Human Basement Membrane Heparan Sulfate Proteoglycan Core Protein: A 467-kD Protein Containing Multiple Domains Resembling Elements of the Low Density Lipoprotein Receptor, Laminin, Neural Cell Adhesion Molecules, and Epidermal Growth Factor

Pekka Kallunki and Karl Tryggvason

Biocenter and Department of Biochemistry, University of Oulu, SF-90570 Oulu, Finland

Abstract. The primary structure of the large human basement membrane heparan sulfate proteoglycan (HSPG) core protein was determined from cDNA clones. The cDNA sequence codes for a 467-kD protein with a 21-residue signal peptide. Analysis of the amino acid sequence showed that the protein consists of five domains. The amino-terminal domain I contains three putative heparan sulfate attachment sites; domain II has four LDL receptor-like repeats; domain III contains repeats similar to those in the short arms of laminin; domain IV has lg-like repeats resembling those in neural cell adhesion molecules; and domain V contains sequences resembling repeats in the G domain of the laminin A chain and repeats in the EGF. The domain structure of the human basement membrane HSPG core protein suggests that this mosaic protein has evolved through shuffling of at least four different functional elements previously identified in other proteins and through duplication of these elements to form the functional domains. Comparison of the human amino acid sequence with a partial amino acid sequence from the corresponding mouse protein (Noonan, D. M., E. A. Horigan, S. R. Ledbetter, G. Vogeli, M. Sasaki, Y. Yamada, and J. R. Hassell. 1988. *J. Biol. Chem.* 263:16379–16387) shows a major difference between the species in domain IV, which contains the Ig repeats: seven additional repeats are found in the human protein inserted in the middle of the second repeat in the mouse sequence. This suggests either alternative splicing or a very recent duplication event in evolution. The multidomain structure of the basement membrane HSPG implies a versatile role for this protein. The heparan sulfate chains presumably participate in the selective permeability of basement membranes and, additionally, the core protein may be involved in a number of biological functions such as cell binding, LDL-metabolism, basement membrane assembly, calcium binding, and growth- and neurite-promoting activities.

BASEMENT membranes are thin sheetlike structures that form a highly specialized part of the extracellular matrix located at the immediate proximity to the surface of organ cells such as epithelial, endothelial, and mesenchymal cells where they meet with the underlying interstitial connective tissue. The basement membranes serve a number of functions in vivo, e.g., during embryogenesis, cell differentiation, and cell–cell and cell–matrix interactions. They are also considered important for the correct remodeling and regeneration processes of tissues, and for the filtration of macromolecules by the renal glomerular basement membrane (for reviews see references 19, 64). The basement membranes are composed of several distinct proteins, some of which are specific for these structures such as type IV collagen, laminin, heparan sulfate proteoglycan (HSPG),[1] and entactin/nidogen. Type IV collagen and laminin are complex trimeric proteins which can be present in several isoforms, each protein having at least five different kinds of subunits encoded by distinct genes (6, 18, 27, 29, 42, 43, 45, 46, 52, 53, 58, 59, 64).

HSPGs are integral components of basement membranes. An HSPG containing a core protein with an estimated size between 350 and 500 kD has been isolated from several sources including the basement membrane matrix forming mouse EHS tumor (25), endothelial cells (56), epithelial cells (41), human colon carcinoma cells (30), and fibroblasts (26). Both polyclonal and monoclonal antibodies specific for the large basement membrane HSPG core protein have been shown to stain basement membranes in a variety of tissues (11, 33). Basement membranes have also been reported to contain HSPGs with smaller core proteins (25, 62). Their identity and possible interrelationship is, however, still largely unclear. Some of these small core proteins share antigenic determinants with the large HSPG core protein and they

1. *Abbreviations used in this paper*: EHS, Engelbreth-Holm-Swarm; HS, heparan sulfate; HSPG, heparan sulfate proteoglycan; LDL, low density lipoprotein; N-CAM, neural cell adhesion molecule.

have been suggested to be proteolytic fragments of the latter (34, 36).

Based on rotary shadowing EM images and protein analysis, the large HSPG has an ~60–80-nm-long elongated structure with several globular domains and three heparan sulfate (HS) side chains at one end (37, 41, 50). cDNA clones encoding small parts of the HSPG core protein from mouse tumor (47), human colon carcinoma, and HT1080 fibrosarcoma cells (16, 31) have demonstrated that the core protein contains regions with homology to the rodlike and globular domains of the short arms of the laminin chains. Additionally, the core protein was shown to contain internal repeats similar to Ig repeats in the neural cell adhesion molecule N-CAM (12). The human basement membrane HSPG core protein has been shown to be encoded by a single gene, *HSPG2*, located on chromosome 1p36.1 → p35 (31). In addition to basement membrane HSPGs, distinct cell surface HSPGs have been identified on mesenchymal, epithelial, and neural cells. Among those are syndecan with a 33-kD core protein (60) and fibroblast HSPGs with 48- and 64-kD core proteins (13, 40). These cell surface HSPGs are all distinct gene products and have not been shown to have structural homology to the large basement membrane HSPG.

Several functions including role in cell proliferation and morphogenesis have been suggested for HSPG and, importantly, they may influence the permeability of macromolecules in capillaries and renal glomeruli by virtue of their strong anionic charge (19, 32, 64). It has been postulated that the loss of HS is responsible for increased permeability of the glomerular basement membrane in proteinuria observed in diabetes, and the nephrotic syndrome (49, 65). Previous reports indicate that the HS side chains form a distinct anionic layer, e.g., in the glomerular basement membrane where they have been located to the *lamina rara interna* and *lamina rara externa* (19, 32). Antibodies generated against the core protein have localized it to all layers of the basement membranes (26, 33, 56). These data suggest that this large elongated component may be oriented in the basement membrane in such a fashion that it penetrates all layers with the end, where the HS side chains are attached on one side of the basement membranes (26).

In the present study we describe the complete cDNA-derived primary structure of the large basement membrane HSPG core protein. This protein, which has a molecular mass of 467 kD, has a complex multiple domain structure with homology to the low density lipoprotein (LDL) receptor, laminin, neural cell adhesion molecules, and EGF with three putative attachment sites for HS at the amino-terminal end. The primary structure indicates a variety of functions for this basement membrane-specific component.

Materials and Methods

RNA Isolations and Northern Analysis

Poly(A) RNA was isolated from confluent cultures of human HT1080

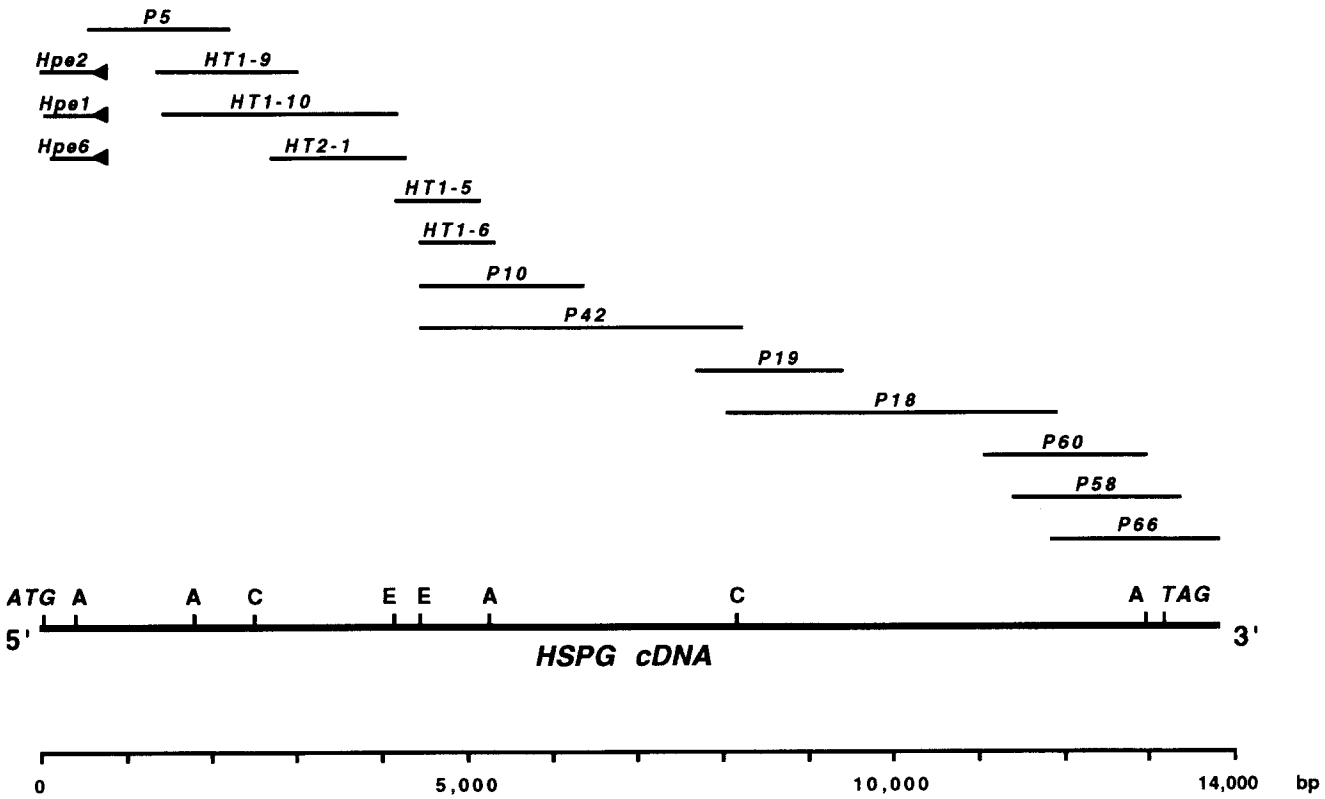


Figure 1. Scheme of 16 cDNA clones encoding the 467-kD human basement membrane HSPG core protein. cDNA clones with an arrow tail represent clones made by primer extension. Location of the ATG translation initiation signal and the 3'-end TAG translation stop codon are shown. Restriction enzyme sites for AccI (*A*), ClaI (*C*), and EcoRI (*E*) are indicated. Scale in base pairs is shown at the bottom.

fibrosarcoma cells as described previously (31). This RNA was used for Northern analysis and for the construction of the cDNA libraries made in this study. For Northern analysis, ~5 μg of poly(A) RNA was electrophoresed on a 0.7% agarose gel in the presence of formaldehyde (39). The RNA was then transferred to nitrocellulose filters, which were hybridized with nick-translated HSPG core protein coding cDNA probes followed by autoradiography.

