to 20 μg of total RNA (30 min at 65 °C in 0.15 M KCl, 10 mM Tris (pH 8.0), 1 mM EDTA). After hybridization the nucleic acids were ethanol precipitated and resuspended in 26 μl of water and 9 μl of 5 × Moloney murine leukemia virus reverse transcriptase buffer (Promega). 1.5 μl of 5 mM dNTP mixture, 6.5 μl of 1 mg/ml actinomycin D (Sigma), and 2 μl (400 units) of Moloney murine leukemia vurus reverse transcriptase (Promega) were added and the reactions incubated for 60 min at 42 °C. After that 105 μl of RNase reaction mixture (100 μg/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 precipitated, 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 syndecan-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 transcription 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 *Xba*I site at −9.4 kb.[2] A comparison of this upstream sequence with all rodent sequences in GenBank revealed the presence of several regions of high homology to repetitive sequences of other genes (between nucleotides −7762 and −2702).

Identification of the Mouse Syndecan-1 Promoter—In order to characterize the regions regulating the transcription activity of the gene, a series of chimaeric constructs containing different 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 syndecan-1 mRNA levels decreased to one-third from the original, after the cells were cultured in serum-deprived conditions (Fig. 1*A*).

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. 1*B*). 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. 1*B*). 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 constructs into NMuMG cells. The deletion of the 5′-flanking region from −2528 to −437 had only a small effect on the promoter activity as seen with constructs p-2.5CAT (−2528 to −137), p-1.0CAT (−1023 to −137), p-830CAT (−830 to −137), p-492CAT (−492 to −95), and p-437CAT (−437 to −95) (Fig. 2). The deletion to −365 (construct p-365CAT) abolished three GT box sequences (GGGGTGGG or GGGGTGTGG) and resulted in a 20–30% increase of promoter activity (Fig. 2). The further deletion to −326 (construct p-326CAT) removed two GC box sequences (GGGGCGGGG) and reduced the promoter activity about 75% (Fig. 2). The short constructs p-326CAT (containing one GT and one GC boxes) and p-289CAT (containing one GC box) showed only low basal transcription activity as did the shortest construct p-271CAT (−271 to −137) including only the putative TATA sequence and the first and the second transcription initiation sites (Fig. 2). Stable expressing NMuMG cell clones were prepared by co-transfecting these chimaeric promoter-CAT constructs with a plasmid pBGS containing a neomycin-resistance gene (6). The CAT activities of the stably transfected clones were in accordance with the data from tran-

[2] The nucleotide sequence of the mouse syndecan-1 gene in EMBL data bank (msyn-1gen; accession number Z22532) has now been supplemented with the upstream sequence to −9.4 kb described in this article.

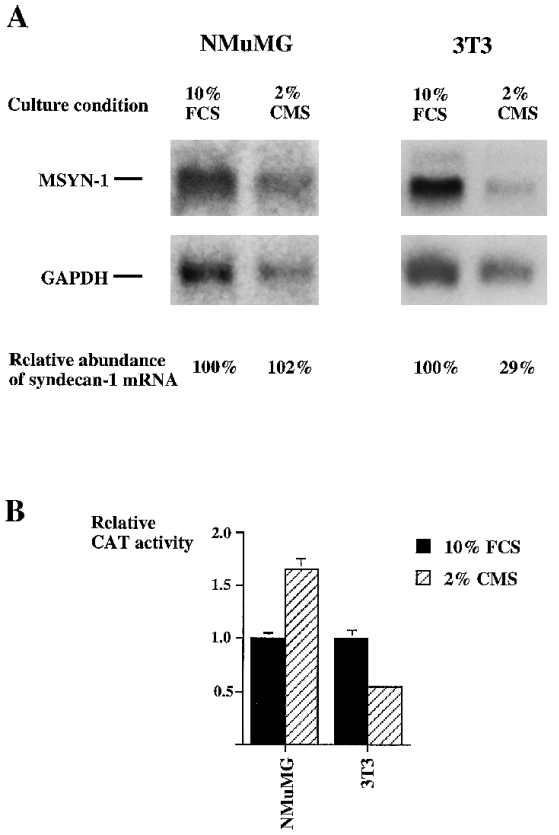

FIG. 1. **Mouse syndecan-1 RNA levels and promoter activity in control and serum-deprived NMuMG and 3T3 cells.** *A*, cells were grown for 48 h in the presence of Dulbecco's modified Eagle's medium supplemented with either 10% FCS or 2% carboxymethyl-Sephadex medium (*CMS*). Syndecan-1 RNA levels were analyzed by Northern blotting using PM-4 probe (1). Rat glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) probe was used to control the RNA loading. *Below*, the relative abundance of syndecan-1 RNA of serum-deprived cells compared to control cells is presented by percentages. *B*, relative activity of the 2.5-kb long syndecan-1 promoter fragment in control and serum-deprived NMuMG and 3T3 cells. Both cell lines were stably co-transfected with a promoter-CAT chimaeric construct and a neomycin resistance gene construct. The effect of serum deprivation for promoter activity is demonstrated as relative CAT activities of cell extracts. The amount of extracts used in CAT reactions were adjusted by their protein amounts.

sient transfections (data not shown). These results demonstrated that the single GT/GC box sequence (between nucleotides −282 and −295) with the downstream promoter region was not able to activate a high level transcription of the reporter gene in NMuMG cells. At least one of the GC boxes located between nucleotides −326 and −365 was also needed for high level transcription activity of syndecan-1 promoter. The more upstream regions (−365 to −437) contained some negatively acting *cis*-elements.

