Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 7683–7687, October 1986
Biochemistry

Primary structure of an extracellular matrix proteoglycan core protein deduced from cloned cDNA

TOM KRUSIUS AND ERKKI RUOSLAHTI

Cancer Research Center, La Jolla Cancer Research Foundation, 10901 North Torrey Pines Road, La Jolla, CA 92037

Communicated by Marilyn G. Farquhar, July 7, 1986

ABSTRACT **The core protein of a small chondroitin/dermatan sulfate proteoglycan expressed by human fibroblasts and present in extracellular matrices in association with collagen has been cloned from a λgt11 fibroblast cDNA library. cDNA clones were isolated by use of antibodies specific for the intact proteoglycan and antibodies against a peptide synthesized on the basis of the amino-terminal sequence of the core protein. A 1.8-kilobase cDNA was found to code for a prepro core protein composed of a signal peptide, a propeptide, and a mature core protein of 329 amino acids. The amino-terminal amino acid sequence deduced from the cDNA sequence was identical to that previously obtained by protein sequencing. The core protein contains three Ser-Gly dipeptide sequences, of which one is substituted with glycosaminoglycan. A protein data base homology search established the core protein sequence is a unique sequence distinct from published amino acid sequences. RNA blot hybridizations, performed using the cloned cDNA as a probe, revealed two related transcripts of 1.6 and 1.9 kilobases in RNA from both human fibroblast and placental tissue. Hybridization of genomic DNA restriction fragments suggested that there is one gene for the core protein of this proteoglycan and possibly one other closely related gene. Availability of the cloned cDNA for the proteoglycan now makes it possible to apply methods of molecular biology to study the collagen-binding and cell attachment-inhibiting properties of this proteoglycan.**

Proteoglycans are abundant molecules in extracellular matrices, in basement membranes, and at cell surfaces (1, 2). They exist in many different forms, some of which are tissue- or cell-type specific (1, 2). The structural relationships and functional properties of proteoglycans are poorly understood. This is mostly due to a lack of detailed structural information about the core proteins which primarily determine the specific properties of a proteoglycan.

A small chondroitin/dermatan sulfate proteoglycan, or group of proteoglycans, is of particular interest due to its abundance and functional properties. This proteoglycan is the major sulfated product of fibroblasts and is abundant in the extracellular matrices of connective tissues (2). It binds to type I collagen, affecting fibril formation (3, 4), and inhibits the cell attachment-promoting activity of collagen and fibronectin by binding to these molecules near their cell-binding site (5). At least the binding to collagen appears to be mediated by the core protein, since treatment of the proteoglycan with chondroitinase, which removes most of the glycosaminoglycan, does not eliminate the binding to collagen (ref. 4 and unpublished results).

This proteoglycan has a core protein of M_r 38,000 and usually carries one glycosaminoglycan chain (2, 6, 7). It is not known whether the core protein is a single protein or whether it might consist of a family of closely related molecules. Recent work on bovine cartilage proteoglycans suggests that at least two core protein species are present in this type of proteoglycan (8). To gain more information about the core protein structure, we have turned to cDNA cloning.

In this report, we describe the cloning and sequencing of the cDNA encoding the core protein of the major chondroitin/dermatan sulfate proteoglycan produced by human fibroblasts. We call the proteoglycan PG40 to denote the molecular weight of its core protein precursor. The amino acid sequence of PG40 deduced from the cDNA is quite different from the other known amino acid sequence for an extracellular matrix proteoglycan, that from a rat yolk sac tumor (9). DNA and RNA blot hybridization analyses indicate that no more than two genes code for PG40 and proteoglycans closely related to it.

METHODS

Construction of λgt11 cDNA Library. RNA was extracted from human embryonic fibroblast cell line IMR-90 by the guanidinium isothiocyanate method (10) and poly(A)$^+$ RNA was isolated by chromatography on oligo(dT)-cellulose (11). Synthesis of the first cDNA strand was carried out using oligo(dT) as a primer for reverse transcription (12), and second-strand synthesis was carried out using ribonuclease H and DNA polymerase I (13). The double-stranded cDNA was made blunt-ended with nuclease S1, and endogenous *Eco*RI sites were protected with *Eco*RI methylase (12). *Eco*RI linkers were ligated to cDNA by use of bacteriophage T4 DNA ligase. After *Eco*RI digestion, cDNA was size-fractionated on Sepharose 4B, and cDNA larger than 0.5 kilobases (kb) was used for ligation to λgt11 arms previously restricted with *Eco*RI and treated with calf intestinal alkaline phosphatase (12). Phage DNA was packaged using the Packagene System (Promega Biotec, Madison, WI).

