Molecular Cloning of Syndecan, an Integral Membrane Proteoglycan

Scott Saunders, Markku Jalkanen, Susan O'Farrell, and Merton Bernfield

Department of Pediatrics, Stanford University School of Medicine, Stanford, California 94305

Abstract. We describe cDNA clones for a cell surface proteoglycan that bears both heparan sulfate and chondroitin sulfate and that links the cytoskeleton to the interstitial matrix. The cDNA encodes a unique core protein of 32,868 D that contains several structural features consistent with its role as a glycosaminoglycan-containing matrix anchor. The sequence shows discrete cytoplasmic, transmembrane, and NH_2-terminal extracellular domains, indicating that the molecule is a type I integral membrane protein. The cytoplasmic domain is small and similar in size but not in sequence to that of the β-chain of various integrins. The extracellular domain contains a single dibasic sequence adjacent to the extracellular face of the transmembrane domain, potentially serving as the protease-susceptible site involved in release of this domain from the cell surface. The extracellular domain contains two distinct types of putative glycosaminoglycan attachment sites; one type shows sequence characteristics of the sites previously described for chondroitin sulfate attachment (Bourdon, M. A., T. Krusius, S. Campbell, N. B. Schwartz, and E. Ruoslahti. 1987. *Proc. Natl. Acad. Sci. USA.* 84:3194–3198), but the other type has newly identified sequence characteristics that potentially correspond to heparan sulfate attachment sites. The single N-linked sugar recognition sequence is within the putative chondroitin sulfate attachment sequence, suggesting asparagine glycosylation as a mechanism for regulating chondroitin sulfate chain addition. Both 5′ and 3′ regions of this cDNA have sequences substantially identical to analogous regions of the human insulin receptor cDNA: a 99-bp region spanning the 5′ untranslated and initial coding sequences is 67% identical and a 35-bp region in the 3′ untranslated region is 81% identical in sequence. mRNA expression is tissue specific; various epithelial tissues show the same two sizes of mRNA (2.6 and 3.4 kb); in the same relative abundance (3:1), the cerebrum shows a single 4.5-kb mRNA. This core protein cDNA describes a new class of molecule, an integral membrane proteoglycan, that we propose to name syndecan (from the Greek *syndein*, to bind together).

THE cellular behavior responsible for the development, repair, and maintenance of tissues is regulated, in large part, by interactions between the cells and their extracellular matrix. These interactions are mediated by cell surface molecules acting as receptors that bind the large insoluble matrix molecules and induce responses that result in changes of cellular phenotype. Several proteins associated with the cell surface can bind matrix components (Buck and Horwitz, 1987). These proteins differ in their specificity and affinity and in their mode of association with the cell surface. Some bind cells to single matrix ligands while others, such as some members of the integrin super family (Hynes, 1987; Ruoslahti and Pierschbacher, 1987), appear to have multiple matrix ligands. Of the various matrix-binding proteins at the cell surface, only the integrins are known to be integral membrane proteins. Indeed, the integrin fibronectin receptor codistributes both with extracellular fibronectin and with intracellular cytoskeletal components (Chen et al., 1985; Damsky et al., 1985), apparently via an association of the receptor's cytoplasmic domain with the cytoskeletal protein talin (Horwitz et al., 1986).

We have been studying a lipophilic proteoglycan containing both heparan sulfate and chondroitin sulfate (Rapraeger et al., 1985) that is found at the surface of mouse mammary epithelial cells and that behaves as a high affinity receptor specific for multiple components of the interstitial matrix. The proteoglycan binds the epithelial cells via its heparan sulfate chains to collagen types I, III, and V (Koda et al., 1985), fibronectin (Saunders and Bernfield, 1988), and thrombospondin (Sun, X. et al., 1989). When its extracellular domain (ectodomain) is cross-linked at the cell surface, it associates intracellularly with the actin cytoskeleton (Rapraeger et al., 1986), and the isolated proteoglycan binds directly or indirectly to F-actin (Rapraeger and Bernfield, 1982). Cultured epithelial cells shed the ectodomain from their apical surfaces as a nonlipophilic proteoglycan that contains all of the glycosaminoglycan of the intact molecule and polarize the proteoglycan exclusively to their basolateral surfaces (Rapraeger et al., 1986), a location consistent with its matrix receptor function. Upon suspension of these cells, the ectodomain is cleaved from the cell surface; the proteoglycan is not replaced while the cells are suspended (Jalkanen et al., 1987). The proteoglycan is mainly on epithelia in mature tissues (Hayashi et al., 1987), and we have proposed

that it is a matrix anchor that stabilizes the morphology of epithelial sheets by linking the cytoskeleton to the extracellular matrix (Bernfield et al., 1985).

The cell surface proteoglycan undergoes substantial regulation; its size, glycosaminoglycan composition, and location at the cell surface vary between epithelial types, and its expression changes during development. The proteoglycan is located exclusively at the basolateral cell surface of simple epithelia but surrounds stratified epithelial cells (Hayashi et al., 1987). At basolateral cell surfaces, it appears to contain two heparan sulfate and two chondroitin sulfate chains, but where it surrounds cells, it contains only a single heparan sulfate chain and a single small chondroitin sulfate chain (Sanderson and Bernfield, 1988). In self-renewing epithelial cell populations, such as the epidermis or vagina, the proteoglycan is lost when the cells terminally differentiate (Hayashi et al., 1988). In embryos, the proteoglycan is transiently lost when epithelia change their shape and is transiently expressed by mesenchymal cells undergoing morphogenetic tissue interactions (Thesleff et al., 1988; Trautman, M. S., J. Kimelman, and M. Bernfield, manuscript submitted for publication).

Heparan sulfate proteoglycans are ubiquitous on the surfaces of adherent cells and bind various ligands including extracellular matrix, growth factors, proteinase inhibitors, and lipoprotein lipase (Fransson, 1987). However, no structure is known for the core protein of any such cell surface proteoglycan. Therefore, using a library from mouse mammary epithelial cells, we have molecularly cloned and sequenced full-length cDNAs for the cell surface proteoglycan matrix receptor and have assessed the expression of its mRNA in various tissues. The 311-amino acid core protein has a unique sequence that contains several structural features consistent with its role as a matrix anchor and as an acceptor of two distinct types of glycosaminoglycan chains. The expression of its mRNA is tissue type specific and both the 5′ and 3′ untranslated regions of its cDNA show substantial sequence homology to those of the human insulin receptor cDNA. This core protein cDNA defines a new class of matrix receptor, an integral membrane proteoglycan, for which we propose the name syndecan (from the Greek *syndein*, to bind together).

Materials and Methods

cDNA Libraries

Normal murine mammary gland (NMuMG)[1] cells (passages 13–22) were maintained in bicarbonate-buffered DME (Gibco Laboratories, Grand Island, NY) as previously described (David and Bernfield, 1979). For preparation of poly(A) RNA, cells were plated on 245 × 245-mm tissue culture plates (Nunc, Roskilde, Denmark) at ~20% confluent density and grown to 80–90% confluency (3–4 d). After brief washing with ice-cold PBS, the cells were solubilized in RNA extraction buffer (4 M guanidine isothiocyanate in 5 mM sodium citrate, pH 7.0, 0.1 M β-mercaptoethanol, and 0.5% *N*-lauryl sarcosine), and total RNA was prepared by CsCl density centrifugation (Chirgwin et al., 1979). Poly(A) RNA was purified by chromatography on oligo(dT)-cellulose (type 3; Collaborative Research Inc., Waltham, MA) and used in the commercial synthesis (Stratagene, La Jolla, CA) of cDNA by the S1 method (Huynh et al., 1985). After the addition of Eco RI linkers, those cDNA >1 kb long were isolated by gel filtration chromatography, inserted into the Eco RI sites of λ gt-10 and the expression vector λ gt-11, and packaged. A portion of the gt-11 library was amplified for later study, while the remainder was screened immediately without expansion.

A primer extension cDNA library was prepared using the RNase H method (Gubler and Hoffman, 1983). First-strand cDNA was synthesized from 10 μg liver poly(A) RNA (prepared as described below) using reverse transcriptase (Life Sciences Inc., St. Petersburg, FL) specifically primed with 1 μg of an 18-bp oligonucleotide containing sequence derived from near the 5′ end of PM-4. The second strand was synthesized using RNase H (Bethesda Research Laboratories, Gaithersburg, MD) and DNA polymerase Klenow fragment (Boehringer Mannheim Biochemicals, Indianapolis, IN). The cDNA was methylated with EcoRI methylase and then ligated with synthetic Eco RI linkers (New England Biolabs, Beverly, MA). Excess linkers were removed by Eco RI digestion and the cDNA was purified on agarose gel electrophoresis and recovered by electroelution. The resulting cDNA was inserted into λ gt-10 (Promega Biotec, Madison, WI) and packaged using Giga pack Gold (Stratagene).