Identification of the Putative Enhancer Areas of the Mouse Syndecan-1 Gene—Transient transfections were done to elucidate possible enhancer regions further modifying the transcription activity of the proximal promoter in constant cell culture conditions. The gene area extending from −9.4 to the +15 kb was cloned as six fragments into a vector containing the mouse syndecan-1 promoter (−1310 to −137) upstream from the CAT reporter gene. Again, epithelial NMuMG and mesenchymal 3T3 cells cultured in 10% FCS were used in the experiments. None of these constructs showed any enhancer activity as compared to the construct pSynPromCAT, which included only the promoter region. On the contrary, the construct pSynPron-Xb2/5′CAT containing the 5′-flanking region from −4375 to −2394 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*I site (−137) to the *Bam*HI site (−388), and the longer probe XD to the *Dra*I 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. 3*B*). 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. 3*B*). Footprints D (−350 to −337) and E (−362 to −354) located at the GC box sequences (GGGGCGGGG), and footprints C (−296 to −284) and F (−383 to −371) covered the GT box sequences (GGGGTGGG) (Fig. 3*C*). 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 GAGGCGTGG mismatching one nucleotide (tymidine) to the consensus DNA binding sequence of Sp1 (Fig. 3*C*). 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. 3*C*), 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′-dACCCA-GGGGCGGGGCCCGAGGGGTGGA), and E (5′ dGCTC-CAAGGGGCGGGGCAACCCAGGGG), and a GT box sequence within probe F (5′-dGGATCCCTGGGGGGTGGGGCGCTC-CAA). 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-

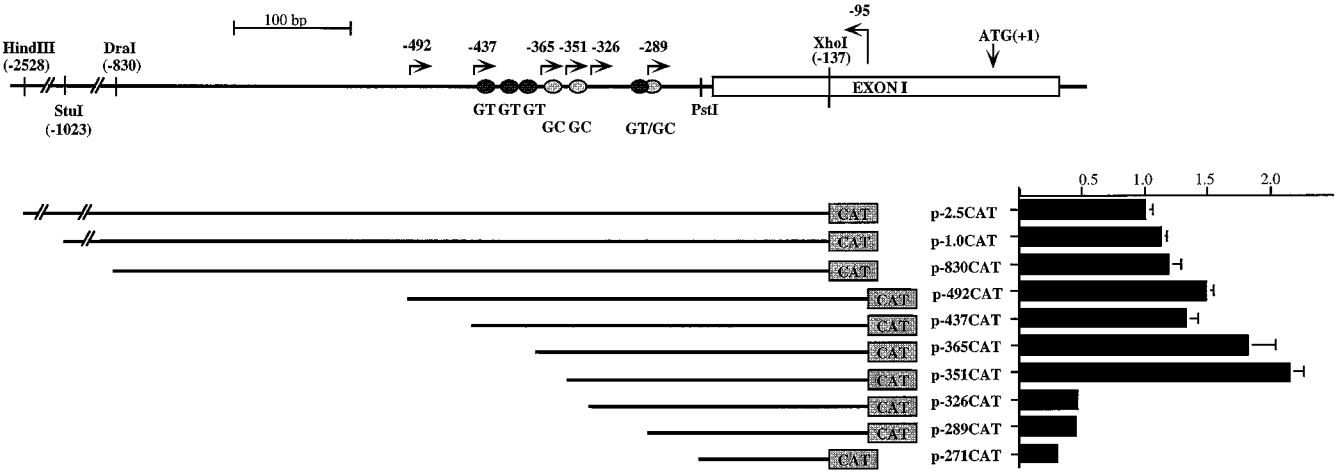

FIG. 2. **Deletion analysis of the mouse syndecan-1 promoter region.** In the *top* of the figure a schematic presentation of the 5′-flanking region of the gene is shown. The *open box* indicates exon I. The *horizontal arrows* show the locations of oligonucleotide primers used in PCR reactions to generate the promoter fragments. Restriction enzymes (*Xho*I, *Pst*I, *Dra*I, *Stu*I, and *Hin*dIII) used in the preparation of the other constructs are also shown. The *vertical arrow* indicates the beginning of the translated sequence, which is counted as +1 in the nucleotide numbering of the sequence. The *closed circles* represent the location of the consensus binding sequences for Sp1 (GC and GT boxes). The *scale bar* represents 100 bp. *Below*, the structures and the names of the chimaeric constructs are shown. The *lines* represent the syndecan-1 promoter fragments and the *hatched boxes* the CAT gene. The 5′- and 3′-ends of the promoter fragments are drawn in scale to the map in the top of the figure. The relative CAT activities of the constructs transfected into NMuMG cells are represented by the *black columns* in the diagram.