One nanogram of cDNA produced 7000 recombinants, and the library comprised 1.3×10^6 independent recombinants. Before amplification of the library in *Escherichia coli* Y1090 r^-m^+ (Promega Biotec), the recombinants accounted for more than 90% of the total phage. By immunological selection using various antibodies, 28 recombinants have been isolated from the library so far, and the average length of the inserts has been 1.2 kb.

Immunological Screening of Phage Library. A rabbit antiserum, which was raised against PG40 isolated from human fetal membranes and reactive with the main proteoglycan of cultured fibroblasts (6), was absorbed with *E. coli* Y1089 lysogen lysate as well as with human fibronectin and serum proteins coupled to Sepharose. Antibodies were affinity-purified on a column of fetal membrane proteoglycan coupled to Sepharose. These affinity-purified antibodies were used to screen 150,000 recombinants by the plaque method described by Young and Davis (14). Peroxidase-labeled goat anti-rabbit IgG second antibody (Bio-Rad) was used to visualize positive plaques. The identity of the isolated clones was confirmed by using a peptide antiserum raised against the amino-terminal

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: kb, kilobase(s).

peptide Asp-Glu-Ala-Ser-Gly-Ile-Gly-Pro-Glu-Val-Pro-Asp-Asp-Arg-Asp (6) of the proteoglycan. The peptide antiserum was absorbed with human serum proteins coupled to Sepharose and with bacteria. Antibodies were affinity-purified on a peptide-Sepharose column. Details of the production and reactivity of this antiserum will be published elsewhere.

Subcloning and DNA Sequencing. cDNA inserts, isolated by *Eco*RI digestion of recombinant phage DNA followed by preparative electrophoresis in agarose (12), were subcloned into pGEM1 vector and M13mp19 phage (12). Recombinant plasmids were isolated by the alkaline-lysis method (15). M13 deletion libraries of both strands of the inserts were constructed by the single-stranded M13 method of Dale *et al.* (16), using a kit from International Biotechnologies (New Haven, CT). Sequencing of the M13 deletion mutants was performed by the dideoxynucleotide chain-termination procedures of Sanger *et al.* (17), using deoxyadenosine 5'-[α-[^{35}S]thio]triphosphate as a tracer. cDNA sequences not found in the deletion libraries were sequenced by the dideoxynucleotide chain-termination method, using oligonucleotide primers which were synthesized, on the basis of the preceding sequences, with an Applied Biosystems (Foster City, CA) DNA synthesizer. Because of an interfering compression in the sequencing gel, a *Xho* II–*Pvu* II fragment (from base 312 to base 501) was also sequenced by the Maxam and Gilbert method (18). Homology searches were done utilizing the BIONET system of Intelligenetics (Palo Alto, CA) and the MicroGenie program supplied by Beckman.

RNA and DNA Blot Analysis. Total RNA or poly(A)$^+$ RNA isolated from cultured cells or from human placenta as described above was fractionated in 0.8% agarose/2.2 M formaldehyde gels, transferred to nitrocellulose filters, and hybridized with radioactively labeled probes (19). Hybridization was at 52°C in the presence of 50% formamide (19). Final washings were carried out at 68°C using 0.2× SSC/0.1% NaDodSO$_4$. (1× SSC is 0.15 M NaCl/0.015 M sodium citrate.)

Genomic DNA was isolated from human placenta (20), digested with restriction enzyme, fractionated in 0.8% agarose gel, and transferred to nitrocellulose filters (21). Hybridization of DNA blots was performed as described for RNA blots, but hybridization was carried out at 42°C or 48°C, and washings were done at 52°C or 68°C (low- and high-stringency conditions, respectively).