Isolation of cDNA Clones

The preparation of a rabbit serum antibody to the ectodomain of NMuMG syndecan has been described elsewhere (Jalkanen et al., 1988). For screening clones in λ gt-11, the immunoserum was first absorbed against *Escherichia coli* proteins to reduce background. Briefly, a 500-ml culture of *Escherichia coli* strain Y1090 was grown to saturation in the presence of 50 μg/ml ampicillin. After centrifugation, the cells were resuspended in 50 ml Tris-buffered saline/Triton (TBST, 10 mM Tris, pH 7, 150 mM NaCl, 0.3% Triton X-100), sonicated, and after addition of 100 μl immunoserum (1:500 dilution), incubated overnight at 4°C. This mixture was centrifuged for 10 min at 4,000 rpm and used to screen expressed λ gt-11 cDNA clones (Young and Davis, 1983) by detection with alkaline phosphatase–conjugated goat anti–rabbit IgG (Promega Biotec). Four antibody reactive clones were identified and plaque purified. Northern and clonal Southern hybridization experiments allowed grouping of these clones into three distinct sets of related clones. Two of these sets produced fusion proteins that reacted with immunoserum affinity purified against the ectodomain of syndecan. A clone from one of these sets, PM-4, was found to contain a sequence that exactly matched the partial amino acid sequence of a cyanogen bromide–cleaved fragment of the ectodomain of syndecan.

Additional screening of both the primer extended liver and NMuMG λ gt-10 libraries was performed using radiolabeled fragments from the 5′ end of PM-4. cDNA fragments isolated from SeaPlaque agarose (FMC BioProducts, Rockland, ME) were labeled with ^{32}P by random oligonucleotide priming (Feinberg and Vogelstein, 1984) and used as described by Maniatis et al. (1982).

Subcloning and DNA Sequencing

Purified lambda DNA was prepared from positively selected clones by Lambdasorb immunoprecipitation (Promega Biotec). Fragments released by restriction endonuclease digestions were isolated by electrophoresis after excision from SeaPlaque agarose (FMC BioProducts). These isolated fragments were subcloned directly, in the presence of agarose (Struhl, 1985), to either pGEM 3 and 4 for in vitro transcription, or M13 mp18 and mp19 (Messing, 1983) for sequence analysis.

DNA sequencing was performed by the dideoxy chain termination method (Sanger et al., 1977), using a modified T7 DNA polymerase (Sequenase; United States Biochemical Corp., Cleveland, OH). Sequence was generated from both ends of subcloned restriction fragments using universal M13 sequencing primers. The internal sequence of large fragments as well as the complementary strands of all fragments were determined using oligonucleotide primers synthesized in accordance with preceding sequences. Sequencing artifacts generated as the result of G-C compression were avoided by determining all sequences using both dGTP and the nucleotide analogue deoxyinosinetriphosphate (dITP).

Northern Blots

RNA for Northern analysis was prepared from the following: NMuMG cells, adult liver, newborn skin, midpregnant mammary gland, adult cerebrum, and skeletal and cardiac muscle. Excised tissues were ground to a fine powder in the presence of liquid nitrogen and transferred directly to RNA extraction buffer (see above); the NMuMG cells were extracted after washing with PBS as described above. The samples were vigorously vortexed, an equal volume of 10 mM Tris, pH 8.0, 1 mM EDTA, and 1% SDS added, and subsequently extracted exhaustively with 24:24:1 Tris-saturated phenol/chloroform/isoamyl alcohol followed by a single extraction with

1. *Abbreviations used in this paper*: NMuMG, normal murine mammary gland; TBST, Tris-buffered saline/Triton.

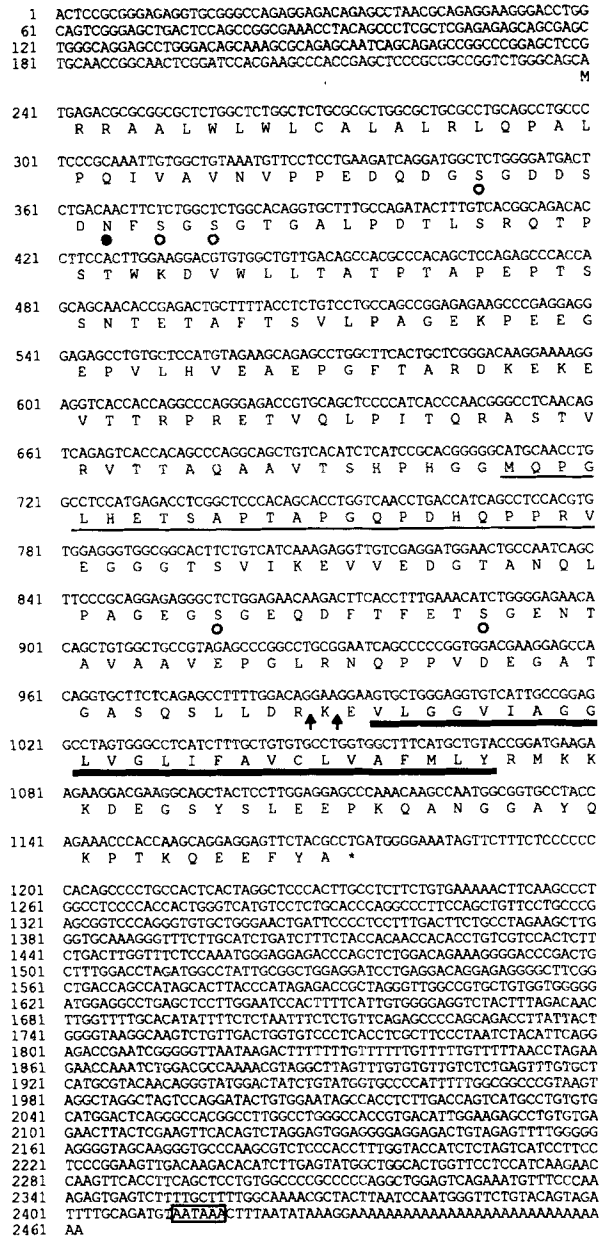

Figure 1. Sequence of syndecan cDNA clones, corresponding amino acid sequence, and important structural features. The putative transmembrane sequence is indicated by thick underlining while that sequence corresponding to the experimentally determined peptide sequence is denoted by thin underlining. Glycosylation sites are identified by circles; (o) corresponds to potential glycosaminoglycan attachment sites and (●) to a potential asparagine-linked glycosylation site. Arrows mark the putative site of ectodomain cleavage from the cell surface. The polyadenylation signal sequence AATAAA is boxed.

24:1 chloroform/isoamyl alcohol. After precipitation with an equal volume of 2-propanol, and resuspension in 10 mM Tris, pH 7.5, 1 mM EDTA, RNA was precipitated by addition of 0.33 vol of 10 M LiCl. Poly(A) RNA was prepared by oligo d(T) chromatography as described above.

For Northern analysis, 2 μg of each poly(A) RNA sample was separated by electrophoresis in 1.2% agarose-formaldehyde gels in the presence of MOPS (Sigma Chemical Co., St. Louis, MO) acetate buffer, pH 7.0 (Maniatis et al., 1982). After alkali treatment (Danielsen et al., 1986) and neutralization in transfer buffer (0.025 M sodium phosphate, pH 6.5), the gel was blotted to GeneScreen (New England Nuclear, Boston, MA) and the RNA immobilized by UV cross-linking (Church and Gilbert, 1984). Hybridization probes were prepared by in vitro transcription of the 5' Eco RI–Sac I fragment of PM-4 subcloned into pGEM 3 (Melton et al., 1984). Blots were prehybridized at 61°C in 50% formamide, 1% SDS, 5X SSPE, 0.1% ficoll, 0.1% polyvinylpyrrolidone, and 100 μg/ml denatured salmon sperm DNA. Hybridization was for 16 h at 61°C in the same buffer containing 5×10^6 cpm/ml of RNA probe. Filters were washed 2 × 15 min at room temperature in 5% SDS/1X SSPE and 6 × 30 min at 67°C in 1% SDS/0.1X SSPE. Molecular sizes were determined relative to ethidium bromide–stained molecular mass markers (Bethesda Research Laboratories) and 18-S and 28-S ribosomal RNA.

Preparation and Use of Antibodies to Synthetic Peptides

A 7-amino acid (^{14}C-labeled) synthetic peptide, corresponding to the predicted COOH terminus of syndecan (Fig. 1), was generously provided by Dr. Russell Doolittle (University of California at San Diego). The NH_2-terminal lysine of this peptide was cross-linked by glutaraldehyde to keyhole limpet hemocyanin (Calbiochem-Behring Corp., San Diego, CA) for immunization and BSA (fraction V, Sigma Chemical Co.) for screening as described by Doolittle (1986). Briefly, 10 mg carrier protein was dissolved in 0.5 ml of 0.4 M phosphate, pH 7.5, mixed with 7.5 μmol of peptide in 1.5 ml water and 1.0 ml of 20 mM glutaraldehyde (Ted Pella, Inc., Tustin, CA) was added dropwise with stirring over the course of 5 min. After continuous stirring at room temperature for 30 min, 0.25 ml of 1 M glycine was added to block unreacted glutaraldehyde and the stirring resumed for an additional 30 min. The product was dialyzed exhaustively against PBS and incorporation determined by TCA precipitation and liquid scintillation counting. This procedure resulted in the attachment of 17 mol of synthetic peptide/mole of carrier protein.