Construction and Screening of cDNA Libraries

Three separate cDNA libraries were made from the HT1080 poly(A) RNA. Two libraries in the λgt11 vector were made by Clontech Laboratories Inc. (Palo Alto, CA); a specific primed library and an oligo(dT), and a random primed library. The specific library was primed with two oligonucleotides made based on short amino acid sequences from the HSPG core protein as described previously (31). A third primer-specific library in the λgt10 vector was made in our laboratory to obtain the 5′-end clones using a primer H57 (5′ CAG GAC TGG CTC CTC ACA ATT GAG 3′) complementary to bases 734–757 (see Fig. 3). Additionally, a commercial oligo(dT)-primed HT1080 cDNA library available from Clontech was screened to obtain 3′-end clones.

The first clone for the human basement membrane HSPG core protein HT2-1, isolated from the specific library made by Clontech, has been characterized previously (31). The HT2-1 insert was used to screen the oligo(dT) and random-primed library and several clones were isolated. Nick-translated fragments from the 5′-ends and 3′-ends of these clones were used in further screening of the libraries.

Characterization of the Human Basement Membrane HSPG cDNA Clones

All cDNA clones were subcloned into M13 and pUC vectors. The nucleotide sequence was determined from both strands by the dideoxy chain termination method (57) using sequenase (United States Biochemical Corp., Cleveland, OH) and universal or sequence-specific oligonucleotide primers. In some cases, the sequences were obtained from nested deletion of large cDNA clones, made using an ExoIII-S1 nested deletion kit (Pharmacia Fine Chemicals, Piscataway, NJ). Most of the sequences were checked later using Taq polymerase (AmpliTAQ, Perkin-Elmer Cetus Instrs., Norwalk, CT), and fluorescent primers or "dyedeoxy-nucleotides," and an automatic DNA sequencer (Applied Biosystems, Inc., Foster City, CA, or Pharmacia Fine Chemicals).

Computer Analysis of Sequences

The nucleotide sequences were initially analyzed using the Microgenie program package (Beckman Instruments, Inc., Berkeley CA). Multiple alignments were performed with a program MULTALIN (INRA, France) (39). Databank searches were carried out using the program FASTA in the GCG Package (Genetics Computer Group, Madison, WI) (14, 51).

Results

Isolation and Characterization of cDNA Clones

A total of 16 cDNA clones (Fig. 1) were isolated and characterized to provide the entire primary structure of the large HSPG core protein. First, screening of the random-primed and oligo(dT)-primed HT1080 cDNA library with the previously described human HT2-1 HSPG cDNA clone (31) yielded the 5′-end and 3′-end overlapping clones HT1-10, HT1-9, HT1-5, and HT1-6 which were all sequenced. 5′-end and 3′-end fragments of the latter were then used to isolate the 5′-end clone P5 and the 3′-end clone P10. Screening of the same library with the 3′-end of P10 resulted in the isolation of P42, P19, and P18. No further clones reaching upstream of P5 or downstream of P18 could be isolated from this cDNA library. To obtain clones spanning the 5′-end of the mRNA, a primer-extended cDNA library was made from HT1080 cell poly(A) RNA using a synthetic oligonucleotide with sequence from the 5′-end of the clone P5. Screening of this library yielded the 5′-end clones Hpe1, Hpe2, and Hpe6.

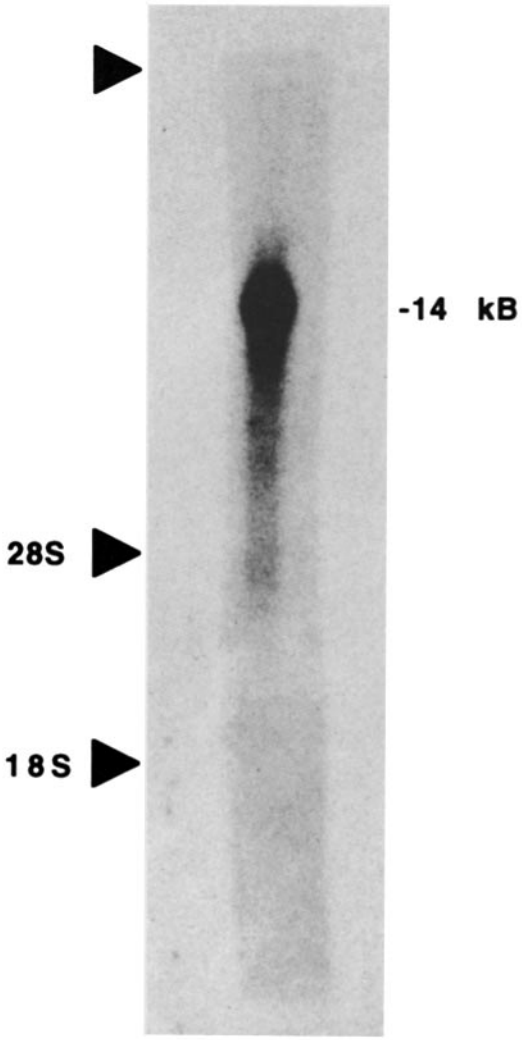


Figure 2. Northern blot analysis of poly(A)-enriched RNA. 5 μg of poly(A)-enriched RNA from human HT1080 fibrosarcoma cells was electrophoresed on 0.7% agarose gels and transferred to a nitrocellulose filter as described in Materials and Methods. Hybridization was carried out using the ^{32}P-labeled human HSPG cDNA HT2-1 (Fig. 1) as probe. The size of the mRNA was estimated by using the sizes of mRNAs for the human laminin A (9 kb), B1 (6 kb), and B2 chains (5.5 and 7.5 kb), and ribosomal 28S and 18S as references.

The 3′-end cDNA clones P58, P60, and P66 could be isolated from an oligo(dT)-primed HT1080 cell cDNA library purchased from Clontech (catalog No. HL10486). The 16 overlapping clones span a total of ~14 kb (Fig. 1) and they were also shown to hybridize to an mRNA with the same size (Fig. 2). Partial restriction map based on the nucleotide sequence (Fig. 3) is shown in Fig. 1.

Nucleotide and Amino Acid Sequences

The complete nucleotide sequence together with the predicted amino acid sequence is shown in Fig. 3. The entire sequence contained in the 16 overlapping clones is 13,793 nucleotides with 40 nucleotides of a putative 5′-end untranslated sequence, 13,179 nucleotides of an open reading frame, and 574 nucleotides of a 3′-end untranslated sequence. The open reading frame encodes 4,393 amino acid residues starting