Restriction fragments were labeled directly in low-melting-point agarose by using random oligonucleotide primers, according to Feinberg and Vogelstein (22), and a kit from Pharmacia.

RESULTS

Isolation of cDNA Clones for PG40 Core Protein. To select a cell line suitable for the preparation of mRNA coding for PG40 core protein, supernatants from several fibroblast lines were examined, by immunoprecipitation with specific antibodies, for their production of the proteoglycan. PG40 was one of the major secreted proteins in IMR-90 human embryonal fibroblast cultures, containing up to 2.5% of all the radioactive methionine incorporated into secreted proteins from these cells. The IMR-90 cells were, therefore, used as the source of RNA for a cDNA library, which was prepared in the expression vector λgt11. Screening of this library with affinity-purified antibodies against the proteoglycan produced a number of positive clones. The clones were purified through secondary and tertiary screenings, and 15 clones were isolated from single plaques. Six of these clones were also positive when tested with affinity-purified antibodies against the amino-terminal peptide of PG40. The insert size of these clones varied from 0.8 kb to 2.0 kb.

The clone with the largest insert (clone 5E, 2.0 kb) was chosen for sequence analysis. The insert contained two endogenous *Eco*RI sites resulting in one *Eco*RI fragment of 1.8 kb and two smaller fragments, each about 0.15 kb. The 1.8-kb fragment contained the entire coding sequence of the proteoglycan. The two 0.15-kb fragments were found to be located in tandem at the 3' end of the cDNA. RNA blot hybridizations revealed that these fragments did not hybridize with the core protein transcript, and they were, therefore, not studied further.

cDNA Sequence and Inferred Amino Acid Sequence of PG40 Core Protein. The sequencing strategy and the restriction map of the 1.8-kb insert are shown in Fig. 1. The complete sequence of the cDNA was established by sequencing both strands. Fig. 2 shows the cDNA sequence and the inferred amino acid sequence. The cDNA is 1778 nucleotides long and includes the complete coding region, 5' and 3' flanking sequences, and the *Eco*RI linkers. Only one open reading frame adequate to code for the core protein is present. This reading frame codes for 359 amino acids corresponding to a protein with a molecular weight of 39,739. The initiation site for translation is defined by an AUG codon and the presence of three upstream stop codons in the same reading frame.

The amino acid sequence inferred from the cDNA starts with a hydrophobic peptide of 16 residues that has features of a typical signal peptide (24). Both the nucleotide sequence upstream from the translational initiation site and the amino acid sequence in the region around the possible signal-peptidase cleavage site are in good agreement with the published consensus sequences for such regions (25, 26). An apparent propeptide precedes the amino-terminal amino acid sequence of the mature core protein. The sequence deduced from the cDNA is identical to that previously obtained by protein sequencing for the amino terminus (6), confirming the identity of the cDNA clone. The mature core protein is composed of 329 amino acids and has a molecular weight of 36,319.

The 3' flanking sequence of the insert in clone 5E is 0.6 kb long but does not extend to a poly(A)$^+$ tail. However, a consensus termination and polyadenylylation sequence (AATAAA) is present 63 bases downstream from the A of the stop codon. An AATATA sequence, which in some rare cases serves as a polyadenylylation signal (27), occurs twice (bases 1193–1198 and 1222–1227) downstream from the AATAAA sequence. Furthermore, a CACTG sequence (bases 1374–1378) and a TTCAA sequence (bases 1448–1452), both of which have been proposed to have a role in the formation of the 3' terminus of mRNA (28–30) are present downstream from the possible polyadenylylation signals.