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 (ATTCGATCGGGGCGGGGCGAGC), but not by the AP-2 (GATCGAACTGACCGCCCGCGGCCCGT) or NFκB (AGTTGAGGGGACTTTCCCAGGC) binding site oligonucleotides (Fig. 6*A*). 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. 6*B*). Moreover, in all supershift experiments using the polyclonal anti-Sp1 antiserum with probes D, E, 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′ part 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′-dCCTAGGAGGCGTGGAAGGGGGTGT) covering the 3′ part of footprinted region G included a possible Sp1 binding sequence (GAGGCGTGG) mismatching only one nucleotide from the consensus binding sequence for Sp1 (G/TG/AGGCG/TG/AG/AG/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 supershift complex (Fig. 8, *lane 7*). 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 labeling 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′- dGGGCGGCTAGTTTTGCAACTG-CAGAG-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.

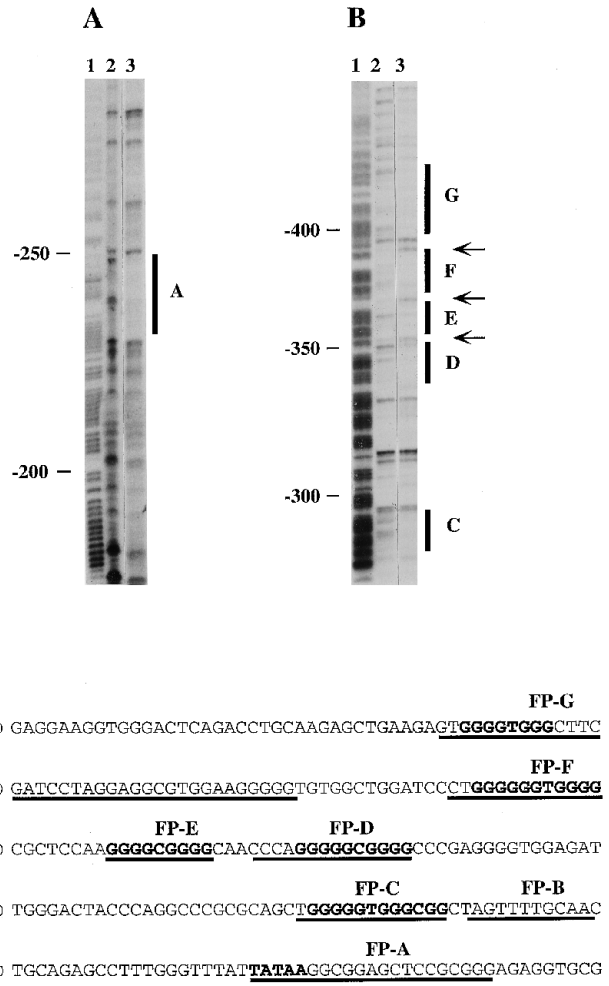

FIG. 3. **DNase I footprinting analysis of the mouse syndecan-1 proximal promoter.** *A*, a *Xho*I-*Bam*HI restriction fragment containing promoter sequences from −137 to −388 was labeled on the sense strand at the *Xho*I end and used as a probe. The probe was incubated either without (*lane 2*) or with (*lane 3*) the nuclear extract prepared from epithelial NMuMG cells. The nucleotide positions (*numbers*) in the *left* of the panel are based on the G + A chemical sequencing reaction ladder shown in *lane 1*. In the *right* of the panel the region protected from DNase I digestion are indicated by the *bar* and *letter*. *B*, a *Xho*I-*Dra*I restriction fragment (−137 to −830) labeled on the sense strand at the *Xho*I site was used in DNase I footprinting analysis as described in *A*. *C*, a nucleotide sequence of the promoter region extending from −470 to −221. The footprinted regions, marked *FP-A*, *-B*, *-C*, *-D*, *-E*, *-F*, and *-G*, are *underlined*. The GC/GT-box sequences and the TATA sequence are indicated by *boldface* letters.