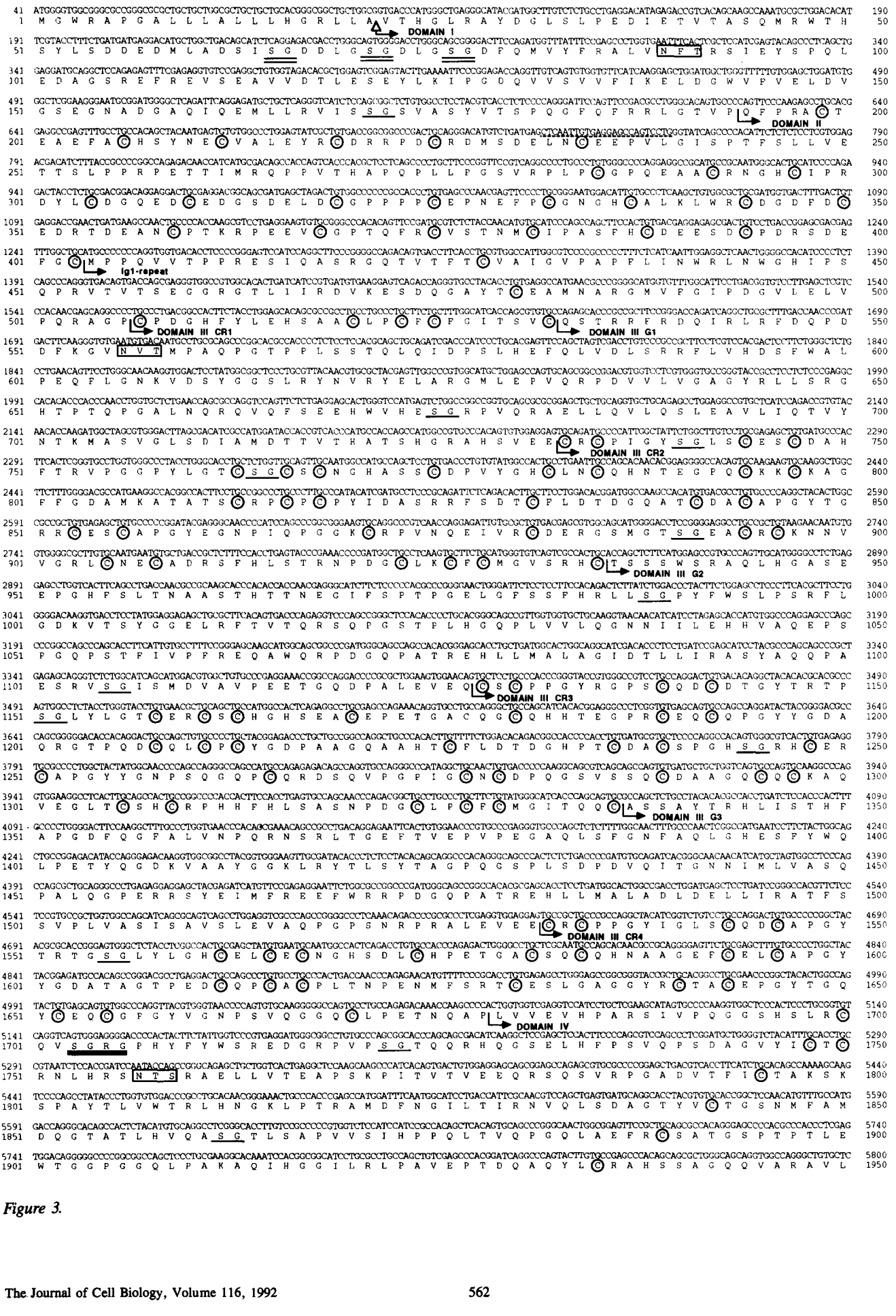
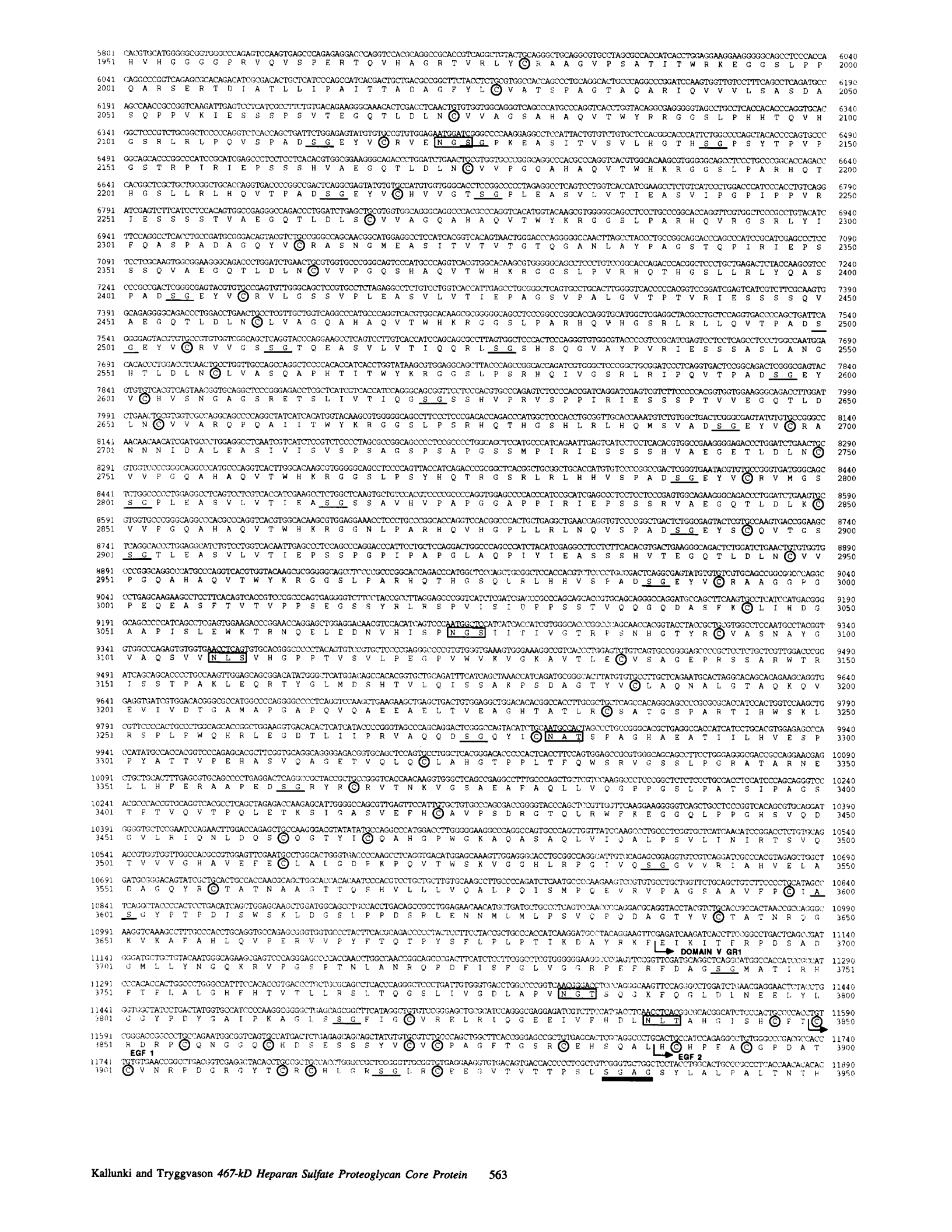

DOMAIN I
DOMAIN II
fg1-repeat
DOMAIN III CR1
DOMAIN III G1
DOMAIN III CR2
DOMAIN III G2
DOMAIN III CR3
DOMAIN III G3
DOMAIN III CR4
DOMAIN IV

Figure 3.


DOMAIN V GR1
EGF 1
EGF 2

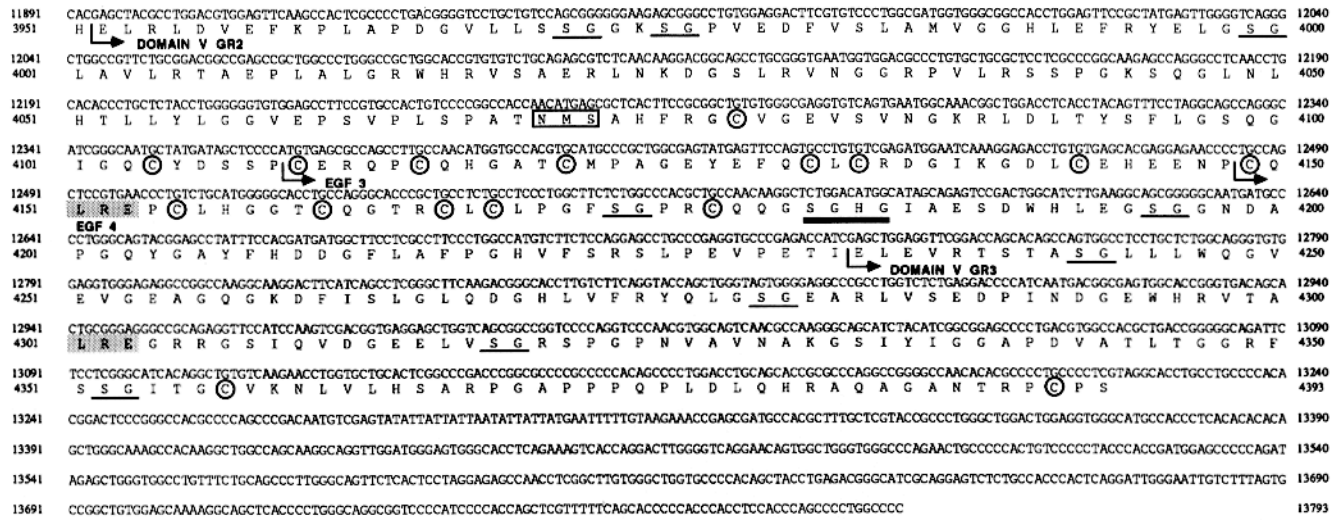


Figure 3. Nucleotide sequence of the human 467-kD basement membrane HSPG core protein cDNA clones and the inferred amino acid sequence. (*First line*) Nucleotide sequence of cDNA clones. (*Second line*) Nucleotide-derived amino acid sequence. The putative signal peptidase cleavage site is indicated by an open arrow. Borders of structural domains are indicated by bent arrows. Ser-Gly doublets (*SG*) which are potential attachment sites for glycosaminoglycan side chains are underlined. The three amino-terminal end Ser-Gly sequences which are the most likely attachment sites for the large HS chains (see text) are highlighted by double underlines. Ser-Gly-X-Gly sequences are marked by a bold underline. Cysteine residues are circled. Potential attachment sites (Asn-X-Ser or Asn-X-Thr) for oligosaccharides are shown by rectangles. Two Leu-Arg-Glu sequences that may mediate motor neuron attachment are shown by shaded rectangles. The sequence of the oligonucleotide used for preparation of primer extended cDNA clones is indicated by a dotted line. These sequence data are available from EMBL/GenBank/DDBJ under accession number X62515.

with the ATG codon for methionine at nucleotide 41, and ending at the translation stop codon TAG at nucleotide 13,219. The predicted translation initiation site obeys the vertebrate 50/75 Consensus Rule with C at −1 and −2 and G at −3 and +4 (8). Although the first 40 nucleotides encode a sequence in-frame with the presumed methionine initiator, it is likely that the consensus sequences described above determine the initiation site for translation. This assumption is supported by the fact that numerous 5′-end cDNA clones generated by primer extension did not reach further upstream.

