Expression of Collagen XVIII and Localization of Its Glycosaminoglycan Attachment Sites*

Received for publication, September 10, 2002, and in revised form, November 8, 2002
Published, JBC Papers in Press, November 13, 2002, DOI 10.1074/jbc.M209276200

Sucai Dong, Gregory J. Cole‡, and Willi Halfter§

From the Department of Neurobiology, University of Pittsburgh, Pittsburgh, Pennsylvania 15261 and  
Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University,  
Raleigh, North Carolina 27606

Collagen XVIII is the only currently known collagen that carries heparan sulfate glycosaminoglycan side chains. The number and location of the glycosaminoglycan attachment sites in the core protein were determined by eukaryotic expression of full-length chick collagen XVIII and site-directed mutagenesis. Three Ser-Gly consensus sequences carrying glycosaminoglycan side chains were detected in the middle and N-terminal part of the core protein. One of the Ser-Gly consensus sequences carried a heparan sulfate side chain, and the remaining two had mixed chondroitin and heparan sulfate side chains; thus, recombinant collagen XVIII was a hybrid of heparan sulfate and chondroitin proteoglycan. In contrast, collagen XVIII from all chick tissues so far assayed have exclusively heparan sulfate side chains, indicating that the posttranslational modification of proteins expressed in vitro is not entirely identical to the processing that occurs in a living embryo. Incubating the various mutated collagen XVIIIIs with retinal basement membranes showed that the heparan sulfate glycosaminoglycan side chains mediate the binding of collagen XVIII to basement membranes.

The collagens type XVIII and XV are members of the multiplexins, a collagen subfamily that is characterized by multiple alternating collagen and non-collagenous domains in the protein sequence (1, 2). Collagen XVIII came into the public spotlight by the discovery that the C-terminal peptide of collagen XVIII, named endostatin, has anti-angiogenic and anti-tumor activities (3, 4). The anti-angiogenic activity of the peptide led to the idea that collagen XVIII might be involved in the development of the vascular system. The targeted deletion of collagen XVIII in mice (5) and site-directed mutagenesis of proteins expressed in vitro is not entirely identical to the processing that occurs in a living embryo. Incubating the various mutated collagen XVIIIIs with retinal basement membranes showed that the heparan sulfate glycosaminoglycan side chains mediate the binding of collagen XVIII to basement membranes.

Experimental Procedures

Molecular Cloning of Chick Collagen XVIII DNA by Nested PCR—To obtain cDNA sequences covering the entire length of collagen XVIII, a 2.4-kb chick collagen XVIII 3' cDNA (p10d; Ref. 10) was extended by nested PCR from a random-primed E6 chick amnion library (α-ZapII; Stratagene, La Jolla, CA). The nucleotide sequence of the T3 or T7 promoter from the library vector was used as the forward primer, and two nucleotide sequences designed from the internal 5' cDNA sequences of p10d were used as the reverse and nested reverse primer. The first round of PCR contained 4 μl of supernatant of the boiled and centrifuged library (2 × 10^6 plaque-forming unit/ml) and elongase mix as the DNA polymerase (Invitrogen) in a total volume of 40 μl. Bands with the expected molecular weights were extracted from a 1% agarose gel and re-amplified in the second round of PCR with the T7 or T3 forward primer and the nested reverse primer. The second round of PCR was carried out using Pfu turbo DNA polymerase (Stratagene) following the recommendations provided by the manufacturer. The re-amplified DNA was extracted from the agarose gel, ligated into pPCR-Script AMP SK(+) vector, and transformed into Epicharact coli (Ep. coli)/XL 10-Gold ultracompetent cells using the PCR-Script AMP cloning kit (Stratagene). Plasmids from white colonies were purified and sequenced in the University of Pittsburgh Sequencing Facility. The sequences were analyzed using the Sequencer software (DNA Codes Inc., Ann Arbor, MI) and compared with published databases using the

* This work was supported by National Institutes of Health Grant NS33981-02. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF083440.

To whom correspondence should be addressed: Dept. of Neurobiology, University of Pittsburgh Medical School, E1402 BST, Pittsburgh PA 15261. Tel.: 412-648-9424; Fax: 412-648-1441; E-mail: whalf@pitt.edu.
Blast search algorithm (17). The amino acid sequence was deduced from the DNA sequence using Sequencher. Three clones were obtained (P537TP64, P217TP21, INT27ANT22) that covered the remaining collagen XVIII cDNA.