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*C*, *FP-B*). To study the transcription initiation of the gene, the TATA sequence (TTTATTATAA) 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(ΔTATA)CAT in Fig. 10*A*) and the FP-B deletion about 30% (construct p(ΔB)CAT in Fig. 10*A*), 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(ΔTATA)CAT, and p(ΔB)CAT. The sequence for oligonucleotide primer was selected from the pCAT-basic cloning vector, so it hybridized only to the RNA transcribed from the CAT constructs. The two most abundant

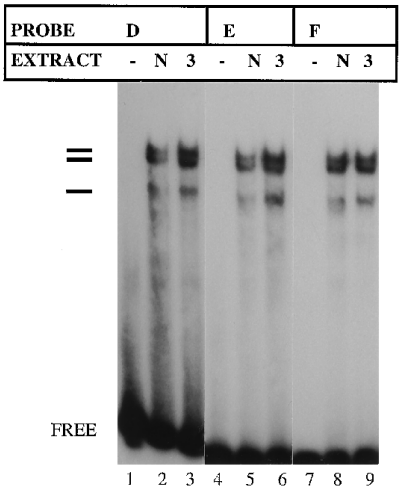

FIG. 4. **Gel mobility shift analysis of footprinted regions D, E, and F including GC or GT box sequences.** End-labeled double stranded oligonucleotide probes covering footprinted regions D (*lanes 1–3*), E (*lanes 4–6*), and F (*lanes 7–9*) were incubated without (*lanes 1, 4,* and *7*) or with (*lanes 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.

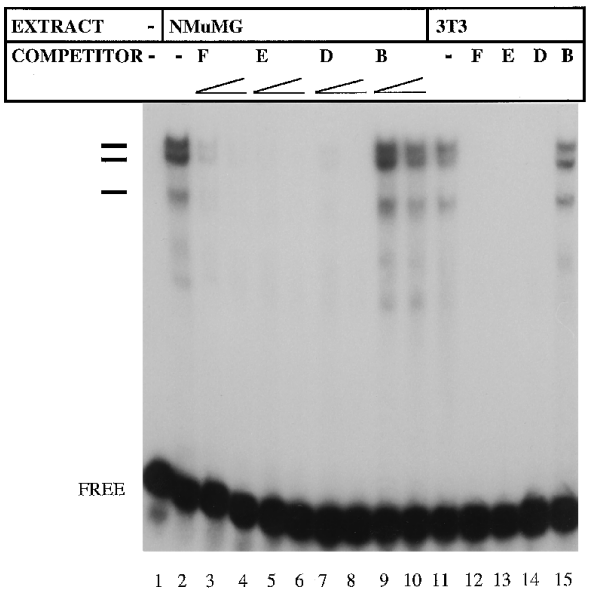

FIG. 5. **Gel mobility shift/competition analysis of the footprinted regions including GC/GT boxes.** The end-labeled double stranded oligonucleotide covering footprinted region E was used as a probe in all of the assays. In *lane 1* the probe only was loaded. The nuclear extracts from NMuMG (*lanes 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- (*lines 3, 5, 7,* and *9*) and 100-fold (*lines 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.

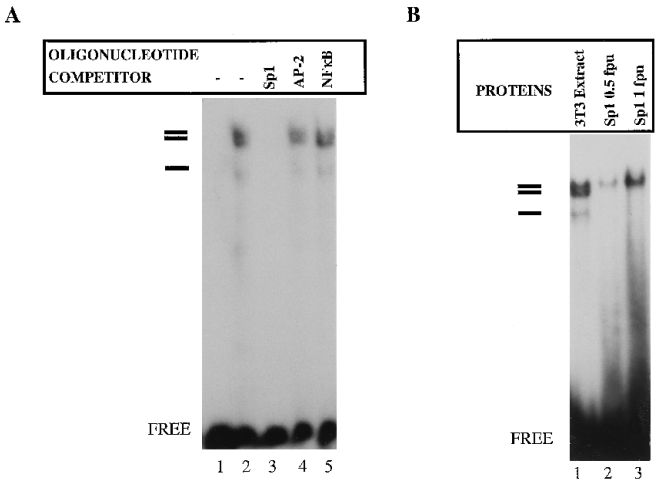

FIG. 6. **Gel mobility shift analysis demonstrating that Sp1 binds to the GC/GT boxes.** *A*, nuclear extract from NMuMG cells was incubated with the unlabeled double stranded competitor oligonucleotide before the probe was added. The double stranded oligonucleotide covering footprinted region E was used as a probe. In *lane 1* the probe only and in *lane 2* the probe only with nuclear extract were loaded. The competitor oligonucleotides including consensus binding sites for Sp1 (*lane 3*), AP-2 (*lane 4*), and NFκB (*lane 5*) were used in 100-fold molar excess. The specifically shifted protein complexes are indicated by *horizontal lines* in the *left* of the panel. *B*, the slowest migrating complex generated from the 3T3 cell nuclear extract with oligonucleotide probe E (*lane 1*) co-migrates with the retarded complex produced with pure Sp1 protein (*lane 2, and 3*). 0.5 (*lane 2*) and 1.0 (*lane 3*) footprinting units of the Sp1 protein were used.