For immunization, 1.25 mg of synthetic peptide–keyhole limpet hemocyanin conjugate in 0.5 ml PBS, pH 7.5, was mixed with 0.5 ml complete Freund's adjuvant (Gibco Laboratories). The emulsion was delivered by intramuscular injections, 0.1 ml in each of 10 sites, into a 3-mo-old New Zealand white rabbit. After 2 wk, the immunization was repeated with an identical quantity of immunogen. 10 d later, the rabbit was injected with Innovar 0.125 ml/kg subcutaneously and was bled from the central auricular artery. Innovar was reversed with Nalline 0.2 ml/kg, and serum was prepared from the collected blood.

The native lipophilic form of syndecan (Sanderson and Bernfield, 1988) and the nonlipophilic medium ectodomain form (Jalkanen et al., 1987) were isolated and purified as described elsewhere and assessed for their reactivity to the immune sera. A cationic nylon membrane, GeneTrans (Plasco Inc., Woburn, MA), was placed into an immunodot apparatus (V&P Scientific, San Diego, CA) and samples of intact syndecan and the ectodomain (0.5, 5, 50, and 500 ng) were loaded on the membrane using mild vacuum. After loading, remaining binding sites on the membrane were blocked by a 1-hr incubation in a solution containing 0.5% BSA, 3% Carnation instant nonfat dry milk, 10 mM Tris (Sigma Chemical Co.), pH 8.0, 0.15 M NaCl, and 0.3% Tween-20. Incubation with immune serum was performed at dilutions of 1:200 for the anticytoplasmic domain, and 1:500 for the antiectodomain in 10 mM Tris, pH 7.4, 0.15 M NaCl, and 0.3% Tween-20 (TBST) for 30 min at room temperature. The membrane was washed for 60 min at room temperature with 10 changes of TBST and then incubated for 30 min with 1:7,500 dilution of alkaline phosphatase goat anti-rabbit IgG (Promega Biotec). After washing for 60 min with 10 changes of TBST, the immobilized alkaline phosphatase was visualized with nitro blue tetrazolium (Promega Biotec) 330 μg/ml and 5-bromo-4-chloro-3-indolyl phosphate (Promega Biotec) 165 μg/ml in 100 mM Tris, pH 9.5, 100 mM NaCl, and 5 mM $MgCl_2$.

Results

Isolation of Syndecan cDNA Clones

Proteoglycan purified from normal murine mammary gland (NMuMG) cell-conditioned medium is both immunologically (Jalkanen et al., 1987, 1988) and chemically (Jalkanen et al., 1987; Weitzhandler et al. 1988) indistinguishable from the trypsin-released ectodomain of syndecan. Therefore, we used a serum antibody directed against polypeptide determinants on this proteoglycan ectodomain to identify

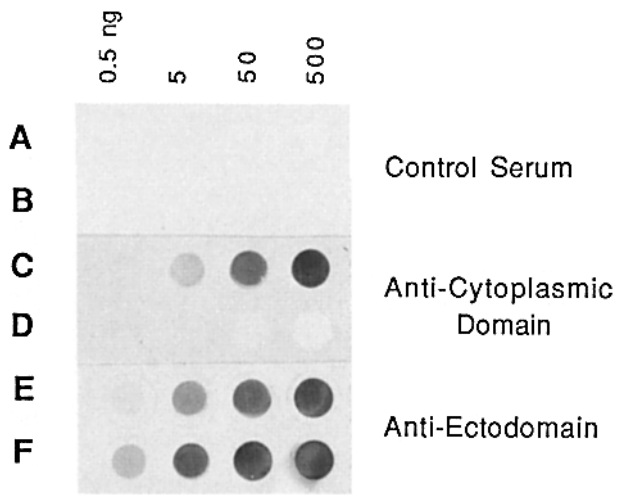

Figure 2. Antibodies to a synthetic peptide corresponding to the cytoplasmic domain distinguishes intact syndecan from the ectodomain. Purified intact syndecan (*A*, *C*, and *E*) and the medium ectodomain (*B*, *D*, and *F*) were loaded on Genetrans cationic nylon membrane at dilutions of 0.5, 5, 50, and 500 ng. The membrane was treated with sera and the bound antibody detected with alkaline phosphatase-conjugated anti-rabbit IgG. (*A* and *B*) Control serum; (*C* and *D*) anticytoplasmic domain serum; and (*E* and *F*) antiectodomain serum. The anticytoplasmic domain serum reacts only with the intact syndecan molecule extracted from the cell surface, whereas the antiectodomain serum reacts both with the intact molecule and the nonlipophilic medium ectodomain.

syndecan cDNA clones in a λ gt-11 expression library prepared from NMuMG cells. The unamplified portion of the NMuMG λ gt-11 cDNA library was initially screened with whole immune serum previously absorbed against host Y 1090 *Escherichia coli*. Screening of 7.5×10^5 recombinants yielded four positive cDNA clones. Hybridization analysis revealed that these four clones represented the cDNA of three distinct gene products, only two of which produced fusion proteins that reacted with immune serum affinity purified against the ectodomain of syndecan. One of these, PM-4, is a 2.1-kb clone that appears to code for the core protein of syndecan.

Proof of the identity of PM-4 was obtained by two independent methods. First, DNA sequence analysis revealed a derived protein sequence that matched the partial amino acid sequence of a cyanogen bromide-cleaved fragment of the ectodomain of syndecan (Weitzhandler, 1988; Fig. 1). Second, syndecan purified from NMuMG cells reacted with an immune serum prepared against a synthetic peptide containing the COOH-terminal seven amino acids (Lys-Gln-Glu-Glu-Phe-Tyr-Ala) of the PM-4-derived protein sequence. This immune serum failed to react with the ectodomain that lacks the putative cytoplasmic domain (Fig. 2). Furthermore, this serum does not cross-react with any other cellular proteins as assessed by Western blotting of total cell extracts (data not shown).

Further screening of 1.5×10^6 recombinants from the amplified NMuMG λ gt-11 cDNA library produced no additional positive clones. This might be explained by retarded growth of clones expressing the core protein caused by toxic influences on the growth of *E. coli* (Huynh et al., 1985). Indeed, infection with the λ gt-11 clone PM-4 consistently yielded low phage titers. PM-4 was, therefore, subcloned into a plasmid vector (pUC-13) to facilitate its characterization.

Two approaches were used to isolate the complete coding region of syndecan. First, to select clones containing an extended 5′ end, a λ gt-10 cDNA library prepared from NMuMG cells was screened with a 250-bp Eco RI-Hinc II probe from the 5′ end of PM-4. This screening yielded two clones, 4-19b and 4-15, which were further characterized (Fig. 3). Second, a primer extended λ gt-10 cDNA library, prepared with liver poly(A) RNA and a synthetic oligonucleotide complementary to a site near the 5′ end of PM-4 (positions 848–865 in Fig. 1), was screened with the same 250-bp Eco RI-Hinc II probe. Several independent clones were characterized from this library; each contained a 5′ sequence identical with that of clone 4-19b (Fig. 3).

Syndecan cDNA Sequence

The clones were sequenced in M13 by the method of Sanger et al. (1977), using a strategy summarized in Fig. 3. The sequence was established by sequencing both strands from a minimum of two of the independently isolated clones shown in Fig. 3. Because the syndecan cDNA contains a large number of guanosine- and cytosine-rich regions, sequences were obtained with both dGTP and dITP to avoid potential sequencing artifacts caused by these regions.

The cDNA (Fig. 1) has the following features. The first AUG is at position 240. This putative initiation codon is preceded by two inframe termination codons (TAA and TGA at positions 39 and 72, respectively) and followed by a 930-base open reading frame that ends at position 1,173 with a TGA termination codon. Following the putative coding region are 1,243 bases of 3′ untranslated sequence that ends with the poly(A) addition sequence (AATAAA) followed 14 bases later by a poly(A) stretch. Because each of the primer extended clones has the same 5′ end as the largest cDNA clone from the NMuMG library, M-4-19b, this sequence likely includes the complete 5′ untranslated region of syndecan.

Syndecan Protein Structure

Starting at the indicated AUG (Fig. 1), the syndecan cDNA codes for a protein of 311 amino acids containing two hydrophobic stretches. This sequence corresponds with a predicted molecular mass of 32,868 D and is unique; no statistically significant similarities were detected when this sequence was compared with either the National Biomedical Research Foundation's Protein Sequence or the translated National Institutes of Health GenBank databases. The derived sequence suggests several domains and structural features; their presumed arrangement is summarized in Fig. 6.