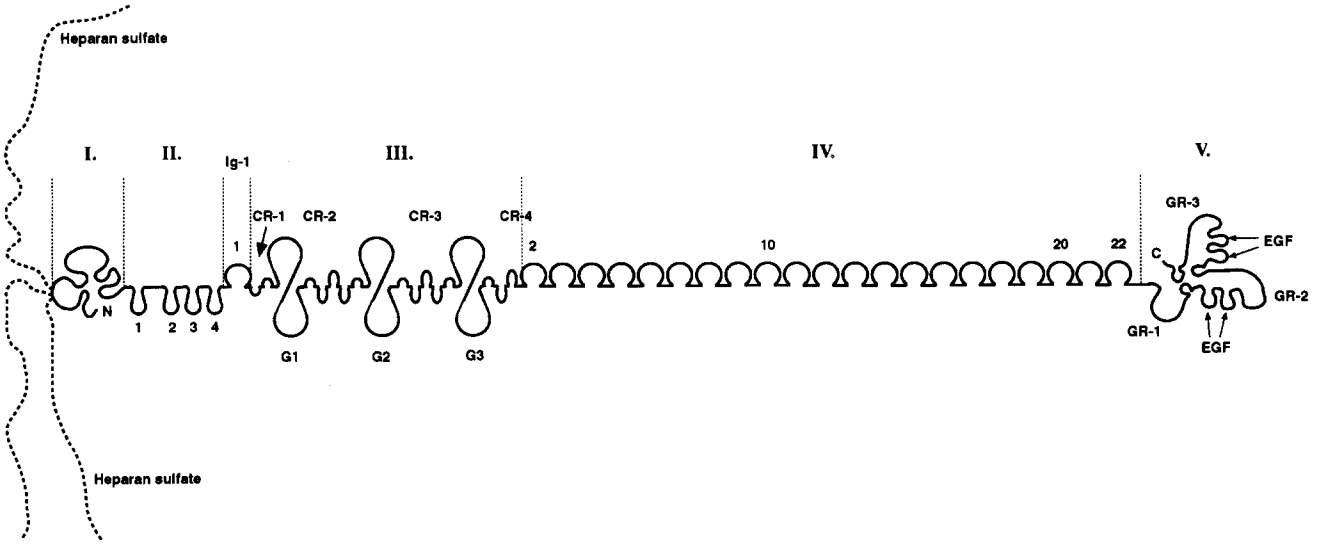


Figure 4. Schematic illustration of the multidomain structure of the human basement membrane HSPG based on EM, analysis of the amino acid sequence, and comparison with structurally related proteins. Five relatively well-defined domains can be observed: domain I, a globular amino-terminal domain containing three SG-linked HS side chains that have been visualized by electron microscopy (41, 50); domain II, containing four copies of internal repeats similar to the ligand-binding domain of the LDL receptor; a short region between domains II and III with one copy of a globular structure similar to the disulfide-linked internal repeats of immunoglobulin; domain III, with homology to the short arm of laminin chains, including four presumably rodlike cysteine-rich subdomains (CR-1 to CR-4), and three globular subdomains (G1, G2, and G3); domain IV with 21 copies of internal Ig-like repeats; and domain V with considerable similarity to the carboxyl-terminal end globular domain of the laminin A chain. This domain has three distinct internal globular repeats (GR-1, GR-2, and GR-3) separated by two doublets of internal repeats similar to parts of EGF. The molecular model is not drawn to scale.

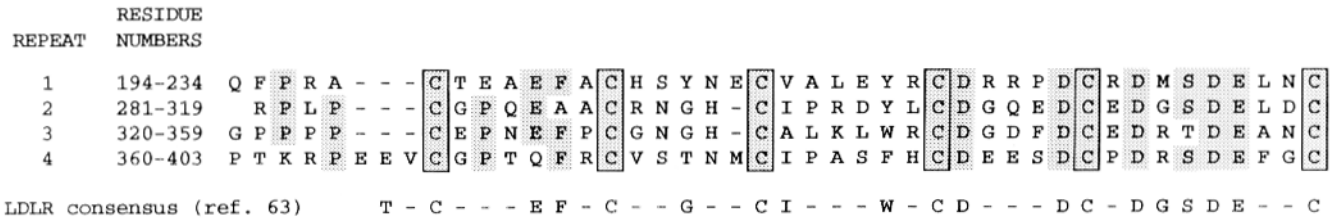


Figure 5. Alignment of amino acid sequences of the four cysteine-rich repeats similar to the ligand-binding domain of the LDL receptor. Residues present in >50% of the repeats are highlighted by shade. All cysteines shown in closed shaded boxes are conserved as in the LDL receptor (63). As found for the LDL receptor each repeat sequence has a net anionic charge of the residues of the carboxyl-terminal end.

Also, the amino acid sequence starts with a 21-residue signal peptide–like segment containing a typical hydrophobic leucine-rich core. The signal peptidase cleavage site was predicted using the computer program SIGSEQ (The Rockefeller University, New York) based on the method of von Hejne (66). The highest probability for the cleavage site was obtained between residues Alanine21 and Valine22. The core protein proper contains 4,372 residues with a calculated molecular weight of 466,876.

The entire sequence contains a total of 52 Ser-Gly amino acid doublets which are putative glycosaminoglycan attachment sites. These sequences are scattered throughout the polypeptide chain (Fig. 3). Three of those located at the amino terminus conform at least partly with the consensus sequence previously identified in proteoglycans such as versican, syndecan, and glypican (13, 60, 69). Three others are similar to the second consensus sequence identified in many proteoglycans (5). Potential attachment sites for Asn-X-Ser/Thr-linked oligosaccharides are present at 10 locations, also distributed in a random manner (Fig. 3). No Arg-Gly-Asp triplets which can confer cell binding (55) were found in the HSPG core protein. The Leu-Arg-Glu sequence which has been reported to mediate motor neuron attachment (28) was found at two locations in the carboxyl-terminal globular domain.

Multidomain Structure

Homology searches of the data banks together with a systemic search for internal repeats demonstrated that the protein has five distinct domains: domain I, a globular amino-terminal domain with putative HS attachment sites; domain II, with four copies of sequences similar to the LDL receptor ligand binding repeats; a central domain III, resembling the short arm of laminin chains with three globules and four regions of cysteine-rich repeats; domain IV, containing Ig repeats; and domain V, a large globular carboxyl-terminal domain resembling a part of the G domain of the laminin A chain (Fig. 4). Domains II and III were separated by one copy of an Ig repeat–like sequence. This domain structure is in accordance with EM and protein analyses, which indicate that the core protein folds into ~60–80-nm-long linear array of globular domains with extensive secondary structure (37, 50).

Domain I: Putative HS Attachment Domain (Residues 22–193)

At the amino-terminal end of the core protein proper there is a 172-residue sequence which has no homologous counterparts in the data banks. The amino acid sequence contains three Ser-Gly-Asp sequences which are potential attachment sites for HS side chains. This sequence is similar to the putative HS attachment sequences in the cell surface HSPGs syndecan (60), and glypican (13), and in chondroitin sulfate proteoglycan versican (69).

Domain II: LDL Receptor–like Domain (Residues 194–403)

A second distinct domain of 210 residues contains 4 copies of ~40-residue-long repeats that have high sequence identity with the repeats in the ligand-binding domains of the LDL receptor (63). As seen in Fig. 5, these repeats are very hydrophilic. The first repeat is separated from the other three by ~40 residues of a unique sequence. In addition to the LDL receptor, these repeats are found in the LDL receptor–related protein and in the terminal complement components such as C9 (15). Each repeat contains a consensus sequence (Fig. 5) which has six conserved cysteines and negatively charged acidic residues in the hydrophilic part characteristic for the ligand-binding regions of the LDL receptor (63).

Ig-like Repeat 1 (Ig-1) (Residues 404–506)

After the LDL receptor–like domain, there is one 103-residue-long segment which is homologous to the Ig-like repeats in the neural cell adhesion molecules. This repeat contains two conserved cysteines and it has, in general, considerable sequence similarity with the 21 individual Ig repeats in domain IV of the HSPG core protein (see below, and Figs. 4 and 7). The cysteines are predicted to form an intrachain disulfide bond and this region probably folds into a globular structure similar to Ig domains of the C-2 type (67).

Domain III: Laminin Short Arm–like Domain (Residues 507–1,678)

This 1,172-residue-long domain resembles the short arm of laminin chains (Fig. 4). It contains four subdomains (CR-1–CR-4) which consist of internal cysteine-rich repeats similar to those typically found in domains III and V of the short arms of the laminin A and B chains (18, 29, 42, 43, 46, 52, 53, 58, 59). These subdomains are separated by globular domains—G1, G2, and G3—which are homologous to domain IV in the short arms of the laminin chains. As in laminin A chain and laminin B2 chain, where the globular domain IV is apparently inserted in the middle of a cysteine repeat between the third and fourth cysteines, the G subdomains are in the middle of a cysteine repeat. The sequences

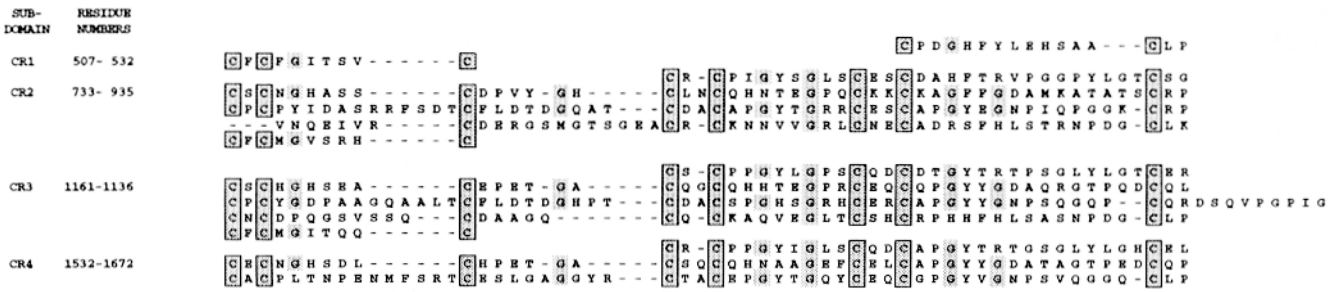


Figure 6. Alignment of the amino acid sequences of internal cysteine-rich repeats. The amino acids are aligned according to the sequence of similar internal repeats found in the laminin A and B chains. The HSPG contains a total of 15 repeats in four different subdomains: CR-1, CR-2, CR-3, and CR-4. Six of the repeats contain only about one-half of an actual repeat sequence. Amino acids present in >50% of the repeats are shaded. The cysteines shown in closed shaded boxes are always conserved as they are in laminin. Additionally, glycine residues that are thought to be located at turns in each loop-containing rodlike subdomains are usually conserved.