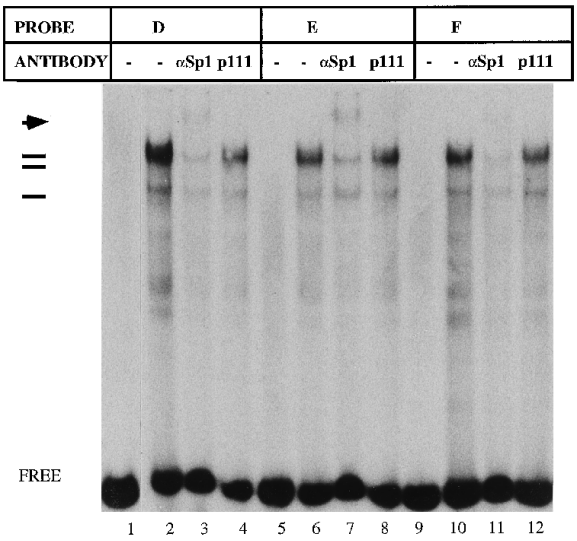

FIG. 7. **Characterization of GC/GT box binding proteins by immunosupershift.** A polyclonal rabbit anti-Sp1 peptide antibody (*lanes 3, 7,* and *11*) or as a control a polyclonal antibody against human neurofibromin (anti-P111; *lanes 4, 8,* and *12*) were incubated for 10 min with the nuclear extract from NMuMG cells before the oligonucleotide probe was added. End-labeled oligonucleotide probes covering footprinted regions *D* (*lanes 1–4*), E (*lanes 5–8*), and E (*lanes 9–12*) were used. In *lanes 1, 5,* and *9*, pure probes, and in *lanes 2, 6,* and *10*, the probes incubated with nuclear extract without antibody were loaded. The specifically shifted protein complexes are indicated by *horizontal lines* in the *left* of the panel and the supershifted complex by *horizontal arrows*.

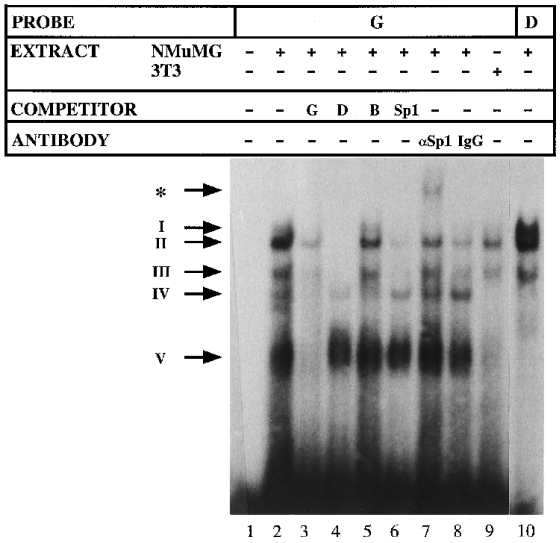

FIG. 8. **Gel mobility shift analyses of footprinted region G.** The end-labeled double stranded oligonucleotide probe G covering the 3′ part of footprinted region G (*lanes 1–9*) and the oligonucleotide probe D (*lane 10*) were used. In *lane 1*, only probe G was loaded. Nuclear extracts from NMuMG (*lanes 2–8* and *10*) and 3T3 cells (*lane 9*) were used. The five complexes generated from NMuMG cell extract (*lane 2*) are indicated by *roman numbers* and *arrows* in the *left* of the panel. The competition assays were carried out with probes D, E, and F as described in the legend to Fig. 5. The unlabeled oligonucleotide G (*lane 3*), oligonucleotides covering footprinted regions D (*lane 4*) and B (*lane 5*), and Sp1 consensus binding site oligonucleotide (*lane 6*) were used as competitors in 100-fold molar excess to probe. A polyclonal rabbit anti-Sp1 peptide antibody (*lane 7*) and as a control rabbit IgG (*lane 8*) were used in supershift experiments. The supershifted complex is indicated by an *asterisk* and *arrow* at the *left* of the panel. From nuclear extract of 3T3 cells only the three slowest migrating complexes were produced (*lane 9*). Oligonucleotide probe D was used to demonstrate that the three largest complexes produced with probe G are co-migrating with complexes binding to GC/GT boxes (*lane 10*).

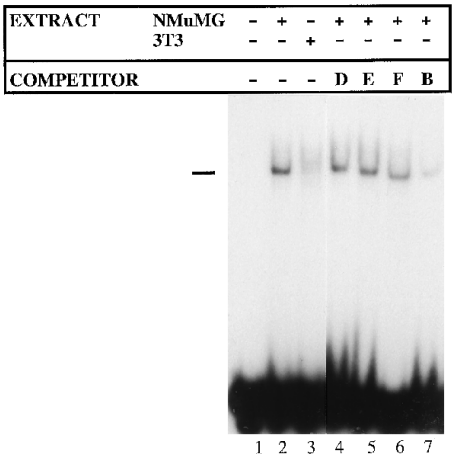

FIG. 9. **Gel mobility shift analyses of footprinted region B.** The end-labeled double stranded oligonucleotide covering footprinted region B was used as a probe. Probe only was loaded in *lane 1*. With nuclear extract from NMuMG cells a single protein complex shift was formed (*lane 2*), whereas from 3T3 cell extract only a faintly visible band was produced (*lane 3*). The unlabeled competitor oligonucleotides D (*lane 4*), E (*lane 5*), F (*lane 6*), and B itself (*lane 7*) were used in 100-fold molar excess as in previous shift assays. The specifically shifted protein complex is indicated by a *horizontal line*.