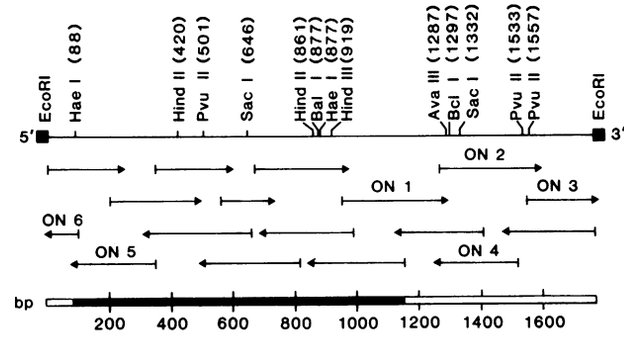

FIG. 1. Composite restriction endonuclease map and sequencing strategy of cDNA clone 5E coding for the core protein of PG40. The sequences of the coding and noncoding regions are indicated by the solid and open bars, respectively. Arrows indicate the 3' end of M13 phage deletion mutants and the length of the sequence read. ON indicates oligonucleotide-primed sequencing. bp, Base pairs.

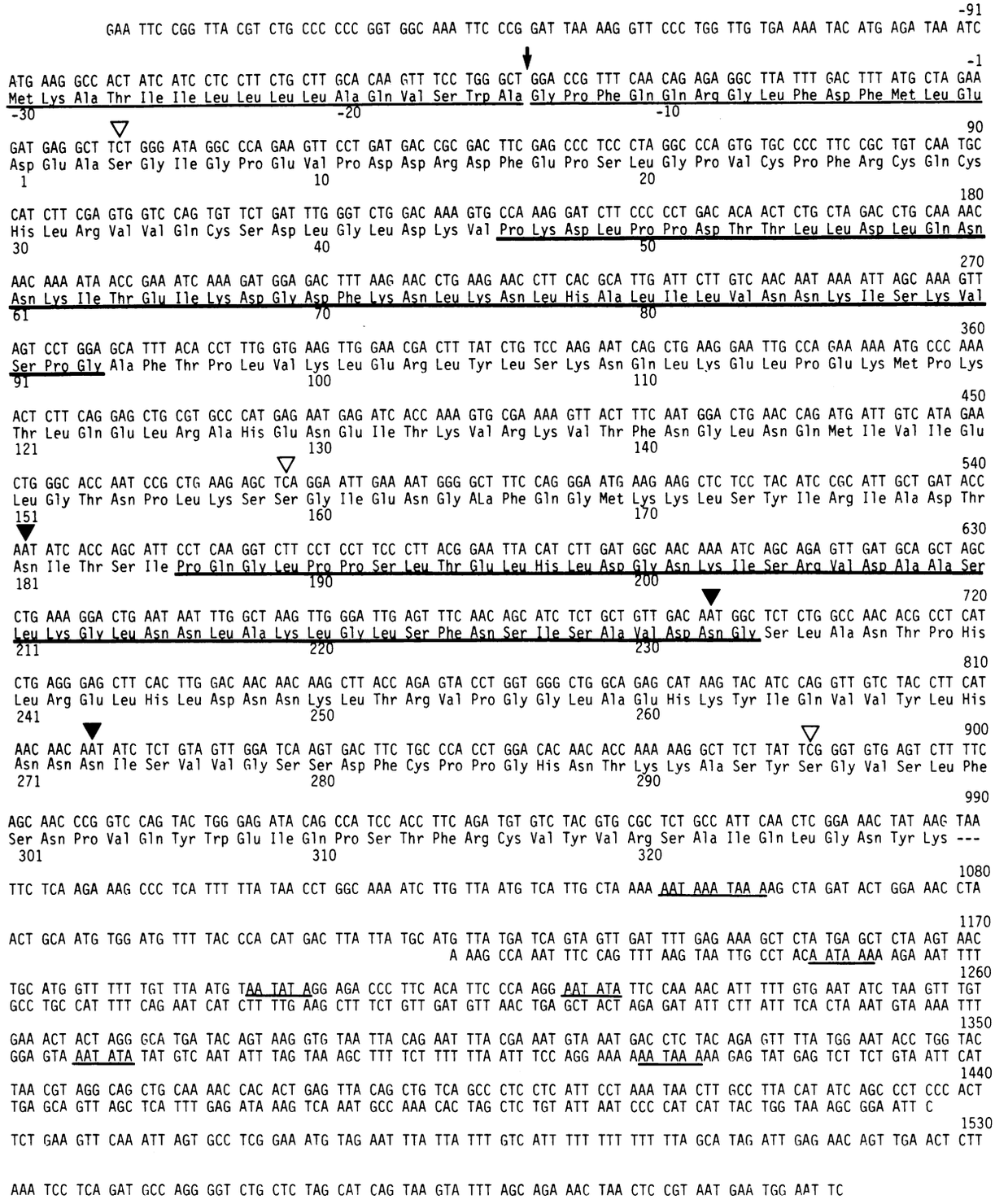

FIG. 2. Nucleotide sequence and deduced amino acid sequence of PG40 core protein cDNA. Nucleotide and amino acid residues are numbered from the amino terminus of the mature core protein. Nucleotides upstream from the amino terminus and amino acids in the signal and propeptide sequences are indicated by negative numbers. Signal and propeptide sequences are shown underlined (thin line) and the possible signal-peptidase cleavage site is indicated by an arrow. Potential glycosylation sites for N-linked glycans and glycosaminoglycans are shown by solid and open triangles, respectively. Serine at position 4 is known to carry a glycosaminoglycan (23). A 48-residue homologous sequence present twice in the core protein is shown underlined (bold line). The alternative 3′ flanking sequence contained in clone 3C (lower line) is shown below the sequence of clone 5E (upper line) starting at base 1122 where the sequences diverge. Possible polyadenylylation signal sequences are underlined.

To confirm the mRNA structure inferred from clone 5E, inserts from three other clones were subcloned into M13 phage and partially sequenced. Clone 12E was found to contain the complete coding sequence including the signal peptide and propeptide sequences. Clones 12D and 3C also contain these sequences, but 12D lacks the extreme 3′ end of the coding sequence, whereas clone 3C is truncated at the 5′ end. The partial sequences obtained from these clones for the 5′ flanking sequence and the coding region are in complete agreement with the sequence of clone 5E. However, clones 12E and 3C were found to contain a different 3′ flanking sequence than clone 5E. For the first 134 bases, the 3′