The first hydrophobic stretch consists of 12 amino acids, including a cysteine, beginning shortly after the presumptive start methionine. Because syndecan is oriented with its NH_2 terminus outside of the plasma membrane (Weitzhandler et al., 1988), this hydrophobic sequence could represent a signal peptide. The NH_2 terminus of the mature syndecan core protein is blocked (Weitzhandler et al., 1988), and, therefore, it has not been possible to determine the NH_2 terminus directly. A likely site for signal peptidase cleavage fol-

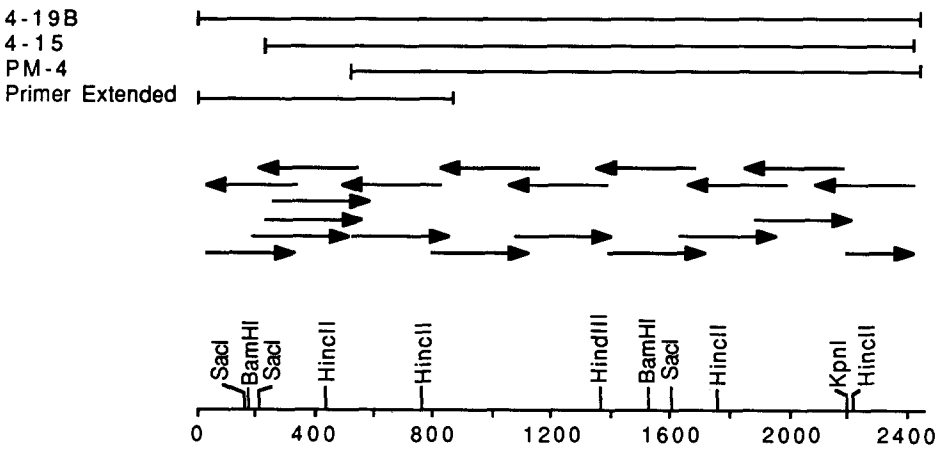

Figure 3. Restriction map and sequencing strategy of syndecan cDNA clones. The size and orientation of syndecan cDNA clones 4-19 B, 4-15, PM-4, and the primer extended clones are shown next to a restriction map of the cDNAs. The short arrows summarize the direction and length of sequencing of various portions of the cDNAs.

lows amino acid residue 17 (Fig. 1) in the predicted sequence. Cleavage at this site would generate an NH_2-terminal glutamine that could readily cyclize forming a pyrrolidone carboxylyl residue and thus a blocked NH_2 terminus, as exists in a number of eukaryotic proteins (Allen, 1981).

The second hydrophobic stretch is a sequence near the COOH terminus that has characteristics of a transmembrane domain (Fig. 1, *thick underline*). This sequence is a highly hydrophobic stretch of 25 residues followed immediately by a series of highly charged residues, consistent with the stop transfer signals found following most membrane spanning domains (Sabatini et al., 1982). This domain also contains the only cysteine and one of the four tyrosines in what we presume is the mature protein sequence.

The putative transmembrane domain defines two hydrophilic domains of the syndecan core protein, a putative extracellular domain consisting of ~235 amino acids, and a smaller putative cytoplasmic domain consisting of 34 amino acids. This orientation with respect to the plasma membrane is confirmed by the reactivity of immune serum directed either against a peptide containing the COOH-terminal seven amino acids or against the ectodomain of syndecan. The anti-COOH-terminus immune serum recognizes the hydrophobic native form of syndecan (Fig. 2 *C*) but is unreactive with the nonhydrophobic ectodomain (Fig. 2 *D*). In contrast, the anti-ectodomain immune serum recognizes both forms of the molecule (Fig. 2, *E* and *F*).

The putative cytoplasmic domain contains three tyrosine residues, but the sequences adjacent to these tyrosines are not similar to the presently identified consensus sequences for tyrosine phosphorylation (Hunter and Cooper, 1985). This domain presumably has protein binding activity because the intact proteoglycan, but not the ectodomain, cosediments with F-actin (Rapraeger and Bernfield, 1982) and because syndecan associates with the actin-containing cytoskeleton when cross-linked at the cell surface (Rapraeger et al., 1986).

The putative extracellular domain has several sequence characteristics that correspond with the known properties of this proteoglycan. The ectodomain of syndecan is shed by cleavage from its membrane anchor (Jalkanen et al., 1987), and an indistinguishable molecule is released from the cell surface by mild trypsin treatment (Jalkanen et al., 1987; Weitzhandler et al., 1988). The only dibasic sequence (Arg-Lys) in this extracellular domain is located adjacent to the putative transmembrane domain at residues 250–251 (Fig. 1, *arrows*). This location places the cleavage site adjacent to the plasma membrane. The putative extracellular domain lacks cysteine, thus eliminating disulfide bridges as a means of generating secondary structure in this molecule.

The ectodomain contains both heparan sulfate and chondroitin sulfate chains (Rapraeger et al., 1985). The serine hydoxyl group of ser-gly sequences are the attachment sites for these glycosaminoglycan chains (Rodén, 1980; Dorfman, 1981). Syndecan possesses five such potential glycosaminoglycan attachment sites, all within the putative extracellular domain; three such serines are clustered near the NH_2 terminus at residues 37, 45, and 47, and the remaining two are clustered near the membrane at residues 207 and 217 (Fig. 1, *open circles*).

The ectodomain from NMuMG cells is insensitive to digestion by *N*-glycosidase F, as assessed by PAGE (Weitzhandler et al., 1988). The putative extracellular domain contains a single canonical sequence for the attachment of N-linked oligosaccharide (Fig. 1, *solid circle*). The serine in this Asn-Xaa-Ser sequence is the putative glycosaminoglycan attachment serine at position 45.

Northern Blot Analysis of Syndecan Expression in Tissues

Poly(A) RNA was prepared from NMuMG cells and a series of tissues (newborn skin, adult liver, midpregnant mammary gland, adult cerebrum, and adult skeletal and cardiac muscle). Immunohistological studies of these tissues showed that the epithelial components of skin, liver, and midpregnant mammary gland react with a monoclonal antibody (281-2) directed against the ectodomain, while cerebrum and striated muscle do not (Hayashi et al., 1987). Northern blot analysis of these poly(A) RNA preparations reveals two mRNA bands in NMuMG cells (Fig. 4 *A*) as well as in skin, liver, and mammary gland tissues (Fig. 4 *B*); one band is at 2.6 and the other at 3.4 kb. The apparent lower level of expression found in midpregnant mammary gland, as compared with skin and liver, is consistent with the relative paucity of epithelial cells in the mammary gland. Longer exposures of the Northern blot shown in Fig. 4, as well as others containing larger quantities of poly(A) RNA, verify that the mam-

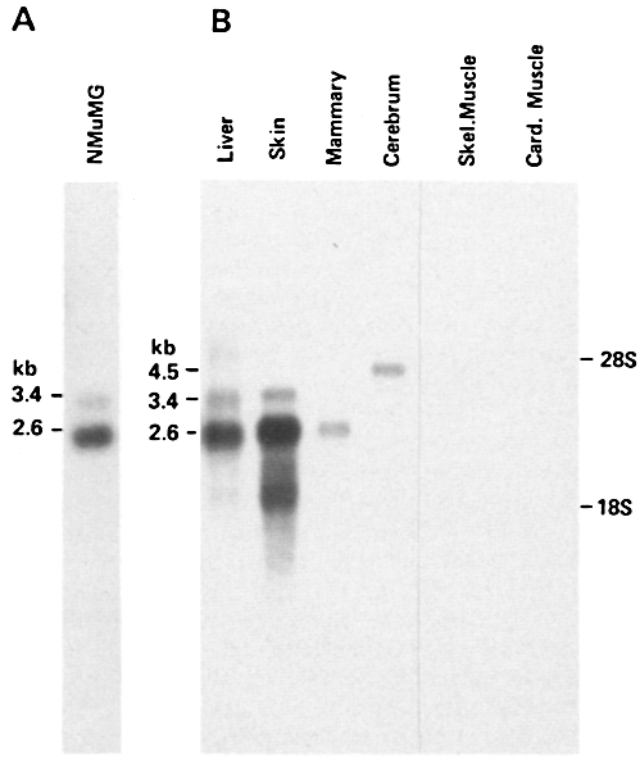

Figure 4. Northern blot analysis for syndecan mRNA. NMuMG cells (*A*) and several tissues (*B*) were assessed for expression of syndecan mRNA. 2 μg of poly(A) RNA from each source was separated by electrophoresis in 1.4% agarose-formaldehyde gel in MOPS buffer, transferred, and hybridized as described in Materials and Methods. NMuMG cells, and liver, skin, and mammary gland tissue reveal messages at 2.6 and 3.4 kb. Hybridization to cerebral RNA reveals a unique message at 4.5 kb, while cardiac and skeletal muscle shows the absence of syndecan message. *A* and *B* represent 4- and 15-h exposures, respectively.

mary gland expresses both the 2.6- and the 3.4-kb messages (data not shown). Scanning densitometry shows that these two messages are present at a nearly constant relative abundance of 3:1 (2.6 kb/3.4 kb) in NMuMG cells and in skin, liver, and mammary gland tissues (data not shown). As expected from the immunohistology, neither of these mRNAs was present in detectable amounts in cerebrum and striated muscle tissues (skeletal and cardiac). Interestingly, however, Northern analysis consistently detected a distinct 4.5-kb mRNA in the cerebrum (Fig. 4 *B*). The relationship of this message to that of syndecan awaits further investigation.