of the internal repeats of the four cysteine-rich subdomains are aligned in Fig. 6, in accordance with the sequence alignment of the repeats of laminin. According to this alignment, CR-1 contains two half repeats, CR-2 and CR-3 each contain two half and three complete repeats, and CR-4 has one half and two complete repeats. As can be seen in Fig. 6, each complete internal repeat has eight conserved cysteines and typically also several glycines that are believed to be located

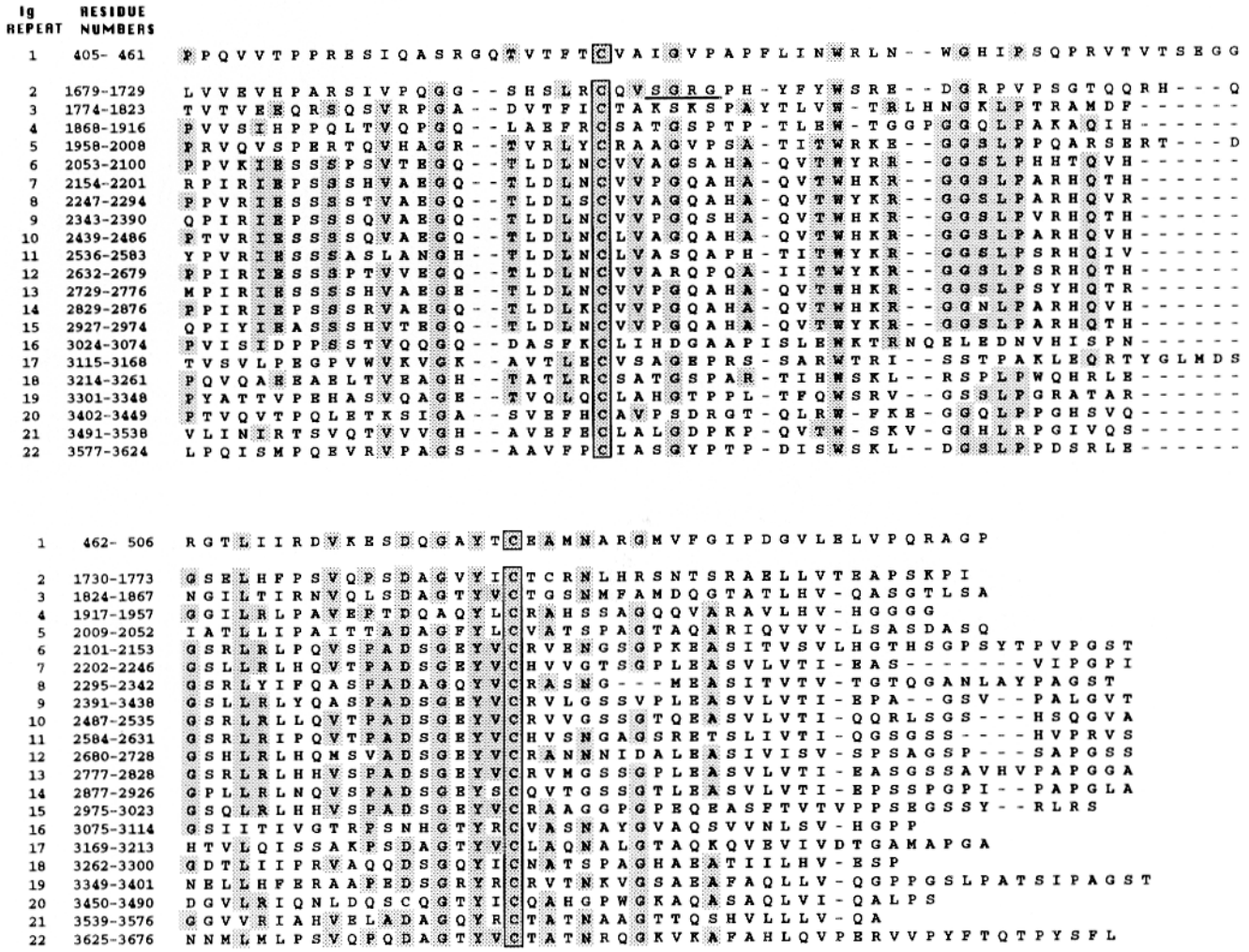


Figure 7. Sequence comparison of the 22 repeat motifs similar to the internal repeats of Igs. The first repeat (Ig-1) is located separately from the others which all are located consecutively following each other (see Fig. 4). The sequences were aligned using the MULTALIN program (Materials and Methods). The two cysteine residues which are always conserved are highlighted by shaded boxes and amino acids conserved in >50% of cases are shaded. It can be observed that, in addition to the two cysteines, a tryptophan residue is always conserved and similarly glycine residues believed to be located at turns of the tertiary structure are usually conserved.

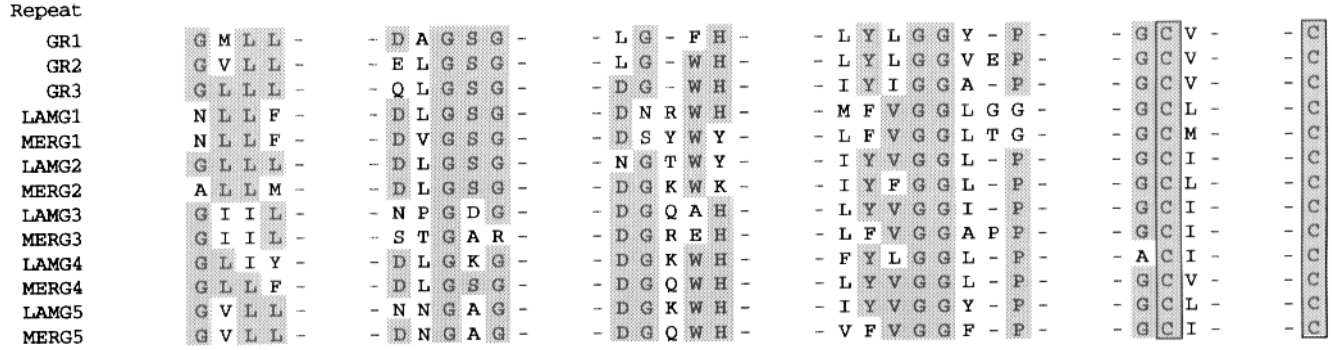


Figure 8. Comparison of conserved sequences of globular subdomains GR-1, GR-2, and GR-3 from domain V with those from globular subdomains of the human laminin A chain and merosin. The sequences were first aligned using the MULTALIN program and regions with substantial sequence identity shown. Residues conserved in ≥50% of locations are highlighted by a shaded pattern.

at the turning points of the small loops that presumably make up each rigid rodlike CR subdomain as described for laminin. The three globular domains G1 (residues 533–732), G2 (936–1,127), and G3 (1,337–1,531) between the cysteine-rich regions share between 30 and 40% sequence similarity and they bear a closest resemblance with the laminin A and B2 chain domain IV. This sequence similarity is ~25–31% and the domains are also almost identical in size. Based on the similarity to laminin these domains can be predicted to adopt a globular conformation similar to that found in the laminin chains. These domains probably form part of the necklace-like structure at basement membrane HSPG core protein identified in electron micrographs, characterized by globules that are connected by short rodlike segments (50). The overall structure of this domain in the HSPG core protein is most similar to the short arm of the A chain. The size of globular domain IV in laminin has been estimated to be ~4–5 nm in diameter and the cysteine repeats have been estimated to form ~2-nm-long elements (3). The size of the lamininlike domain in HSPG core protein can thus be calculated to be ~40 nm. This is comparable to the size of the short arm in laminin A chain, which is estimated to be ~45–50 nm.

Domain IV: Ig-like Domain (Residues 1,679–3,688)

Next to the laminin-like CR-4 subdomain towards the carboxyl terminus there is a well-defined domain of 2,010 residues that contains 21 consecutive copies of homologous repeats which are similar to the Ig repeats in N-CAM (12). As does the single Ig repeat located at the amino-terminal end of the core protein (above), all the 21 repeats are ~100 residues in length (Fig. 7). The repeats have a conserved sequence which includes two cysteine residues and the sequence around the second cysteine has the highest degree of conservation. As the single Ig repeat, these repeats are likely to fold into an Ig domain of C-2 type. The structure of this Ig-like domain is an ellipsoid with dimensions of ~4 × 2.5 × 2.5 nm and in N-CAM five such repeats form a linear, rodlike structure of ~17–18 nm (4). Therefore, the domain IV alone would be at least 60–80 nm in length if the repeats form a rodlike tertiary structure. The amino acid sequences of all the Ig-like repeats of the human HSPG core protein are aligned in Fig. 7 where residues with ≥50% conservation are shown in boxes. It can be noted that glycine residues are frequently conserved. In the middle of this domain there are repeats which have an extremely high sequence identity. A search was made in the sequence data banks for the conserved sequences of these repeats. However, this search did not reveal any protein sequences with substantial sequence identity. Only homology to proteins of the Ig superfamily was observed. Detailed comparison of the Ig repeats in HSPG with neural cell adhesion molecules shows that the highest degree of identity is ~25% over 428 amino acids to the TAG-1 (20), but the identity to contactin (54)—23% over 499 amino acids and L1 (44); 23% over 543 amino acids—is not much lower, with N-CAM (2) having ~19.5% identity over 437 amino acids. For the comparison, the PDGF receptor (68) has a 21% identity over 405 amino acids.