flanking sequences of all three clones are identical, but after base 1121, the sequences of clones 12E and 3C, while identical with one another, diverge from that of clone 5E. The 3′ flanking sequence of clone 3C is shown in Fig. 2. It includes the consensus polyadenylylation sequence AATAAA twice (bases 1267–1272 and 1319–1324) but does not extend to a poly(A)$^+$ tail.

Properties of the PG40 Core Polypeptide. The pI value estimated from the deduced amino acid sequence for the core protein is high (9.8) due to the presence of several short sequences enriched in basic amino acids. The amino acid sequence also includes three Ser-Gly dipeptide sequences. A glycosaminoglycan chain is attached to one such serine, the one at position 4 (23). In addition, three glycosylation sites for possible *N*-glycosidic substitution are present, in agreement with the previously reported number of *N*-glycans in the proteoglycan (7).

Two repeats of a homologous sequence of 48 residues (amino acids 46–93 and 186–233) are present in the core protein. The homology between these repeats on nucleotide and protein levels is 46% and 44%, respectively. The homology between the first 18 amino acids in the two repeats is even higher, 56%. The amino acid sequence of PG40 has no significant extended homology with the rat yolk sac tumor proteoglycan previously cloned in our laboratory (9). A protein data base (National Biomedical Research Foundation) homology search showed that the core protein has a unique sequence that lacks any significant homology with any previously published sequences.

RNA and DNA Blot Analysis. Blot hybridization analysis of human fibroblast RNA, using the cloned 1.8-kb fragment as a probe, revealed transcripts of 1.6 and 1.9 kb at about equal abundance (Fig. 3, lane 1). Both transcripts are also present in human placental RNA (lane 2), whereas neither transcript was found in a human epidermal carcinoma cell line (lane 3). Since the protein coding sequence in each of the four clones analyzed appears to be identical and two different 3′ flanking sequences are present in the cDNA clones, it seems likely that the presence of the two transcripts is due to heterogeneity in the 3′ flanking sequence.