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 coworkers (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. 10*B*, *lane 2*). On the other hand, the deletion of region B eliminated only

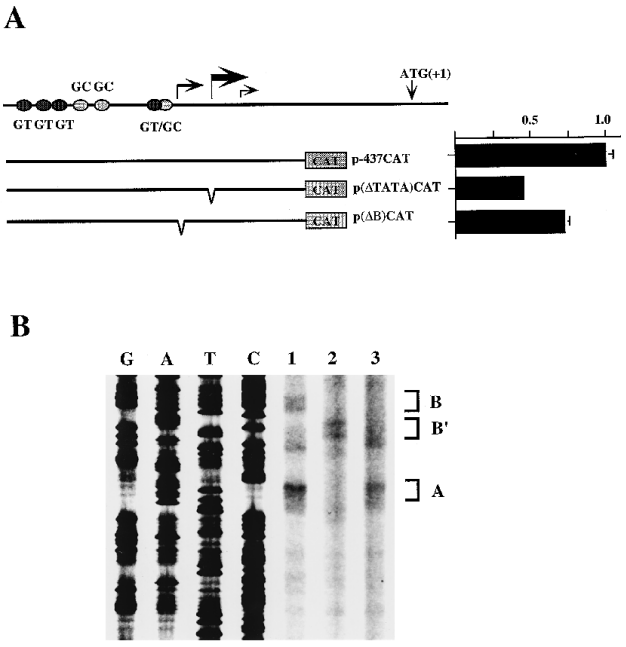

FIG. 10. **Deletion of the promoter elements effect on the transcription activity and start site usage.** *A*, in the top of the figure a schematic presentation of the promoter region is shown. The *filled circles* represent the location of the footprinted sequences for Sp1 (GC and GT boxes). The *horizontal arrows* show the location of transcription initiation sites. The size of the *arrow* indicates the relative abundance of each initiation site (25). The *vertical arrow* indicates the beginning of the translated sequence. *Below*, the structures and the names of the chimaeric constructs are shown. The *lines* represent the syndecan-1 promoter fragments and the *hatched boxes* the CAT genes. The relative CAT activities of the constructs transiently transfected into NMuMG cells are represented by the *black columns* in the diagram. *B*, primer extension analyses of the RNAs transcribed from the promoter-CAT construct. A primer complementary to the 5′-end of the CAT gene was hybridized to total RNA from NMuMG cell clones, which were stably transfected with constructs p-437CAT (*lane 1*), p(ΔTATA)CAT (*lane 2*), and p(ΔB)CAT (*lane 3*). The primer extension products are indicated by *brackets*. The product *A* located at position −252 bp and the lower band of the product *B* at −281 bp. *B′* indicates the shortened extension product B as a result of the TATA sequence deletion. Sequencing reaction (*lanes G, A, T,* and *C*) was done with the same primer as the primer extension reactions, thus the complementary nucleotides of the transcription initiation sites can be read directly from the sequencing ladder.

the usage of initiation site −280 (Fig. 10*B*, *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 CAT assays when the upstream region (up to −9.4 kb) and the first intron (to +15 kb) were studied, whereas some suppressor elements were located upstream to the promoter (−2.4 to −4.4 kb).

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 a binding site 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 this 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 that 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 two GC box sequences, which were all protected by nuclear proteins in DNase I footprinting assays. GC boxes are typical Sp1 transcription factor binding sites found in many genes. Also, functional GT box elements have been found in 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. Interestingly, they all bound a similar set of three protein complexes from nuclear extracts of both epithelial NMuMG and mesenchymal 3T3 cells. The transcription factor Sp1 proved to be one of the binding proteins, as demonstrated by competition assays, by co-migration of pure Sp1 protein, and by supershifts with specific antibodies. We propose that the two uncharacterized complexes represent other members of the Sp1 multigene family, since their binding was easily competed by Sp1 consensus binding site oligonucleotide. Today at least four separate Sp1-like nuclear factors (Sp1, Sp2, Sp3, and Sp4) have been characterized (33, 44). They all bind to both GC and GT box sequences, even though Sp1 and Sp3 bind GT boxes with higher affinity than GC boxes. A highly conserved Sp1-like zinc finger region for DNA binding and a glutamine-rich region for transactivation are typical for all these proteins. Interestingly, it has been shown that Sp3 represses Sp1-mediated transcription activation by competition

(45). This might explain the slight transcription inhibition seen with the longer syndecan-1 promoter constructs, which contained footprinted region G. This region bound the two other proteins more strongly than Sp1. Some other zinc finger DNA-binding proteins have also been shown to bind to GC and GT boxes. For example, EGR-1, a member of an early growth response gene family binds to a GT box sequence of the interleukin 2 gene promoter and activates the gene expression (43).