Domain V: Laminin A Chain–like Domain (Residues 3,689–4,393)

At the carboxyl-terminal end of the HSPG core protein there is a 705-residue-long domain that resembles the large carboxyl-terminal end globular G domain of the A chains of laminin (18, 46, 59). In the HSPG core protein this domain

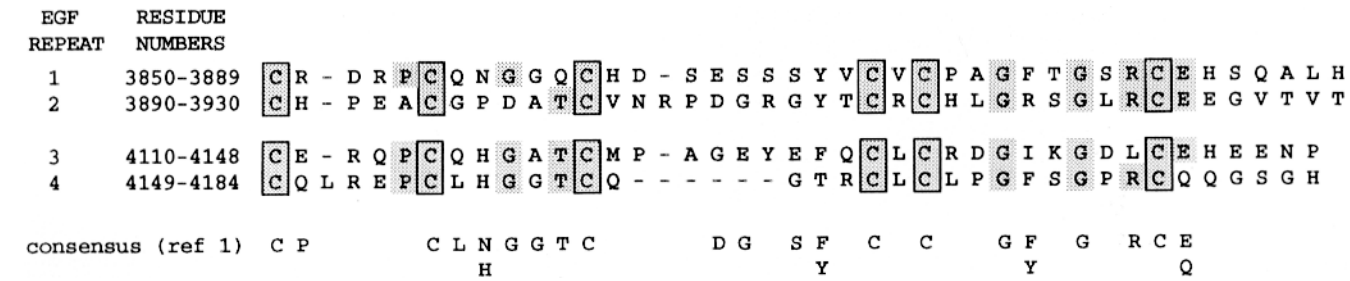


Figure 9. EGF motifs in human basement membrane HSPG core protein. Residues conserved >50% of locations are highlighted by shade. Cysteine residues are boxed. The consensus sequence derived from the EGF repeats (1) is shown below.

has three homologous, presumably globular, regions termed here GR-1, GR-2, and GR-3 (Fig. 4), which are similar to the five repeats in the laminin A chain G domain. Comparison of the amino acid sequence of GR-1, GR-2, and GR-3 revealed that their sequence identity is ~33%. The best alignment of these repeats with similar repeats in laminin and merosin gives an identity of ~30%. Each GR subdomain has two cysteines and several other amino acids conserved in the corresponding repeats of the laminin A chain and merosin as shown in Fig. 8. The sequences in the subdomains also share some homology with sex hormone binding globulin (21). These subdomains in the HSPG core protein most likely fold into globular structures similar to those in the G domain of laminin. The two conserved cysteines in the fifth subdomain of laminin have been shown to be linked (3) and this may be the situation also in the other subdomains. In laminin these globules are estimated to be ~3.5 nm in diameter (3).

The GR repeats in the HSPG core protein are each separated by two copies of sequences similar to internal repeats of EGF (1, 23). The sequences that are similar to repeats in the EGF contain each ~40 residues and have 6 conserved cysteines that form disulfide bonds (Fig. 9). Additionally, three glycine residues presumably located at turning points in the secondary structure are also conserved. The EGF-like regions are proposed to fold into a typical structure with three disulfide-linked loops (1).

Comparison of The Human Core Protein Sequence to Partial Sequences From Mouse

Two regions of the large mouse HSPG core protein amino acid sequence have already been determined from cDNA clones (47). One of the mouse sequences which corresponds to human residues 942–1,603 from domain II showed an overall sequence identity of 87% (data not shown). The other continuous 731-residue mouse sequence has the highest sequence identity with two separate regions of the Ig-like repeats of domain IV of the human protein, residues 1,873–1,999 and 2,674–3,280 which correspond to internal repeats Ig4-5 and Ig12-18, respectively. As discussed below, this difference could be due to alternative splicing of the primary transcripts or actual different sizes of the mouse and human proteins.

Discussion

The present study provides the first complete amino acid sequence of the large basement membrane HSPG core protein. The primary structure of the human protein elucidated here shows that this 467-kD core protein has a complex and interesting multidomain composition with domains and subdomains partially resembling elements of unrelated proteins including the LDL receptor, laminin, neural cell adhesion molecules, and EGF. These data suggest that the HSPG core protein has versatile properties and biological functions. The molecular weight of the actual core protein was calculated to be 466,876, which is consistent with previous molecular size estimates made by SDS-PAGE. These studies have revealed sizes of 400–500-kD for the core proteins from mouse, bovine, and man (25, 26, 50, 56). Allowing some increase in the apparent molecular weight due to Asn-linked glycosylation, the molecular mass of 500 kD would seem to be a close estimate. The length of the entire molecule can be estimated from the domain structure to be least 100–150 nm. This is significantly longer than the length of 60–80-nm estimated from electron micrographs. In these electron micrographs there appears to be, however, some flexibility in the core protein, especially in the second half which does not contain the side chains, apparently corresponding to the domain IV in the core protein. This could allow some folding which would lead to an underestimation of the length of the elongated molecule.

In addition to the large HSPG, basement membranes have been reported to contain several smaller size HSPGs with core proteins ranging from 21 to 350 kD (41, 62). Some of these proteins seem to be immunologically related with the 467-kD core protein and they might even be proteolytic products. For example, it has been reported (34, 41) that the glomerular basement membrane and calf lens capsule contain related 200–250 and 340-kD HSPG core proteins as the predominant species. Additionally, the glomerular basement membrane was shown to contain an immunologically related 400-kD component. These different sizes of immunologically related protein species may be cell- or tissue-specific variants of the same gene products or proteolytic fragments. As described below, however, there may be evidence indicating alternative splicing of the primary transcription product that can give rise to different size core proteins. There is also evidence that basement membranes contain low molecular mass HSPGs genetically distinct from the 467-kD HSPG. Soroka and Farquhar (62) have described a basement membrane HSPG with a core protein of ~40 kD. This protein was shown to be immunologically unrelated to the 467-kD basement membrane HSPG core protein. The actual genetic diversity as well as potential size variety produced through alternative splicing remains to be shown.

Location of HS Attachment Sites

The amino acid sequence of the human 467-kD HSPG core protein was shown to contain a total of 52 Ser-Gly sequences which have been suggested to serve as attachment sites for glycosaminoglycans. These sequences are located quite randomly throughout the polypeptide chain (see Fig. 3). Comparison of the Ser-Gly sequences and their adjacent amino acids revealed three short sequences in the amino-terminal end domain I, whose sequences are similar to the predicted HS attachment sites in the HSPGs syndecan (60) and glypican (13). All the sequences contain a Ser-Gly doublet followed by an acidic residue aspartate or glutamate and are preceded by a neutral residue, glycine, isoleucines, or threonine. A similar consensus sequence is also found in the chondroitin sulfate proteoglycan versican (69). This consensus sequence might be used by the same xylosyl transferase since the polysaccharide linkage region is identical in chondroitin sulfate proteoglycans and HSPGs (for review see 38). The location of these three potential HS attachment sites in domain I would also be in agreement with previous protein and EM data locating three side chains to the end of the core protein.

In addition to the three consensus sequences in domain I, there are three sites which are similar to the second consensus sequence Ser-Gly-X-Gly for glycosaminoglycan attachment found in many proteoglycans. One of those is located in the second immunoglobulin repeat of domain IV and

the other two are present in domain V at similar locations at the initiation of subdomains GR-2 and GR-3 (see Fig. 3). Although these subdomains are homologous to corresponding subdomains in the large carboxyl-terminal end G domain of the laminin A and merosin chains this consensus sequence is not present in laminin. These consensus sequences do not, however, contain the typical acidic residues preceding Ser-Gly (5) which makes them weaker candidates for being glycosaminoglycan attachment sites than the potential sites in domain I.

LDL Receptor Ligand Binding Motifs in the Basement Membrane HSPG

At present, it is not known whether the large basement membrane HSPG binds LDL and if so what physiological significance it has. However, it has been suggested that the anionic HS side chains of proteoglycans can, as such, bind both LDL and lipoprotein(a), possibly binding to the same regions that participate in binding apo B to its receptor (7). As shown in this study, the core protein contains LDL receptor ligand binding-like regions close to the HS side chains containing domain. Consequently, the HS side chains and the LDL receptor-like domains of the basement membrane HSPG may, together, bind lipoproteins and, thus, contribute to their binding to the immediate subendothelial matrix.

Laminin-like Domain III

Domain III was shown to have a structure resembling the globular and cysteine-rich domains in the short arms of laminin. Unlike the cysteine-rich subdomains in the laminin chains, the CR-1 subdomain of HSPG begins with a half repeat. The cysteine-rich sequences and the surrounding globular structures are closely homologous, which indicates that they may have duplicated as units. It has been suggested that the 9-amino acid insertion in the CR-3 subdomain is the site where duplication occurred (47). This would indicate that the original duplicated unit would contain one cysteine repeat, the cysteine repeat with the globular domain inserted in the middle of it and two full cysteine repeats. Interestingly, there is a deletion of at least three amino acids including two conserved cysteines in the CR-2 subdomain exactly at the boundary of the postulated duplicating unit. The CR-1 subdomain has only a half left of the first repeat.

Several biological functions have been assigned to the short arms of laminin. Proteolytic fragments containing the short arms of laminin have been found to mediate cell attachment, and several peptides comprising regions of laminin chains have been reported to have cell binding activity. The laminin-like domain III in HSPG does not, however, contain any sequences, such as RGD or YIGSR found in the laminin chains and reported to promote cell adhesion (22, 55). The cysteine repeats in the laminin rodlike regions have limited homology to EGF and they have been indicated to have growth-factor activity (48). The laminin-like cysteine repeats in the HSPG core protein could also serve the same function. It is also possible that the laminin-like region in HSPG core protein is involved in the interactions of the protein core with other basement membrane components.

N-CAM-like Domain IV

Domain IV was shown to consist of 21 repeats similar to those in the Ig superfamily. Closest homology was found to the family of neural cell adhesion molecules. These repeats have a high sequence identity (see Fig. 7). Most notably repeats 6–15 share a particularly high degree of sequence identity not typical for other repeats of the HSPG core protein. This indicates that the original repeats have diverged early from the other cell adhesion molecules and the similarity between the repeats suggests that they have later duplicated several times. Comparison with the similar domain in the mouse cDNA clone BPG7 reported earlier (47) reveals that there are several sequences which are conserved in mouse and human. However, there is a significant difference between the human and mouse sequences in this region of the protein, since it appears that the human protein contains seven repeats inserted into the middle of the second Ig-like repeat in the mouse amino acid sequence. This suggests either a very recent duplication event in the evolution or alternative splicing. According to these data the molecular mass of the mouse protein would be considerably smaller than that of the human protein, 396 instead of 467 kD. However, no alternative mRNAs, which would account for such large regions ~2 kb, have been observed so far in either human (see Fig. 2.) or mouse mRNAs. The borders where the possible duplication occurred are in the middle of Ig repeats suggesting that this could be the exon-intron boundary. This would be similar to the Ig repeat in coding exons in the N-CAM gene (12), but different in Ig exons in antibodies, which code for a full repeat. The exon-intron structure has been considered as evidence for the early divergence between the neural cell adhesion and immunoglobulins. In this sense the Ig-like repeats in HSPG core protein would also appear to belong to the family of neural cell adhesion molecules.