Discussion

We report here the first characterization of cDNA clones for an integral membrane proteoglycan that represents a class of multi-ligand matrix receptors. This proteoglycan, which we call syndecan (from the Greek, *syndein*, to bind together), binds cells to collagen types I, III, and V (Koda et al., 1985), fibronectin (Saunders and Bernfield, 1988), and thrombospondin (Sun, X. et al., 1989).

Size of the Syndecan Core Protein

The derived protein sequence of syndecan is unique. Comparisons with the National Biomedical Research Foundation and the translated NIH-Genebank databases detected no statistically significant similarities. The nascent polypeptide sequence is 311 amino acids and has a molecular mass of 32,868 D. Treatment of syndecan with heparitinase and chondroitinase ABC generates a protein with relative mobility of ~69 kD versus globular molecular mass markers on a gradient SDS-PAGE system (Sanderson et al., 1988). Treatment of the ectodomain with anhydrous HF for 1.5 h at 0°C (Mort and Lamport, 1977), yields a protein that migrates as a broad band at ~46 kD (Weitzhandler et al., 1988). These core protein sizes are larger than would be predicted based on the cDNA and any incompletely removed carbohydrate. Syndecan is not a disulfide cross-linked dimer. Its migration on SDS-PAGE is unchanged after DTT treatment; its cyanogen bromide cleavage product produces a single signal during amino acid sequencing; and the single cysteine in the predicted mature protein is located in the transmembrane domain. It also does not appear to be cross-linked by lysyl oxidase- or transglutaminase-mediated reactions because β-aminopropionitrile (Franzblau and Faris, 1982) and monodansylcadaverine (Bowness, 1987) treatments of NMuMG cells do not change its mobility on SDS-PAGE in preliminary studies (data not shown). Proteins with regions rich in proline, alanine, and highly charged amino acids have highly extended conformations and anomalously slow mobilities in SDS-PAGE (Guest et al., 1985). These amino acids are abundant in syndecan and a Chou and Fasman secondary structure prediction is consistent with large regions of extended conformation. In vitro translation of synthetic mRNA corresponding to the coding region of syndecan (Sac I–Hind III fragment of clone 4-19b) produces a nascent polypeptide of ~45 kD (data not shown). Therefore, while we have not excluded the possibility of other posttranslational modifications, the bulk of the size difference probably reflects anomalous gel migration on SDS-PAGE.

Syndecan Contains Distinct Functional Domains

Cell biological studies of syndecan have implied three functional domains; an extracellular domain, and by inference, transmembrane and cytoplasmic domains (Rapraeger and Bernfield, 1985; Rapraeger et al., 1986). The amino acid sequence derived from the syndecan cDNA demonstrates these domains directly.

The transmembrane domain was inferred from the physical properties of syndecan. Syndecan is mobile in the plane of the membrane (Rapraeger et al., 1986) and is lipophilic, demonstrated by vectorial intercalation into liposomes (Rapraeger and Bernfield, 1985), binding to hydrophobic materials (Jalkanen et al., 1987), and extraction with detergents (Rapraeger et al., 1986). However, lipophilicity does not assure that anchorage to the cell surface is via a transmembrane domain. Many proteins are anchored to the membrane via phosphatidyl inositol. Indeed, such an anchorage has been suggested for a proteoglycan on the surface of rat hepatocytes (Ishihara et al., 1987). However, the derived COOH-terminal sequence of syndecan contains both a characteristic transmembrane domain and a 34–amino acid putative cytoplasmic domain. These are not present in the proteins linked to the membrane by phosphatidyl inositol (Ferguson and Williams, 1988).

The cytoplasmic domain was inferred from properties in-

dicating that syndecan associates with the actin cytoskeleton. Syndecan co-sediments with F-actin (Rapraeger and Bernfield, 1982), resists detergent extraction from cells both at low pH and when cross-linked at the cell surface by antibodies, and colocalizes in polarized NMuMG cells with the actin-containing cytoskeleton (Rapraeger et al., 1986). An immune serum generated against a synthetic peptide from the COOH-terminus of the derived protein sequence reacts with native syndecan extracted from NMuMG cells but not with the ectodomain (Fig. 2), providing direct evidence for the cytoplasmic domain.

Syndecan and the β_1-chain of integrin contain small cytoplasmic domains, 34 and 47 amino acids, respectively (Tamkun et al., 1986; Argraves et al., 1987), and both are thought to associate with the cytoskeleton after the binding of their extracellular matrix ligands (Chen et al., 1985; Damsky et al., 1985; Rapraeger et al., 1986). The sequences of these domains show no obvious similarities, and the differences likely reflect distinct functions. It will be interesting to assess whether the cytoplasmic domains of various cell surface proteoglycans show the same extent of conservation as is found among the β-chains of various integrins (Hynes, 1987).

The ectodomain is released from NMuMG cell surfaces during cell culture, rapidly in response to cell rounding, or by mild trypsin treatment. The putative extracellular domain of syndecan contains a single dibasic site near the plasma membrane at which cleavage of syndecan from the cell surface undoubtedly occurs. Because the endogenously shed ectodomain of syndecan is indistinguishable from the trypsin-released form (Jalkanen et al., 1987; Weitzhandler et al., 1988), a cell surface trypsinlike protease might be involved. Shedding during cell culture is from the apical surface of NMuMG cells and is markedly reduced when syndecan polarizes to the basolateral surface (Saunders, S., and M. Bernfield, unpublished results). However, when these cells are released from the substratum, destroying their polarity, the ectodomain is rapidly shed. These results suggest that a cell surface protease is involved and identification of the putative cleavage site will now allow more detailed investigation of this activity.

Putative Attachment Sites for Heparan Sulfate and Chondroitin Sulfate

Syndecan isolated from several sources is a hybrid proteoglycan, containing both chondroitin sulfate and heparan sulfate. These chains are linked via a xyloside to serine residues in proteins (Rodén, 1980; Dorfman, 1981). Regulating the elaboration of both chondroitin sulfate and heparan sulfate chains on the same core protein is a significant problem because the initial four saccharides are identical. The synthesis of both types of chains is initiated by a xylosyltransferase that resides in either the endoplasmic reticulum or the Golgi region (for review see Farquhar, 1985) and by three Golgi-localized glycosyltransferases (Geetha-Habib et al., 1984). Specific chain elongation subsequently involves the sequential action of an *N*-acetylgalactosaminyltransferase and a glucuronosyltransferase for chondroitin sulfate, and an *N*-acetylglucosaminyltransferase and a glucuronosyltransferase for heparan sulfate. This specific chain elongation must involve recognition of unique structural features of the core protein, indicating that distinct peptide sequences might exist at chondroitin sulfate versus heparan sulfate attachment sites. The presence of both chondroitin sulfate and heparan sulfate on syndecan provides the opportunity to assess the relationship between these attachment sites. Based on the core protein sequence of three chondroitin sulfate proteoglycans (PG-19, PG-40, and invariant chain) and the reactivity of a xylosyltransferase with synthetic peptides, Bourdon et al. (1987) proposed that the xylose acceptor sequence for chondroitin sulfate in these proteoglycans is acidic-acidic-Xaa-Ser-Gly-Xaa-Gly. Syndecan contains five ser-gly sequences; the two in its single Ser-Gly-Ser-Gly repeat closely match this acceptor sequence (Fig. 5 *A*). Interestingly, although this consensus acceptor sequence is located near the NH_2 terminus of syndecan and near the COOH terminus of invariant chain, it is distant from the plasma membrane on both proteins.

Syndecan contains three potential ser-gly glycosaminoglycan attachment sites that contain some features of this consensus acceptor sequence as well as unique features (Fig. 5 *B*). Though each of these three sequences retains an acidic amino acid two residues NH_2 terminal to the acceptor ser-gly, they lack the consensus glycine that is two residues COOH terminal to the ser-gly. This omission does not preclude this sequence from serving as a xylosyltransferase acceptor because it is also omitted from the gly-ser site of type IX collagen (Huber et al., 1988). The unique feature of these three sequences is the consistent finding of an acidic amino acid COOH terminal to the ser-gly (Fig. 5 *B*). In contrast, the analogous amino acids in the chondroitin sulfate proteoglycans PG-19, PG-40, and invariant chain are either uncharged or hydrophobic. These three sites in syndecan may represent unique recognition sequences for the elongation of heparan sulfate chains.