Genomic DNA blots were made to analyze the genes related to PG40 cDNA. Filters were hybridized with either the cloned cDNA or with restriction fragments prepared from it. The 1.8-kb cDNA hybridized to two or three fragments produced by a variety of restriction enzymes. Shorter probes, corresponding either to the coding sequence or to the 3′ flanking sequence, revealed only one or two strongly hybridizing restriction fragments (Fig. 4), suggesting that there is only one gene for the core protein. In addition, most restriction enzymes produced a fragment that gave weak hybridization. This latter set of fragments was detected even when stringent hybridization and washing conditions were used (see *Methods*). These DNA fragments could have originated from the PG40 gene, if they have such short overlap with the probes that weak hybridization results. However, the fact that several restriction enzymes and two probes from different parts of the cDNA revealed such fragments makes this explanation unlikely. It is more likely that these fragments represent either a pseudogene related to the PG40 gene or a gene for a related protein. Hybridization at low-stringency conditions revealed no additional restriction fragments, suggesting that no more than two genes code for the core protein whose cDNA we have cloned or for proteins closely related to it.

DISCUSSION

In the present communication, we describe the cloning and sequencing of the cDNA encoding the core protein of a chondroitin/dermatan sulfate proteoglycan, PG40, from human fibroblasts. The identity of the cDNA clones was established by three criteria. First, affinity-purified rabbit antibodies prepared against the intact proteoglycan reacted with the fusion proteins produced by the clones. Second, some of the fusion proteins also reacted with a rabbit antiserum prepared against a 15 amino acid synthetic peptide from the known amino-terminal sequence of the core protein. Finally, DNA sequencing showed that the cDNA encodes the amino-terminal sequence of the core protein. The peptide antiserum was used because we found that a synthetic oligonucleotide from the amino-terminal sequence hybridized not only to the proteoglycan clones but also to clones that further analysis showed to be unrelated to the proteoglycan.

The core protein of PG40 appears to be synthesized in a

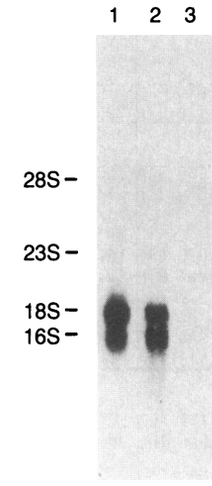

FIG. 3. RNA blot hybridization analysis of PG40 mRNA. RNA from human fibroblasts (lane 1, 2 μg of poly(A)$^+$ RNA), human placenta (lane 2, 10 μg of total RNA), and A431 human epidermal carcinoma cells (lane 3, 10 μg of total RNA) was separated in 0.8% agarose/2.2 M formaldehyde gels, blotted, and hybridized to an *Eco*RI–*Sac* I restriction fragment from the protein-coding region of the proteoglycan. The size markers were calf liver 28S (4.7 kb) and 18S (1.9 kb) rRNA and *E. coli* 23S (3.1 kb) and 16S (1.6 kb) rRNA.

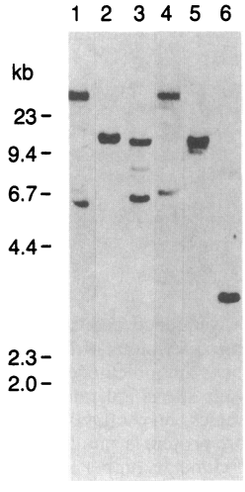

FIG. 4. Genomic DNA blot hybridization analysis. DNA isolated from human placenta was restricted with *Pst* I (lanes 1 and 4), *Bgl* II (lanes 2 and 5), or *Hin*dIII (lanes 3 and 6), fractionated in 0.8% agarose gels, blotted, and hybridized to an *Eco*RI–*Hin*cII restriction fragment extending from the 5′ end of the cDNA to base 862 (lanes 1–3) or to a *Hin*cII–*Eco*RI restriction fragment (bases 863–1773; lanes 4–6). A *Hin*dIII digest of λ phage DNA was used to provide molecular size standards.

prepro form and processed into the mature core protein of 329 amino acids. Interestingly, the core protein of a rat yolk sac tumor proteoglycan core protein, although otherwise entirely different (9), is also synthesized in a prepro form (33). The mature core protein of PG40 has a molecular weight of 36,319, which is in good agreement with the value of 38,000 determined by $NaDodSO_4$/polyacrylamide gel electrophoresis for the nonglycosylated core protein obtained from tunicamycin-treated fibroblasts (7).