The assembly of the transcription initiation machinery in syndecan-1 promoter was directed by initiator elements typically found in TATA-less promoters. Syndecan-1 promoter contains a putative TATA box sequence 23 bp upstream from one of the previously described transcription initiation sites (25). However, the most frequently used transcription initiation site located only 2 bp and the second intense site 31 bp upstream from this TATA box. Deletion of the sequences spanning these two major transcription start sites resulted in a loss of transcription initiation from the site being deleted, while a deletion of 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 to 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 TFIID 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 serglysin genes (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 serglysin has been functionally analyzed (70). The cell-specific regulatory element were 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 Jaan Palgi for participating in the sequencing and transfection works. The skilled technical assistance of Taija Laaksonen, Taina Kalevo-Mattila, and Susanna Pyökäri 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., Koda, J., and Bernfield, M. (1985) *J. Biol. Chem.* **260,** 11046–11052
3. Bernfield, M., Kokenyesi, R., Kato, M., Hinkes, M. T., Spring, J., Gallo, R. L., and Lose, E. J. (1992) *Annu. Rev. Cell Biol.* **8,** 365–393
4. Jalkanen, M., Elenius, K., and Rapraeger, A. (1993) *Trends Glycosci. Glycotech.* **5,** 107–120
5. Elenius, K., and Jalkanen, M. (1994) *J. Cell Sci.* **107,** 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. Yayon, A., Klagsbrun, 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., Elenius, K., Miettinen, H. M., and Jalkanen, M. (1993) *J. Biol. Chem.* **268,** 24215–24222
11. Schlessinger, J., Lax, I., and Lemmon, M. (1995) *Cell* **83,** 357–360
12. Kim, C. W., Goldberger, O. A., Gallo, R. L., and Bernfield, M. (1994) *Mol. Biol. Cell* **5,** 797–805
13. Vainio, S., Jalkanen, M., Vaahtokari, A., Sahlberg, C., Mali, M., Bernfield, M., and Thesleff, I. (1991) *Dev. Biol.* **147,** 322–333
14. Solursh, M., Reiter, R. S., Jensen, K. L., Kato, M., and Bernfield, M., (1990) *Dev. Biol.* **140,** 83–92
15. Vainio, S., Jalkanen, M., Bernfield, M., and Saxén, L. (1992) *Dev. Biol.* **152,** 221–232