The number of chondroitin sulfate chains on syndecan apparently differs in cells of distinct cellular organization (Sanderson and Bernfield, 1988) and changes in response to

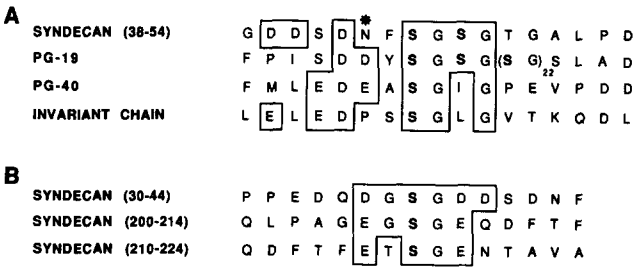

Figure 5. Putative glycosylation sites of the syndecan core protein. The potential glycosaminoglycan attachment sites of the hybrid proteoglycan syndecan are compared with the known attachment sites of the chondroitin sulfate proteoglycans PG-19, PG-40, and invariant chain. The putative glycosylated serine in each case is indicated in bold. Identical or conserved substitutions of acidic residues have been boxed and the position of these residues in syndecan (see Fig. 1) are in parenthesis. (*A*) The potential glycosylation site between residues 38–54 of syndecan shows the sequence Ser-Gly-Xaa-Gly with acidic residues NH_2 terminal to the serine, strikingly similar to the consensus sequence described for attachment of chondroitin sulfate chains (Bourdon et al., 1987). The asterisk denotes the single potential N-linked glycosylation site. (*B*) The three remaining potential glycosylation sites of syndecan between residues 30–44, 200–214, and 210–224 are similar to each other and distinct from the proposed chondroitin sulfate attachment sites. These Ser-Gly are not followed by Xaa-Gly and contain acidic residues both COOH and NH_2 terminal to the serine. These sites are proposed to be heparan sulfate attachment sites.

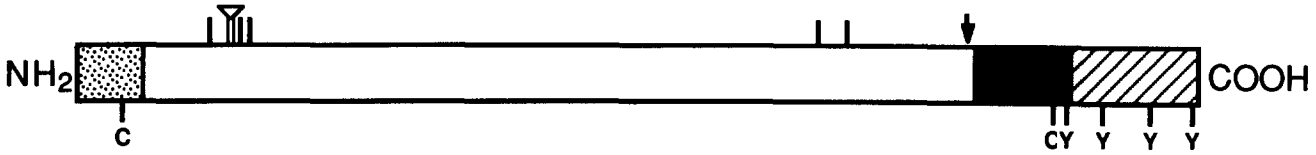

Figure 6. Schematic structure of syndecan core protein. A structural organization is proposed for the syndecan core protein. Domains are indicated by boxes. The putative signal sequence is stippled, the extracellular domain open, the transmembrane domain solid, and the cytoplasmic domain diagonally striped. Potential glycosaminoglycan attachment sites are indicated by the solid bars, and the single potential N-linked glycosylation site is indicated by an open bar and triangle. An arrow indicates the dibasic protease cleavage site, and the only cysteine and tyrosine residues in the predicted core protein are labeled.

TGF-β (Rasmussen and Rapraeger, 1988), implying that each potential glycosaminoglycan attachment site is not always used. A possible novel regulatory mechanism for this variation is suggested by the location in syndecan of its single potential N-linked glycosylation site, Asn-Phe-Ser, at residues 43–45. This site is located within the putative chondroitin sulfate attachment sequence, and the attachment of an N-linked sugar at this site would likely prevent subsequent recognition by the xylosyltransferase.

Structural and Functional Relationship of Syndecan and Integrins

A cell type can contain both syndecan and integrins, and these receptors can bind the cells to virtually identical matrix ligands (Saunders and Bernfield, 1988). However, these receptors are both structurally and functionally distinct. Syndecan binds principally via its heparan sulfate chains, and, while the precise ligand binding sites on the integrins are not well defined, they are apparently generated by combinations of specific α and β chains (Hynes, 1987; Ruoslahti and Pierschbacher, 1987). Whereas many of the matrix-binding integrins bind to an RGD-containing sequence (Pierschbacher and Ruoslahti, 1984), it is unknown whether syndecan binds to specific protein sequence(s). Structurally, the integrins are heterodimers, and each oligomer is much larger than the single polypeptide of syndecan; however, both syndecan and the integrins are transmembrane proteins containing short cytoplasmic domains that associate with the actin cytoskeleton (Chen et al., 1985; Damsky et al., 1985; Rapraeger et al., 1986).

Syndecan is present primarily on epithelial cells in mature tissues (Hayashi et al., 1987) but is also found on cells of the B-cell lineage (Sanderson, R., P. Lalor, and M. Bernfield, unpublished results) and can be expressed in embryonic and undifferentiated mesenchymal cells (Theselff, 1988; Trautman, M. S., J. Kimelman, and M. Bernfield, manuscript submitted for publication). Integrins are expressed on these cells but also on a wide variety of hematopoietic precursor, mesenchymal, and neural cells (Buck and Horwitz, 1987). Cells can contain several integrins with seemingly redundant binding specificities and similar modes of transmembrane action. Why should these cells contain yet another class of matrix receptor with similar binding specificity and mode of action? We favor the hypotheses that (*a*) like the integrins, syndecan represents a collection of receptors that differ in their interactions because of their diverse number, type, and sequence of glycosaminoglycan chains, but that (*b*) matrix binding via the integrins and syndecans has a different role in regulating cell behavior.

Interactions of cells with fibronectin indicate that binding via integrin and syndecan receptors can result in distinct cell behaviors. The fibronectin domain containing the sequence RGDS is recognized by an integrin and promotes cell adhesion; synthetic peptides containing this sequence interfere with cell–fibronectin interactions (Pierschbacher and Ruoslahti, 1984). A heparin-binding domain near the COOH terminus is recognized by syndecan (Saunders and Bernfield, 1988); synthetic peptides containing sequences from this site bind heparin directly and promote cell adhesion (McCarthy et al., 1988). Cell interactions with both the RGD-containing and heparin-binding domains are required for full elaboration of focal contacts and cytoskeletal organization (Izzard et al., 1986; Woods et al., 1986). However, focal contacts do not form on substrata containing a mixture of the two isolated domains (Izzard et al., 1986). These results suggest that the integrin and syndecan class of receptors cooperate when bound to the same polypeptide. Because heparan sulfate induces a conformational change in fibronectin (Osterlund et al., 1985), this cooperation could result from changes induced by syndecan at the RGDS-containing domain. Based on several such studies, we have proposed that syndecan stabilizes cell interactions with fibronectin-containing matrices, whereas the integrins are involved in initial recognition and perhaps cell migration over such matrices (Saunders and Bernfield, 1988).

The notion that syndecan mediates stable cell interactions with matrix is consistent with its structure. Because there are multiple putative heparan sulfate attachment sites, a single syndecan molecule could have multiple interactions with matrix components via these glycosaminoglycan chains. This model could explain the high affinity of syndecan for fibrillar collagen (K_d = 1 nM) (Koda et al., 1985), three orders of magnitude greater than the affinity of integrins (Buck and Horwitz, 1987). Assuming a similar number of interacting receptors, this affinity translates into exceedingly tight associations that presumably are more physically stable than those mediated by lower affinity receptors. The syndecan structure also provides a unique mechanism to reduce the stability of the linkage: a putative protease-susceptible site adjacent to the transmembrane domain that allows its extracellular domain to be cleaved from the cells thus releasing cells from the matrix.