Molecular Cloning of Full-length of the Coding Sequence for Chick Collagen XVIII—The three PCR clones obtained by nested PCR and the previously obtained (10) were connected by PCR to obtain a cDNA for the full-length coding sequence of chick collagen XVIII. The reaction contained 0.07 pmol each of the DNA segments, a forward primer (5′-TTGGCAGTACTGCCGAGAACTGAG-3′) that contains an Nhel cutting site in addition to 3 nucleotides to protect the cutting site at 5′ end, and a reverse primer (5′-TTGGCAGTACTGCCGAGAACTGAG-3′) that contains a XhoI cutting site and 4 protection nucleotides at the 5′ end. The PCR was carried out using Pfu turbo DNA polymerase. The cycling condition was 94°C for 1 min and 72°C for 7 min for 2 cycles; 94°C for 1 min, 55°C for 1 min, and 72°C for 6 min for 2 cycles; and 94°C for 1 min and 72°C for 7 min for 30 cycles, following denaturing at 94°C for 5 min. A 4-kilobase band extracted from the 1% agarose gel ligated into pPCR-Script AMP SK(+) vector was transformed into E. coli XL1-10-Gold ultracompetent cells as described above. One of the clones (Full-12) contained the complete chick collagen XVIII-coding sequence, as confirmed by DNA sequencing.

Site-directed Mutagenesis—The oligonucleotides used to generate the mutant constructs are shown in Table 1. Site-directed mutagenesis (Stratagene). Briefly, a mutated duplicate of the original plasmid was produced by PCR using Pfu DNA polymerase and primers containing the desired mutation. After digestion of the original plasmid by DpnI, E. coli XL10-Gold ultracompetent cells were transformed with the nicked plasmid, which contains the desired mutation.

Construction, Expression, and Purification of Recombinant Collagen XVIII—Plasmid from Full-12 and the mutated Full-12 clones were digested with Nhel and Xhol (New England Biolabs, Beverly, MA) and de-phosphorylated with alkaline phosphatase (0.05 units/pmol ends; New England Biolabs). The digested and de-phosphorylated DNA inserts were isolated from 1% agarose gels and ligated into an E. coli pGEM-T vector using T4 DNA ligase. The DNA was transformed into E. coli DH5α (18). Several clones were obtained (Full-10, Full-11, and Full-12) that contained a 2400-nucleotide site-directed mutation in the predicted collagen XVIII gene.

TABLE I

| Names | Oligonucleotide sequence (from 5′ to 3′) |
|-------|---------------------------------------|
| Forward primer 2 | GAT GAC GCC GAG GCA GCT GGA GAT TTC G |
| Reverse primer 2 | CAT CAA TCT CAG CTT CCG CGT CAT C |
| Forward primer 3 | CAT CTG TCT GCC CGG CGG CGG AAC TTT CCA GAT T |
| Reverse primer 3 | CAT TCT TCC CCG CGG GAT GAT GAT GCC G |
| Forward primer 4 | GAC AGG CAT CTC CCA GCA GGG AAG A |
| Reverse primer 4 | CAT TCT TCC CCG CGG GAT GAT GAT GCC G |
| Forward primer 7 | CAT TGA CAT GGA GGG GCG GGG GGG CTT G |
| Reverse primer 7 | CAC CAC ACC AGG CAG CCA TCT CCA TCC TT |
| Forward primer 8 | GGT GAC ATG GAA GGC GCT GGG CTC GTG CCA CCT TT |
| Reverse primer 8 | GCA AGT GGC AGC CCA CCT TCC ATG TCA CC |

Enzyme Treatments—30 µl of Q-Sepharose-purified recombinant collagen XVIII and 30 µl of vitreous body supernatant were used for the digestion. For collagenase treatment, samples were dialyzed against sodium phosphate buffer (50 mM) containing 0.15 M NaCl (phosphate-buffered saline), pH 7.4, and incubated with 5 milliunits/µl collagenase (Sigma type VII) and 2% bovine serum albumin at 37°C for 1 h. For heat inactivation, the samples were dialyzed against phosphate-buffered saline, pH 7.0, containing 3 mM CaCl2 and treated with 0.125 milliunits/µl heparinase (Seikagaku, Rockville, MA) at 37°C for 2 h. For chondroitinase treatment, the samples were dialyzed against phosphate-buffered saline, pH 8.0, and treated with 0.131 milliunits/µl chondroitinase ABC (Seikagaku) at 37°C for 2 h. For mixed digestion, the samples were dialyzed against phosphate-buffered saline, pH 7.5, containing 3 mM CaCl2 and treated with a mixture of heparinase (0.125 milliunits/µl) and chondroitinase ABC (0.131 milliunits/µl).