The deduced amino acid sequence of the core protein includes three Ser-Gly dipeptides. In the known proteoglycan amino acid sequences, glycosaminoglycan chains are attached to the serine residues of Ser-Gly sequences (31). The serine residue of the Ser-Gly sequence at positions 4 and 5 is known to carry a glycosaminoglycan (23). The two other Ser-Gly sequences may become substituted by a second glycosaminoglycan chain in some of the molecules but remain unsubstituted in most, possibly all, models (23). In contrast to the presence of the dispersed Ser-Gly dipeptides in different parts of PG40 core protein, the rat yolk sac tumor proteoglycan contains a continuous 49 amino acid Ser-Gly repeat, which functions as the attachment site for all of its approximately 14 chondroitin sulfate side chains (9).

It is obvious that a Ser-Gly dipeptide alone cannot be a sufficient signal for glycosaminoglycan attachment, since most proteins contain one or more of them and even in the fibroblast proteoglycan two of them remain unsubstituted. As pointed out earlier, the amino acid sequence around the serine residue that is substituted in this proteoglycan is completely conserved between the human and bovine proteoglycans (6, 32). Moreover, the 12 amino acid sequence that surrounds the substituted serine within the carboxyl terminus of the propeptide and the amino terminus of the mature core protein in PG40 (residues −3 through 9 in Fig. 2) is 50% homologous to a 12 amino acid stretch within the propeptide of the rat yolk sac tumor proteoglycan (33). It is possible that some structural features in the sequences flanking the Ser-Gly dipeptide or located close to such a sequence can serve as a recognition signal for the attachment of the glycosaminoglycan side chain to the serine residues. The availability of the cDNAs for core proteins makes it possible to identify such signals by site-directed mutagenesis and expression of the altered genes in mammalian cells.

The small fibroblast proteoglycan appears to be an abundant and widely distributed component in tissues. Proteoglycans isolated from a variety of tissues are similar to PG40 with regard to core protein size, glycosaminoglycan component, amino acid composition, and antigenicity (6, 8, 32, 34–41), and it is possible that they all represent the same proteoglycan. Our results, in suggesting limited heterogeneity for the core proteins, would be in agreement with such an assumption. Genomic DNA blots were consistent with there being only one gene for the fibroblast core protein and one related gene. RNA blot hybridization of human placental RNA also indicated restricted heterogeneity. Although placenta contains a variety of different cell types, only the two transcripts coding for the fibroblast proteoglycan were detected in placental RNA. Since fibroblasts and other mesenchymal cells are present in most tissues, it may be that many of the small chondroitin/dermatan sulfate proteoglycans isolated from different tissues are products of the gene coding for the fibroblast PG40 proteoglycan. However, because we did find evidence for one related gene and because others (8) have found in cartilage two distinct core proteins with the same size as that of the fibroblast proteoglycan, some limited heterogeneity may exist in this proteoglycan type.

The detailed function of PG40 is unknown, but, since it is an extracellular matrix component and binds to other matrix components, its function is likely to be related to matrix assembly and interactions. The cloned cDNA now provides new possibilities for the analysis of the biological roles of this proteoglycan by methods such as gene transfer.

This work was supported by Grant CA28896 and Cancer Center Support Grant CA30199 from the National Cancer Institute, Department of Health and Human Services. T.K. is the recipient of Fogarty International Fellowship F05 TW0 3477-01 and grants from the Finnish Cancer Foundation, Finnish Cultural Foundation, and the Sigrid Juselius Foundation, Helsinki, Finland.