Members of the Ig superfamily of proteins have been implicated in many functions. Neural cell adhesion molecules especially have been demonstrated to mediate cell–cell adhesion (for review see 17). The Ig domains participate in homophilic binding between cell adhesion molecules. In addition, they may contain sequences important in binding heparin and heparan sulfate such as in N-CAM (30). The function of Ig domains in the HSPG core protein has not been demonstrated, but they could participate in either homophilic binding between two HSPG core proteins, or in heterophilic binding to other proteins, such as cell adhesion molecules. It has been observed previously that the purified core protein has a strong tendency to aggregate in dimers and even stellate structures (25).

Domain V: Elements from the G Domain of Laminin A Chains Separated by EGF-like Repeats

The large HSPG core protein was shown to contain a large presumably globular carboxyl-terminal end domain with homology to the large G domain of the laminin A and merosin chains (18, 46, 59). In laminin, the G domain, which is also located at the carboxyl-terminal end, has been suggested to have a role in the supramolecular assembly of the basement membrane network. It may contain binding sites for type IV collagen and the basement membrane HSPG. Furthermore, the globular domain and fragments of it have been shown to have cell binding activity (61). It is possible that the HSPG domain V shares some of the functions of the laminin G domain. For example, it has been shown by in vitro studies that

the HSPG core proteins aggregate at the ends opposite the HS chains. It is also possible that the two Leu-Arg-Glu sequences (28) have motor neuron attachment activities.

Despite their similarities, the HSPG domain VI has several features that distinguish it substantially from laminin. One feature is that HSPG has only three globular subdomains—GR-1, GR-2, and GR-3—as opposed to five in the laminin G domain. There are two doublets of ~40-residue EGF motifs located between the GR subdomains in the HSPG core, whereas no such repeats are present in the G domain of laminin. Sequences with homology to the EGF motif have been found in several proteins in single or multiple copies. Characteristically these motifs have six conserved cysteines which form disulfide bonds producing a predicted rigid loop structure (1). The constant rigid structure ensures the exposure of biologically active residues such as in the actual EGF where short segments have been shown to provide minimal ligand causing plerotropic proliferative and developmental effects (35). EGF motifs have been found in the LDL receptor; plasminogen activators; laminin; coagulation factors VII, IX, X, and XII; and some other proteins (1). Coagulation factor IX contains two EGF motifs which have been proposed to participate in binding to a cell surface receptor. However, recent studies show that the first EGF domain is not likely to have receptor binding activity (9) and the only remaining function appears to be calcium binding (24). In the absence of protein and protein-cell interaction data the role of the EGF motifs in the HSPG core protein can only be speculated.

The present work provides the possibilities for generating synthetic peptides and recombinant proteins from different regions of the large core protein and specific antibodies to these regions. This will enable us to do actual protein and cell biological experiments to study the different functions of this multidomain protein.

This work was supported in part by grants from the Academy of Finland, The Sigrid Juselius Foundation, and the Finnish Cancer Institute.

Received for publication 26 August 1991 and in revised form 30 September 1991.

References

1. Appella, E., I. T. Weber, and F. Blasi. 1988. Structure and function of epidermal growth factor-like regions in proteins. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 231:1-4.
2. Barthels, D., M.-J. Santoni, W. Wille, C. Ruppert, J.-C. Chaix, M.-R. Hirsch, J. C. Fontecilla-Camps, and C. Goridis. 1987. Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a M_r 79,000 polypeptide without a membrane-spanning region. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:907-914.
3. Beck, K., J. Spring, R. Chiquet-Ehrismann, J. Engel, and M. Chiquet. 1991. Structural motifs of the extracellular matrix proteins laminin and tenascin. *Springer Ser. Biophys.* 7:231-256.
4. Becker, J. W., H. P. Erickson, S. Hoffman, B. A. Cunningham, and G. M. Edelman. 1989. Topology of cell adhesion molecules. *Proc. Natl. Acad. Sci. USA.* 86:1088-1092.
5. 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.
6. Butkowski, R. J., J. P. M. Langeveld, J. Wieslander, J. Hamilton, and B. G. Hudson. 1987. Localization of the Goodpasture epitope to a novel chain of basement membrane collagen. *J. Biol. Chem.* 262:7874-7877.
7. Cardin, A. D., and H. J. R. Weintraub. 1989. Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis.* 9:21-32.
8. Cavener, D. R., and S. C. Ray. 1991. Eukaryotic start and stop translation sites. *Nucleic Acids Res.* 19:3185-3192.
9. Cheung, W.-F., D. L. Straight, K. J. Smith, S.-W. Lin, H. R. Roberts, and D. W. Stafford. 1991. The role of the epidermal growth factor-1 and hydrophobic stack domains of human factor IX in binding to endothelial cells. *J. Biol. Chem.* 266:8797-8800.
10. Corpet, F. 1988. Multiple sequence alignment with hierarchial clustering. *Nucleic Acids Res.* 16:10881-10890.
11. Couchman, J. R., and A. V. Ljubimov. 1989. Mammalian tissue distribution of a large heparan sulfate proteoglycan detected by monoclonal antibodies. *Matrix.* 9:311-321.
12. Cunningham, B. A., J. J. Hemperly, B. A. Murray, E. A. Prediger, R. Brackenbury, and G. M. Edelman. 1987. Neural cell adhesion molecule: structure, immunoglobulin-like domains, cell surface modulation, and alternative RNA splicing. *Science (Wash. DC).* 236:799-806.
13. David, G., V. Lories, B. Decock, P. Marynen, J.-J. Cassiman, and H. Van den Berghe. 1990. Molecular cloning of a phosphatidylinositol-anchored membrane heparan sulfate proteoglycan from human lung fibroblasts. *J. Cell Biol.* 111:3165-3176.
14. Devereux, J., P. Haberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387-395.
15. DiScipio, R. G., M. R. Gehring, E. R. Podack, C. C. Kan, T. E. Hugli, and G. H. Fey. 1984. Nucleotide sequence of cDNA and derived amino acid sequence of human complement component C9. *Proc. Natl. Acad. Sci. USA.* 81:7298-7302.
16. Dodge, G. R., I. Kovalszky, M.-L. Chu, J. R. Hassell, O. W. McBride, H. F. Yi, and R. V. Iozzo. 1991. Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellular expression, and mapping of the gene (HSPG2) to the short arm of human chromosome 1. *Genomics.* 10:673-680.
17. Edelman, G. M., and K. L. Crossin. 1991. Cell adhesion molecules: implications for a molecular histology. *Annu. Rev. Biochem.* 60:155-190.
18. Ehrig, K., I. Leivo, W. S. Argraves, E. Ruoslahti, and E. Engval. 1990. Merosin, a tissue-specific basement membrane protein, is a laminin-like protein. *Proc. Natl. Acad. Sci. USA.* 87:3264-3268.
19. Farquhar, M. G., P. J. Courtoy, M. C. Lemkin, and Y. S. Kanwar. 1982. Current knowledge of the functional architecture of the glomerular basement membrane. *In* New Trends in Basement Membrane Research. K. Kühn, H. H. Schöne, and R. Timpl, editors. Raven Press, New York. 9-30.
20. Furley, A. J., S. B. Morton, D. Manalo, D. Karagogeos, J. Dodd, and T. M. Jessel. 1990. The axonal glycoprotein TAG-1 is an immunoglobulin superfamily member with neurite outgrowth-promoting activity. *Cell.* 61:157-170.
21. Gershagen, S., P. Fernlund, and Å. Lundwall. 1987. A cDNA coding for human sex hormone binding globulin. Homology to vitamin K-dependent protein S. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 220:129-135.
22. Gragt, J., Y. Iwamoto, M. Sasaki, G. R. Martin, H. K. Kleinman, F. A. Robey, and Y. Yamada. 1987. Identification of an amino acid sequence in laminin mediating cell attachment, chemotaxis, and receptor binding. *Cell.* 48:989-996.
23. Gray, A., T. J. Dull, and A. Ullrich. 1983. Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-molecular weight protein precursor. *Nature (Lond.).* 303:722-725.
24. Handford, P. A., M. Baron, M. Mayhew, A. Willis, T. Beesley, G. G. Brownlee, and I. D. Campbell. 1990. The first EGF-like domain from human factor IX contains a high-affinity calcium binding site. *EMBO (Eur. Mol. Biol. Organ.) J.* 9:475-480.
25. Hassell, J. R., P. Gehron-Robey, H.-J. Barrach, J. Wilczek, S. I. Rennard, and G. R. Martin. 1980. Isolation of a heparan sulfate-containing proteoglycan from basement membrane. *Proc. Natl. Acad. Sci. USA.* 77: 4494-4498.
26. Heremans, A., B. Van Der Schueren, B. De Cock, M. Paulsson, J.-J. Cassiman, H. Van Den Berghe, and G. David. 1989. Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclonal antibodies decorate the pericellular matrix of connective tissue cells and the stromal side of basement membranes. *J. Cell Biol.* 109:3199-3211.
27. Hostikka, S. L., R. L. Eddy, M. G. Byers, M. Höyhtyä, T. B. Shows, and K. Tryggvason. 1990. Identification of a distinct type IV collagen a chain with restricted kidney distribution and assignment of its gene to the locus of X chromsosome-linked Alport syndrome. *Proc. Natl. Acad. Sci. USA.* 87:1606-1610.
28. Hunter, D. D., B. E. Porte, J. W. Bulock, S. P. Adams, J. P. Merlie, and J. R. Sanes. 1989. Primary sequence of a motor neuron-selective adhesive site in the synaptic basal lamina protein S-laminin. *Cell.* 59:905-913.
29. Hunter, D. D., V. Shah, J. P. Merlie, and J. R. Sanes. 1989. A laminin-like adhesive protein concentrated in the synaptic cleft of the neuromuscular junction. *Nature (Lond.).* 338:229-234.
30. Iozzo, R. V. 1984. Biosynthesis of heparan sulfate proteoglycan by human colon carcinoma cells and its localization at the cell surface. *J. Cell Biol.* 99:403-417.
31. Kallunki, P., R. L. Eddy, M. G. Byers, M. Kestilä, T. B. Shows, and K. Tryggvason. 1991. Cloning of human heparan sulfate proteoglycan core protein assignment of the gene (HSPG1) to 1p36.1 → p35 and identification of a BamHI restriction fragment length polymorphism. *Genomics.* 11:389-396.
32. Kanwar, Y. S., and M. G. Farquhar. 1979. Anionic sites in the glomerular