Syndecan Expression Is Tissue Specific

Of a wide variety of mature tissues examined with antibody 281-2, syndecan is expressed mainly in epithelia (Hayashi et al., 1987). Northern blot analysis of mRNA revealed two mRNA species at 2.6 and 3.4 kb (constant ratio 3:1, respec-

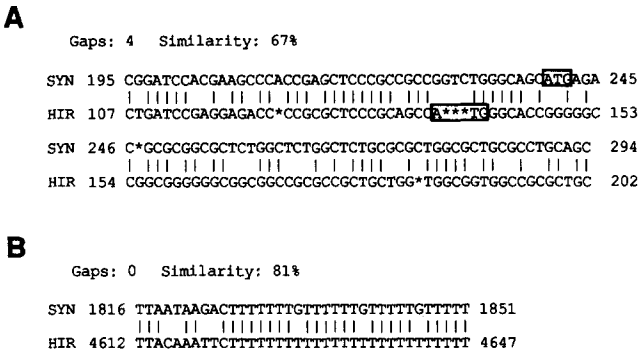

Figure 7. DNA sequence similarities between murine syndecan and human insulin receptor. The DNA sequence of the 5′ untranslated-translational start and 3′ untranslated regions of syndecan (*SYN*) are highly similar with the same regions of the human insulin receptor (*HIR*). Bestfit alignments of these sequences are indicated with asterisks marking gap nucleotides. (*A*) Positions 195–294 of the syndecan cDNA are 67% identical, allowing four small gaps (one 3-nucleotide and three single-nucleotide), with the comparable region of the human insulin receptor cDNA. The putative start ATGs of these two molecules are indicated by boxes. (*B*) Positions 1,816 to 1,851 from the 3′ untranslated region of the syndecan cDNA are 80% identical (no gaps) with a sequence from the 3′ untranslated region of the human insulin receptor cDNA.

tively) in NMuMG cells as well as skin, liver, and midpregnant mammary gland, all containing immunoreactive syndecan. In contrast, these two mRNAs were undetectable in cardiac and skeletal muscle, tissues of mesenchymal origin that do not stain with 281-2. A 4.5-kb mRNA was detected in adult cerebrum, which does not react in fixed tissue sections with the antibody (Hayashi et al., 1987).

The cDNA sequence reported here corresponds to the smaller (2.6 kb) and more abundant of the two mRNAs. Though the relationship between the 2.6- and 3.4-kb mRNAs is unknown, they are likely generated by usage of alternative polyadenylation sites. Probes from both 5′ and 3′ regions of the syndecan cDNA hybridized identically to these two mRNAs in Northern blot analysis (data not shown). Moreover, the primer-extended library contained clones identical to the 5′ end of clone 4-19B. The relationship of the 4.5-kb mRNA identified in cerebrum to the others is unknown because clones have not yet been characterized from cerebral cDNA libraries.

Untranslated Sequences Shared with Insulin Receptor cDNA

Sequence alignments demonstrate a remarkable similarity at the nucleotide level between the mouse syndecan and human insulin receptor cDNA sequences (Ebina et al., 1985). Alignment of these sequences (University of Wisconsin GCG Bestfit program) places the putative start ATGs near the middle of a region of similarity; a 99-bp region of syndecan that spans its 5′ untranslated and initial coding sequences is 67% identical, with four small gaps, to the analogous region of the human insulin receptor (Fig. 7 *A*). The location of this similarity and the large size of the 5′ untranslated regions suggest that these sequences are shared translational control elements, as has been described for the 5′ untranslated region of the mRNAs for ferritin (Aziz and Munro, 1987; Casey et al., 1988) and the B polypeptide of platelet-derived growth factor (Ratner et al., 1987). Remarkably, there is a second region of similarity between these cDNAs in their 3′ untranslated regions; a 35-bp T-rich sequence of syndecan is 80% identical (no gaps) with a sequence of the human insulin receptor (Fig. 7 *B*). These identical sequences in both 5′ and 3′ untranslated regions between the mouse syndecan and human insulin receptor mRNAs suggest that posttranscriptional controls are shared by these two molecules.

The immunohistology, derived structure, and mRNA analyses are consistent with syndecan's proposed function of stabilizing epithelial sheets by anchoring the cells to the matrix. However, heparan sulfate proteoglycans are also known to bind growth factors. Basic fibroblast growth factor binds to heparan sulfate proteoglycans from endothelial cells (Saksela et al., 1988), an activity duplicated by syndecan. Transforming growth factor-β binds to the core protein of a hybrid cell surface proteoglycan found on a variety of cell types (Segarini and Seyedin, 1988). Because insulin is a known growth factor, the sequence similarity between the syndecan and insulin receptor cDNAs makes a growth factor binding function of syndecan more tenable. Binding of syndecan to interstitial matrix at the basolateral surface of simple epithelia might simultaneously anchor the cells to the matrix and serve as a sequestration point for growth factors at this cell surface. This multifunctionality might serve to integrate cellular behavior in response to growth factors and extracellular matrix.

The authors thank G. Stark, J. Nathans, and Russell F. Doolittle for their helpful discussions and Robert Simoni and R. Kumar Kadiyala for their assistance.

This work was supported by National Institutes of Health grant CA-28735 to M. Bernfield. S. Saunders is a predoctoral trainee supported by Public Health Service grant 5T 32 CA 09302 from the National Cancer Institute.

Received for publication 12 November 1988.

References

Allen, G. 1981. Sequencing of proteins and peptides. *In* Laboratory Techniques in Biochemistry and Molecular Biology. T. S. Work and R. H. Burdon, editors. Elsevier Science Publishing Co. Inc., New York. 35–36.

Argraves, W. S., S. Suzuki, H. Arai, K. Thompson, M. D. Pierschbacher, and E. Ruoslahti. 1987. Amino acid sequence of the human fibronectin receptor. *J. Cell Biol.* 105:1183–1190.

Aziz, N., and H. N. Munro. 1987. Iron regulates ferritin mRNA translation through a segment of its 5′ untranslated region. *Proc. Natl. Acad. Sci. USA.* 84:8478–8482.

Bernfield, M., A. Rapraeger, M. Jalkanen, and S. D. Banerjee. 1985. Matrix interactions in epithelial morphogenesis involve a cell surface proteoglycan. *In* Basement Membranes. S. Shibata, editors. Elsevier Science Publishing Co. Inc., New York. 343–352.

Bourdon, M. A., T. Krusius, S. Campbell, N. B. Schwartz, and E. Ruoslahti. 1987. Identification and synthesis of a recognition signal for the attachment of glycosaminoglycans to proteins. *Proc. Natl. Acad. Sci. USA.* 84:3194–3198.

Bowness, J. M. 1987. Cartilage fucoproteins with sites for cross-linking by transglutaminase. *Biochem. Cell Biol.* 65:280–285.

Buck, C. A., and A. F. Horwitz. 1987. Cell surface receptors for extracellular matrix molecules. *Annu. Rev. Cell Biol.* 3:179–205.

Casey, J. L., M. W. Hentze, D. M. Koeller, S. W. Caughman, T. A. Rouault, R. D. Klausner, and J. B. Harford. 1988. Iron-responsive elements: regulatory RNA sequences that control mRNA levels and translation. *Science (Wash. DC).* 240:924–928.

Chen, W. T., T. Hasegawa, C. Hasegawa, C. Weinstock, and K. M. Yamada. 1985. Development of cell surface linkage complexes in cultivated fibroblasts. *J. Cell Biol.* 100:1103–1114.

Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry.* 18:5294–5299.

Church, G. M., and W. Gilbert. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci. USA.* 81:1991-1995.

Damsky, C. M., K. A. Knudsen, D. Bradley, C. A. Buck, and A. F. Horwitz. 1985. Distribution of the CSAT cell-matrix antigen on myogenic and fibroblastic cells in culture. *J. Cell Biol.* 100:1528-1539.

Danielsen, M., J. P. Northrop, and G. M. Ringold. 1986. The mouse glucocorticoid receptor: mapping of functional domains by cloning, sequencing and expression of wild-type and mutant receptor proteins. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:2513-2522.

David, G., and M. Bernfield. 1979. Collagen reduces glycosaminoglycan degradation by cultured mammary epithelial cells: possible mechanism for basal lamina formation. *Proc. Natl. Acad. Sci. USA.* 76:786-790.

Doolittle, R. F. 1986. Of URFS and ORFS: A Primer on How to Analyze Derived Amino Acid Sequences. University Science Books, Mill Valley, CA. 85.

Dorfman, A. 1981. Proteoglycan biosynthesis. *In* Cell Biology of Extracellular Matrix. E. D. Hay, editor. Plenum Publishing Corp., New York. 115-138.

Ebina, Y., L. Ellis, K. Jarnagin, M. Edery, L. Graf, E. Clauser, J. Ou, F. Masiarz, Y. W. Kan, I. D. Goldfine, R. A. Roth, and W. J. Rutter. 1985. The human insulin receptor cDNA: the structural basis for hormone-activated transmembrane signalling. *Cell.* 40:747-758.

Farquhar, M. G. 1985. Progress in unraveling pathways of Golgi traffic. *Annu. Rev. Cell Biol.* 1:447-488.

Feinberg, A. P., and B. Vogelstein. 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Addendum. *Anal. Biochem.* 137:266-267.

Ferguson, M. A. J., and A. F. Williams. 1988. Cell-surface anchoring of proteins via glycosyl-phosphatidylinositol structures. *Annu. Rev. Biochem.* 57:285-320.

Fransson, L. 1987. Structure and functions of cell-associated proteoglycans. *Trends Biochem. Sci.* 12:406-411.

Franzblau, C., and B. Faris. 1982. Biosynthesis of insoluble elastin in cell and organ cultures. *Methods Enzymol.* 82:615-637.