1. Hascall, V. C. & Hascall, G. K. (1981) *Proteoglycans in Cell Biology of Extracellular Matrix*, ed. Hay, E. D. (Plenum, New York), pp. 39–63.
2. Heinegård, D. & Paulsson, M. (1984) in *Connective Tissue Biochemistry*, eds. Piez, K. A. & Reddi, A. H. (Elsevier, New York), pp. 277–328.
3. Scott, J. E. (1984) *Biochem. J.* **218**, 229–233.
4. Vogel, K. G., Paulsson, M. & Heinegård, D. (1984) *Biochem. J.* **223**, 587–597.
5. Rosenberg, L., Choi, H., Poole, A. R., Lewandowska, K. & Culp, L. (1986) *Proc. CIBA Found. Symp.*, in press.
6. Brennan, M. J., Oldberg, Å., Pierschbacher, M. D. & Ruoslahti, E. (1984) *J. Biol. Chem.* **259**, 13742–13750.
7. Glössl, J., Beck, M. & Kresse, H. (1984) *J. Biol. Chem.* **259**, 14144–14150.
8. Rosenberg, L. C., Choi, H. U., Tang, L.-H., Johnson, T. L., Pal, S., Webber, C., Reiner, A. & Poole, A. R. (1985) *J. Biol. Chem.* **260**, 6304–6313.
9. Bourdon, M. A., Oldberg, Å., Pierschbacher, M. & Ruoslahti, E. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1321–1325.
10. Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Fischer, E., Rutter, W. J. & Goodman, H. M. (1977) *Science* **196**, 1313–1319.
11. Singer, R. H. & Penman, S. (1973) *J. Mol. Biol.* **78**, 321–334.
12. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
13. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
14. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198.
15. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
16. Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) *Plasmid* **13**, 31–40.
17. Sanger, F., Nicklen, S. & Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
18. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–559.
19. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201–5205.
20. Blin, N. & Stafford, D. W. (1976) *Nucleic Acids Res.* **3**, 2303–2308.
21. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
22. Feinberg, A. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
23. Chopra, R. K., Pearson, C. H., Pringle, G. A., Fackre, D. S. & Scott, P. G. (1985) *Biochem J.* **232**, 277–279.
24. Watson, M. E. E. (1984) *Nucleic Acids Res.* **12**, 5145–5164.
25. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
26. von Heijne, G. (1984) *J. Mol. Biol.* **173**, 243–251.
27. Tosi, M., Young, R. A., Hagenbüchle, O. & Schibler, U. (1981) *Nucleic Acids Res.* **9**, 2313–2323.
28. Berget, S. M. (1984) *Nature (London)* **309**, 179–182.
29. McLauchlan, J., Gaffney, D., Whitton, J. L. & Clements, J. B. (1985) *Nucleic Acids Res.* **13**, 1347–1368.
30. Urano, Y., Watanabe, K., Sakai, M. & Tamaoki, T. (1986) *J. Biol. Chem.* **261**, 3244–3251.
31. Roden, L., Koerner, T., Olson, C. & Schwartz, N. B. (1985) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **44**, 373–380.
32. Pearson, C. H., Winterbottom, N., Fackre, D. S., Scott, P. G. & Carpenter, M. R. (1983) *J. Biol. Chem.* **258**, 15101–15104.
33. Bourdon, M. A., Shiga, M. & Ruoslahti, E. (1986) *J. Biol. Chem.*, in press.
34. Vogel, K. G. & Peterson, D. W. (1981) *J. Biol. Chem.* **256**, 13235–13242.
35. Damle, S. P., Cöster, L. & Gregory, J. D. (1982) *J. Biol. Chem.* **257**, 5523–5527.
36. Hassell, J. R., Newsome, D. A., Krachmer, J. H. & Rodrigues, M. M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3705–3709.
37. Fransson, L.-A., Cöster, L., Malmström, A. & Sheehan, J. K. (1982) *J. Biol. Chem.* **257**, 6333–6338.
38. Uldbjerg, N., Malmström, A., Ekman, G., Sheehan, J., Ulmsten, U. & Wingerup, L. (1983) *Biochem. J.* **209**, 497–503.
39. Oegema, T. R., Jr., Hascall, V. C. & Eisenstein, R. (1979) *J. Biol. Chem.* **254**, 1312–1318.
40. Salisbury, B. G. J. & Wagner, W. D. (1981) *J. Biol. Chem.* **256**, 8050–8057.
41. Fischer, L. W., Termine, J. D., Dejter, S. W., Jr., Whitson, S. W., Yanagishita, M., Kimura, J. H., Hascall, V. C., Kleinman, H. K., Hassell, J. R. & Nilsson, B. (1983) *J. Biol. Chem.* **258**, 6588–6594.