basement membrane. In vivo and in vitro localization to the laminae rarae by cationic probes. *J. Cell Biol.* 81:137–153.
33. Kato, M., Y. Koike, S. Suzuki, and K. Kimata. 1988. Basement membrane proteoglycan in various tissues: characterization using monoclonal antibodies to the Engelbreth-Holm-Swarm mouse tumor low density heparan sulfate proteoglycan. *J. Cell Biol.* 106:2203–2210.
34. Klein, D. J., D. M. Brown, T. R. Oegema, P. E. Brenchley, J. C. Anderson, M. A. J. Dickinson, E. A. Horigan, and J. R. Hassell. 1988. Glomerular basement membrane proteoglycans are derived from a large precursor. *J. Cell Biol.* 106:963–970.
35. Komoriya, A., M. Hortsch, C. Meyers, M. Smith, H. Kanety, and J. Schlessinger. 1984. Biologically active synthetic fragments of epidermal growth factor: localization of a major receptor-binding region. *Proc. Natl. Acad. Sci. USA.* 81:1351–1355.
36. Ledbetter, S. R., B. Tyree, J. R. Hassell, and E. A. Horigan. 1985. Identification of the precursor protein to basement membrane heparan sulfate proteoglycans. *J. Biol. Chem.* 260:8106–8113.
37. Ledbetter, S. R., L. W. Fisher, and J. R. Hassell. 1987. Domain structure of the basement membrane heparan sulfate proteoglycan. *Biochemistry.* 26:988–995.
38. Lindahl, U., and L. Kjellen. 1987. Biosynthesis of heparin and heparan sulfate. *In* Biology of Proteoglycans. T. M. Wight and R. P. Mecham, editors. Academic Press, NY. 59–104.
39. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. 545 pp.
40. Marynen, P., J. Zhang, J.-J. Cassiman, H. Van den Berghe, and G. David. 1989. Partial primary structure of the 48- and 90-kilodalton core proteins of cell surface-associated heparan sulfate proteoglycans of lung fibroblasts. Prediction of an integral membrane domain and evidence for multiple distinct core proteins at the cell surface of human lung fibroblasts. *J. Biol. Chem.* 264:7017–7024.
41. Mohan, P. S., and R. G. Spiro. 1991. Characterization of heparan sulfate proteoglycan from calf lens capsule and proteoglycans synthesized by cultured lens epithelial cells. Comparison with other basement membrane proteoglycans. *J. Biol. Chem.* 266:8567–8575.
42. Montell, D. J., and C. S. Goodman. 1988. Drosophila substrate adhesion molecule: sequence of laminin B1 chain reveals domains of homology with mouse. *Cell.* 53:463–473.
43. Montell, D. J., and C. S. Goodman. 1989. Drosophila laminin: sequence of B2 subunit and expression of all three subunits during embryogenesis. *J. Cell Biol.* 109:2441–2453.
44. Moos, M., R. Tacke, H. Scherer, D. Teplow, K. Früh, and M. Schachner. 1988. Neural adhesion molecule L1 as a member of the immunoglobulin superfamily with binding domains similar to fibronectin. *Nature (Lond.).* 334:701–703.
45. Morrison, K. E., G. G. Germino, and S. T. Reeders. 1991. Use of the polymerase chain reaction to clone and sequence a cDNA encoding the bovine a3 chain of type IV collagen. *J. Biol. Chem.* 266:34–39.
46. Nissinen, M., R. Vuolteenaho, R. Boot-Handford, P. Kallunki, and K. Tryggvason. 1991. Primary structure of the human laminin A chain. Limited expression in human tissues. *Biochem. J.* 276:369–379.
47. Noonan, D. M., E. A. Horigan, S. R. Ledbetter, G. Vogeli, M. Sasaki, Y. Yamada, and J. R. Hassell. 1988. Identification of cDNA clones encoding different domains of the basement membrane heparan sulfate proteoglycan. *J. Biol. Chem.* 263:16379–16387.
48. Panayotou, G., P. End, M. Aumailey, R. Timpl, and J. Engel. 1989. Domains of laminin with growth-factor activity. *Cell.* 56:93–101.
49. Parthasarathy, N., and R. G. Spiro. 1982. Effect of diabetes on the glycosaminoglycan component of the human glomerular basement membrane. *Diabetes.* 31:738–741.
50. Paulsson, M., P. D. Yurchenco, G. C. Ruben, J. Engel, and R. Timpl. 1987. Structure of low density heparan sulfate proteoglycan isolated from a mouse tumor basement membrane. *J. Mol. Biol.* 197:297–313.
51. Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA.* 85:2444–2448.
52. Pikkarainen, T., R. Eddy, Y. Fukushima, M. Byers, T. Shows, T. Pihlajaniemi, M. Saraste, and K. Tryggvason. 1987. Human laminin B1 chain: a multidomain protein with gene (LAMB1) locus in the q22 region of chromosome 7. *J. Biol. Chem.* 262:10454–10462.
53. Pikkarainen, T., T. Kallunki, and K. Tryggvason. 1988. Human laminin B2 chain. Comparison of the complete amino acid sequence with the B1 chain reveals variability in sequence homology between different structural domains. *J. Biol. Chem.* 263:6751–6758.
54. Ranscht, B. 1988. Sequence of contactin, a 130-kD glycoprotein concentrated in areas of interneuronal contact, defines a new member of the immunoglobulin supergene family in the nervous system. *J. Cell Biol.* 107:1561–1573.
55. Ruoslahti, E., and M. D. Pierschbacher. 1987. New perspectives in cell adhesion: RGD and integrins. *Science (Wash. DC).* 238:491–497.
56. Saku, T., and H. Furthmayr. 1989. Characterization of the major heparan sulfate proteoglycan secreted by bovine aortic endothelial cells in culture. Homology to the large molecular weight molecule of basement membranes. *J. Biol. Chem.* 264:3514–3523.
57. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5467.
58. Sasaki, M., and Y. Yamada. 1987. The laminin B2 chain has a multidomain structure homologous to the B1 chain. *J. Biol. Chem.* 262:17111–17117.
59. Sasaki, M., H. K. Kleinman, H. Huber, R. Deutzmann, and Y. Yamada. 1988. Laminin, a multidomain protein. The A chain has a unique globular domain and homology with the basement membrane proteoglycan and the laminin B chains. *J. Biol. Chem.* 263:16536–16544.
60. Saunders, S., M. Jalkanen, S. O'Farrell, and M. Bernfield. 1989. Molecular cloning of syndecan, an integral membrane proteoglycan. *J. Cell Biol.* 108:1547–1556.
61. Sonnenberg, A., C. S. T. Linders, P. W. Modderman, C. H. Dmasky, M. Aumailley, and R. Timpl. 1990. Integrin recognition of different cell-binding fragments of laminin (P1, E3, E8) and evidence that $\alpha 6\beta 1$ but not $\alpha 6\beta 4$ function as a major receptor for fragment E8. *J. Cell Biol.* 110: 2145–2155.
62. Soroka, C. J., and M. G. Farquhar. 1991. Characterization of a novel heparan sulfate proteoglycan found in the extracellular matrix of liver sinusoids and basement membranes. *J. Cell Biol.* 113:1231–1241.
63. Sûdhof, T. C., J. L. Goldstein, M. S. Brown, and D. W. Russell. 1985. The LDL receptor gene: a mosaic of exons shared with different proteins. *Science (Wash. DC).* 228:815–822.
64. Timpl, R. 1989. Structure and biological activity of basement membrane proteins. *Eur. J. Biochem.* 180:487–502.
65. Vernier, R. L., D. J. Klein, S. P. Sisson, J. D. Mahan, T. R. Oegema, and D. M. Brown. 1983. Heparan sulfate–rich anionic sites in the human glomerular basement membrane. Decreased concentration in congenital nephrotic syndrome. *N. Engl. J. Med.* 309:1001–1009.
66. von Hejne, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* 14:4683–4690.
67. Williams, A. F., and A. N. Barday. 1988. The immunoglobulin superfamily: domains for cell surface recognition. *Annu. Rev. Immunol.* 6: 381–405.
68. Yarden, Y., J. A. Escobedo, W.-J. Kuang, T. L. Yang-Feng, T. O. Daniel, P. M. Tremble, E. Y. Chen, M. E. Ando, R. N. Harkins, U. Francke, V. A. Fried, A. Ullrich, and L. T. Williams. 1986. Structure of the receptor for platelet-derived growth factor helps define a family of closely related growth factor receptors. *Nature (Lond.).* 323:226–232.
69. Zimmermann, D. R., and E. Ruoslahti. 1989. Multiple domains of the large fibroblast proteoglycan, versican. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:2975–2981.