Geetha-Habib, M., S. C. Campbell, and N. B. Schwartz. 1984. Subcellular localization of the synthesis and glycosylation of chondroitin sulfate proteoglycan core protein. *J. Biol. Chem.* 259:7300-7310.

Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene (Amst).* 25:263-269.

Guest, J. R., H. M. Lewis, L. D. Graham, L. C. Packman, and R. N. Perham. 1985. Genetic reconstruction and functional analysis of the repeating lipoyl domains in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli. J. Mol. Biol.* 185:743-754.

Hayashi, K., M. Hayashi, M. Jalkanen, J. H. Firestone, R. L. Trelstad, and M. Bernfield. 1987. Immunocytochemistry of cell surface heparan sulfate proteoglycan in mouse tissues. A light and electron microscopic study. *J. Histochem. Cytochem.* 35:1079-1088.

Hayashi, K., M. Hayashi, E. Boutin, G. R. Cunha, M. Bernfield, and R. L. Trelstad. 1988. Hormonal modification of epithelial differentiation and expression of cell surface heparan sulfate proteoglycan in the mouse vaginal epithelium. *Lab. Invest.* 58:68-76.

Horwitz, A., K. Duggan, C. Buck, M. C. Berkerle, and K. Burridge. 1986. Interaction of plasma membrane fibronectin receptor with talin—a transmembrane linkage. *Nature (Lond.).* 320:531-533.

Huber, S., K. H. Winterhalter, and L. Vaughan. 1988. Isolation and sequence analysis of the glycosaminoglycan attachment site of type IX collagen. *J. Biol. Chem.* 263:752-756.

Hunter, T., and J. A. Cooper. 1985. Protein-tyrosine kinases. *Annu. Rev. Biochem.* 54:897-930.

Huynh, T. V., R. A. Young, and R. W. Davis. 1985. Construction and screening cDNA libraries in l-gt 10 and l-gt 11. *In* DNA Cloning: A Practical Approach. Vol. 1. D. M. Glover, editor. 49-78.

Hynes, R. O. 1987. Integrins: a family of cell surface receptors. *Cell.* 48: 549-554.

Ishihara, M., N. S. Fedarko, and H. E. Conrad. 1987. Involvement of phosphatidylinositol and insulin in the coordinate regulation of proteoheparan sulfate metabolism and hepatocyte growth. *J. Biol. Chem.* 262:4708-4716.

Izzard, C. S., R. Radinsky, and L. A. Culp. 1986. Substratum contacts and cytoskeletal reorganization of Balb/c 3T3 cells on a cell-binding fragment and heparin binding fragments of plasma fibronectin. *Exp. Cell Res.* 165:320-336.

Jalkanen, M., A. Rapraeger, S. Saunders, and M. Bernfield. 1987. Cell surface proteoglycan of mouse mammary epithelial cells is shed by cleavage of its matrix-binding ectodomain from its membrane-associated domain. *J. Cell Biol.* 105:3087-3096.

Jalkanen, M., A. Rapraeger, and M. Bernfield. 1988. Mouse mammary epithelial cells produce basement membrane and cell surface heparan sulfate proteoglycans containing distinct core proteins. *J. Cell Biol.* 106:953-962.

Koda, J. E., A. Rapraeger, and M. Bernfield. 1985. Heparan sulfate proteoglycans from mouse mammary epithelial cells: cell surface proteoglycan as a receptor for interstitial collagens. *J. Biol. Chem.* 260:8157-8162.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.

McCarthy, J. B., M. K. Chelberg, D. J. Mickelson, and L. T. Furcht. 1988. Localization and chemical synthesis of fibronectin peptides with melanoma adhesion and heparin binding activities. *Biochemistry.* 27:1380-1388.

Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* 12:7035-7056.

Messing, J. 1983. New M13 vectors for cloning. *Methods Enzymol.* 101:20-78.

Mort, A. J., and D. T. A. Lamport. 1977. Anhydrous hydrogen fluoride deglycosylates glycoproteins. *Anal. Biochem.* 82:289-309.

Osterlund, E., I. Eronen, K. Osterlund, and M. Vuento. 1985. Secondary structure of human plasma fibronectin: conformational changes induced by calf alveolar heparan sulfate. *Biochemistry.* 24:2661-2667.

Pierschbacher, M. D., and E. Ruoslahti. 1984. Cell attachment activity of fibronectin can be duplicated by small fragments of the molecule. *Nature (Lond.).* 309:30-33.

Rapraeger, A., and M. Bernfield. 1982. An integral membrane proteoglycan is capable of binding components of the cytoskeleton and the extracellular matrix. *In* Extracellular Matrix. S. Hawkes and J. Wang, editors. Academic Press Inc., New York. 265-269.

Rapraeger, A., and M. Bernfield. 1985. Cell surface proteoglycan of mammary epithelial cells: protease releases a heparan sulfate-rich ectodomain from a putative membrane-anchoring domain. *J. Biol. Chem.* 260:4103-4109.

Rapraeger, A., M. Jalkanen, E. Endo, J. Koda, and M. Bernfield. 1985. The cell surface proteoglycan from mouse mammary epithelial cells bears chondroitin sulfate and heparan sulfate glycosaminoglycans. *J. Biol. Chem.* 260:11046-11052.

Rapraeger, A., M. Jalkanen, and M. Bernfield. 1986. Cell surface proteoglycan associates with the cytoskeleton at the basolateral cell surface of mouse mammary epithelial cells. *J. Cell Biol.* 103:2683-2696.

Rasmussen, S., and A. Rapraeger. 1988. Altered structure of the hybrid cell surface proteoglycan of mammary epithelial cells in response to transforming growth factor-β. *J. Cell Biol.* 107:1959-1968.

Ratner, L., B. Thielan, and T. Collins. 1987. Sequences of the 5′ portion of the human c-*sis* gene: characterization of the transcriptional promoter and regulation of expression of the protein product by 5′ untranslated mRNA sequences. *Nucleic Acids Res.* 15:6017-6036.

Rodén, L. 1980. Structure and metabolism of connective tissue proteoglycans. *In* The Biochemistry of Glycoproteins and Proteoglycans. W. J. Lennarz, editor. Plenum Publishing Corp., New York. 267-371.

Ruoslahti, E., and M. D. Pierschbacher. 1987. New perspectives in cell adhesion: RGD and integrins. *Science (Wash. DC).* 238:491-497.

Sabatini, D. D., G. Kreibich, T. Morimoto, and M. Adesnik. 1982. Mechanisms for the incorporation of proteins in membranes and organelles. *J. Cell Biol.* 92:1-22.

Saksela, O., D. Moscatelli, A. Sommer, and D. B. Rifkin. 1988. Endothelial cell-derived heparan sulfate binds basic fibroblast growth factor and protects it from proteolytic degradation. *J. Cell Biol.* 107:743-751.

Sanderson, R. D., and M. Bernfield. 1988. Molecular polymorphism of a cell surface proteoglycan: distinct structure on simple and stratified epithelia. *Proc. Natl. Acad. Sci. USA.* 85:9562-9566.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463-5467.

Saunders, S., and M. Bernfield. 1988. Cell surface proteoglycan binds mouse mammary epithelial cells to fibronectin and behaves as a receptor for interstitial matrix. *J. Cell Biol.* 106:423-430.

Segarini, P. R., and S. M. Seyedin. 1988. The high molecular weight receptor to transforming growth factor-β contains glycosaminoglycan chains. *J. Biol. Chem.* 263:8366-8370.

Struhl, K. 1985. A rapid method for creating recombinant DNA molecules. *Biotechniques.* 3:452-453.

Sun, X., D. F. Mosher, and A. Rapraeger. 1989. Heparan sulfate-mediated binding of epithelial cell surface proteoglycan to thrombospondin. *J. Biol. Chem.* 264:2885-2889.

Tamkun, J. W., D. W. DeSimone, D. Fonda, R. S. Patel, C. Buck, A. F. Horwitz, and R. O. Hynes. 1986. Structure of integrin, a glycoprotein involved in the transmembrane linkage between fibronectin and actin. *Cell.* 46:271-282.

Thesleff, I., M. Jalkanen, S. Vainio, and M. Bernfield. 1988. Cell surface proteoglycan expression correlates with epithelial-mesenchymal interaction during tooth morphogenesis. *Dev. Biol.* In press.

Weitzhandler, M., H. B. Streeter, W. J. Henzel, and M. Bernfield. 1988. The cell surface proteoglycan of mouse mammary epithelial cells: the extracellular domain contains the N-terminus and a peptide sequence present in a conditioned medium proteoglycan. *J. Biol. Chem.* 263:6949-6952.

Woods, A., J. R. Couchman, S. Johansson, and M. Hook. 1986. Adhesion and cytoskeletal organization of fibroblasts in response to fibronectin fragments. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:665-670.

Young, R. A., and R. W. Davis. 1983. Yeast RNA polymerase II genes: isolation with antibody probes. *Science (Wash. DC).* 222:778-782.