16. Brauker, J. H., Trautman, M. S., and Bernfield, M. (1991) *Dev. Biol.* **147,** 285–292
17. Elenius, K., Vainio, S., Laato, M., Salmivirta, M., Thesleff, I., and Jalkanen, M. (1991) *J. Cell Biol.* **114,** 585–595
18. Leppä, S., Härkönen, P., and Jalkanen, M. (1991) *Cell Regul.* **2,** 1–11
19. Leppä, S., Mali, M., Miettinen, H. M., and Jalkanen, M. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 932–936
20. Inki, P., Stenbäck, F., Talve, L., and Jalkanen, M. (1991) *Am. J. Pathol.* **139,** 1333–1340
21. Inki, P., Kujari, H., and Jalkanen, M. (1992) *Lab. Invest.* **66,** 314–323
22. Ridley, R. C., Xiao, H., Hata, H., Woodliff, J., Epstein, J., and Sanderson, R. D. (1993) *Blood* **81,** 767–774
23. Inki, P., Joensuu, H., Grenman, R., Klemi, P., and Jalkanen, M. (1994) *Br. J. Cancer* **70,** 319–323
24. Vihinen, T., Auvinen, P., Alanen-Kurki, L., and Jalkanen, M. (1993) *J. Biol. Chem.* **268,** 17261–17269
25. Hinkes, M. T., Goldberger, O. A., Neumann, P. E., Kokenyesi, R., and Bernfield, M. (1993) *J. Biol. Chem.* **268,** 11440–11448
26. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
27. Ausubel, F. M., Brent, R., Kingston, E. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1993) *Current Protocols in Molecular Biology*, Wiley Interscience, New York
28. Vogel, A., Raines, E., Kariya, B., Rivest, M.-J., and Ross, R. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 2810–2814
29. Chen, C., and Okayama, H. (1987) *Mol. Cell. Biol.* **7,** 2745–2752
30. Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159
31. Fort, P., Marty, L., Piechazyk, M., El Sabrouty, S., Dani, C., Jeanteur, P., and Planchard, J. M. (1985) *Nucleic Acids Res.* **13,** 1431–1442
32. Maxam, A., and Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560
33. Kingsley, C., and Winoto, A. (1992) *Mol. Cell. Biol.* **12,** 4251–4261
34. Salmivirta, M., and Jalkanen, M. (1995) *Experientia* (*Basel*) **51,** 863–872
35. Baciu, P. C., Acaster, C., and Goetinck, P. F. (1994) *J. Biol. Chem.* **269,** 696–703
36. Murre, C., McCaw, P. S., and Baltimore, D. (1989) *Cell* **56,** 777–783
37. Blackwell, T. K., Kretzner, L., Blackwood, E. M., Eisenman, R. N., and Weintraub, H. (1990) *Science* **250,** 1149–1151
38. Myers, R. M., Tilly, K., and Maniatis, T. (1986) *Science* **232,** 613–618
39. Becker, P. B., Gloss, B., Schmid, W., Strähle, U., and Schütz, G. (1986) *Nature* **324,** 686–688
40. Strähle, U., Schmid, W., and Schütz, G. (1988) *EMBO J.* **6,** 3389–3395
41. Danesch, U., Gloss, B., Schmid, W., Schütz, G., and Renkawitz, R. (1987) *EMBO J.* **6,** 625–630
42. Schüle, R., Muller, M., Otsuka-Murakami, H., and Renkawitz, R. (1988) *Nature* **332,** 87–90
43. Skerka, C., Decker, E. L., and Zipfel, P. F. (1995) *J. Biol. Chem.* **270,** 22500–22506
44. Hagen, G., Müller, S., Beato, M., and Suske, G. (1992) *Nucleic Acids Res.* **20,** 5519–5525
45. Hagen, G., Müller, S., Beato, M., and Suske, G. (1994) *EMBO J.* **13,** 3843–3851
46. Javaheri, R., Khachi, A., Lo, K., Zenzie-Gregory, B., and Smale, S. T. (1994) *Mol. Cell. Biol.* **14,** 116–127
47. Pugh, B. F., and Tjian, R. (1991) *Genes & Dev.* **5,** 1935–1945
48. Dynan, W. S., and Tjian, R. (1983) *Cell* **35,** 79–87
49. Gidoni, D., Kadonaga, J. T., Barrera-Saldana, H., Takahashi, K., Chamber, P., and Tjian, R. (1985) *Science* **230,** 511–517
50. Courey, A., and Tjian, R. (1988) *Cell* **55,** 887–898
51. Azizkhan, J. C., Vaughn, J., Christy, R. J., and Hamlin, J. H. (1986) *Biochemistry* **25,** 6228–6236
52. Swick, A. G., Blake, M. C., Kahn, J. W., and Azizkhan, J. C. (1989) *Nucleic Acids Res.* **17,** 9291–9304
53. Blake, M. C., Jambou, R. C., Swick, A. G., Khan, J. W., and Azizkhan, J. C. (1990) *Mol. Cell. Biol.* **10,** 6632–6641
54. Chen, X., Azizkhan, J. C., and Lee, D. C. (1992) *Oncogene* **7,** 1805–1815
55. Saffer, J. D., Jackson, S. P., and Annarella, M. B. (1991) *Mol. Cell. Biol.* **11,** 2189–2199
56. Leggett, R. W., Armstrong, S. A., Barry, D., and Mueller, C. R. (1995) *J. Biol. Chem.* **270,** 25879–25884
57. Pugh, B. F., and Tjian, R. (1990) *Cell* **61,** 1187–1197
58. Wu, R-L., Chen, T-T., and Sun, T-T. (1994) *J. Biol. Chem.* **269,** 28450–28459
59. Inagaki, Y., Truter, S., and Ramirez, F. (1994) *J. Biol. Chem.* **269,** 14828–14834
60. Krishnan, V., Wang, X., and Safe, S. (1994) *J. Biol. Chem.* **269,** 15912–15917
61. Li, R., Knight, J. D., Jackson, S. P., Tjian, R., and Botchan, M. R. (1991) *Cell* **65,** 493–505
62. Watanabe, H., Gao, L., Sugiyama, S., Doege, K., Kimata, K., and Yamada, Y. (1995) *Biochem. J.* **308,** 433–440
63. Naso, M. F., Zimmermann, D. R., and Iozzo, R. V. (1994) *J. Biol. Chem.* **269,** 32999–33008
64. Santra, M., Danielson, K. G., and Iozzo, R. V. (1994) *J. Biol. Chem.* **269,** 579–587
65. Mauviel, A., Santra, M., Chen, Y. Q., Uitto, J., and Iozzo, R. V. (1995) *J. Biol. Chem.* **270,** 11692–11700
66. Fisher, L. W., Heegaard, A.-M., Vetter, U., Vogel, W., Just, W., Termine, J. D., and Young, M. F. (1991) *J. Biol. Chem.* **266,** 14371–14377
67. Cohen, I. R., Grässel, S., Murdoch, A. D., and Iozzo, R. V. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 10404–10408
68. Avraham, S., Austen, K. F., Nicodemus, C. F., Gartner, M. C., and Stevens, R. L. (1989) *J. Biol. Chem.* **264,** 16719–16726
69. Nicodemus, C. F., Avraham, S., Austen, K. F., Purdy, S., Jablonski, J., and Stevens, R. L. (1990) *J. Biol. Chem.* **265,** 5889–5896
70. Avraham, S., Avraham, H., Austen, K. F., and Stevens, R. L. (1992) *J. Biol. Chem.* **267,** 610–617