Northern Blots—Northern blotting was performed as described previously (19). Briefly, total RNA was isolated from E7–E10 chick brain, embryonic liver, and kidney using TRIzol Reagent (Invitrogen). Poly(A)−RNA was purified using PolyATrac mRNA isolation system IV (Promega, Madison WI), and samples of 2 µg of poly(A)+ RNA were separated in 1% agarose gels containing 2.2% formaldehyde, transferred to positively charged nylon membranes (Immobilon N Millipore), and cross-linked with UV. Digoxigenin-labeled cRNA hybridization probe was synthesized from the linearized P127TP21 collagen XVIII clone template using the RNA-polymerase labeling kit (Roche Molecular Biochemicals). The digoxigenin-labeled cRNA hybridized blot was visualized using alkaline phosphatase-conjugated anti-digoxigenin Fab fragment with CSPD® ready-to-use chemiluminescent substrates (Roche Molecular Biochemicals). RNA molecular weight marker II (Roche Molecular Biochemicals) was used to determine the size of the collagen XVIII mRNA.

Binding of Collagen XVIII to Retinal Basement Membranes—A PAPPen (EMS, Washington, PA) was used to circle an ~1 cm2 area of a plastic dish coated with nitrocellulose (20). The hydrophobic ring reduces the volume of the incubation solutions to less than 100 µl. Retinal embryonic liver, and kidney were incubated in calcium- and magnesium-free Hank’s solution, the preparations were incubated with a monoclonal antibody to the Myc epitope (Santa Cruz Biotechnology, Santa Cruz, CA) for 1 h, washed, and visualized with a Cy-3-labeled goat-anti-mouse secondary antibody. Binding of collagen XVIII to tissues was done by

GATT Attachment Sites of Collagen XVIII

Mismatches with the template are indicated by boldface letters. Two single mutants, M2 and M7, were produced using Full-12 as the template with forward and reverse primers 2 and 7, respectively. The PCR system for double mutant DM78 contained M7 as the template and forward and reverse primers 2, 3, and 4, respectively.

SDS-PAGE and Western Blotting—Vitreous body was collected from E5–E7 chick eyes and centrifuged at 15,000 rpm for 5 min. Recombinant collagen XVIII was purified by Q-Sepharose chromatography as described above and dialyzed against calcium- and magnesium-free Hank’s solution. The samples were subjected to 3.5–15% SDS-PAGE under reducing conditions. Some of the samples were also run under non-reducing conditions and without boiling the sample. The proteins were visualized using alkaline phosphatase-conjugated goat anti-mouse IgG (Jackson Immunoresearch, West Grove PA) and 4-nitro blue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (Roche Molecular Biochemicals) as coloring reagents.

E5–E7 chick eyes and centrifuged at 15,000 rpm for 5 min. Recombinant collagen XVIII was purified by Q-Sepharose chromatography as described above and dialyzed against calcium- and magnesium-free Hanks’ solution. The samples were subjected to 3.5–15% SDS-PAGE under reducing conditions. Some of the samples were also run under non-reducing conditions and without boiling the sample. The proteins were visualized using alkaline phosphatase-conjugated goat anti-mouse IgG (Jackson Immunoresearch, West Grove PA) and 4-nitro blue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (Roche Molecular Biochemicals) as coloring reagents.
incubating sections of parafomaldehyde-fixed heads of the E6 chick embryo with the cell culture supernatants.

RESULTS

Extension of the cDNA Sequence of Chick Collagen XVIII by Nested PCR—A cDNA clone (p10d) for chick collagen XVIII was previously obtained by conventional screening of an E5 chick yolk sac library with the 6C4 monoclonal antibody. It covers 2.4 kb of the 3' end of collagen XVIII with 1.24 kb of the 3'-untranslated region (10). To extend the sequence in 5' direction, we sequentially amplified three collagen XVIII cDNAs from a randomly primed amnion library using a PCR procedure. Taking the T3 or T7 primer of the library vector as a forward primer and a specific reverse primer from the 5' end of p10d, we amplified two DNA bands in the first round of PCR with the T3 primer and five bands with the T7 primer. In the second round of PCR using the nested reverse primer, only two DNA bands were obtained with the T7 primer. These bands were slightly smaller in size than the DNA bands from the first round of PCR, as expected from the close proximity of the specific reverse primer and the nested reverse primer. One of the DNA sequences, named INT27ANTI22, was homologous to human and mouse collagen XVIII cDNA. The INT27ANTI22 clone contained a 1270-bp insert and overlapped with p10d by 150 bp. We repeated the procedure using a specific reverse primer and a nested reverse primer designed from INT27ANTI22 to obtain P127TP21 and then from P127TP21 to obtain P537TP64. Both P127TP21 and P537TP64 overlapped with their respective 3' predecessors (Fig. 1). The deduced amino acid sequence of P537TP64 contained the starting Met codon and a 26-residue signal peptide.

The Complete cDNA Sequence of Chick Collagen XVIII—A continuous chick collagen XVIII cDNA sequence was obtained by the alignment of P537TP64, P127TP21, INT27ANTI22, and p10d. The sequence contained 5279 bp with an open reading frame of 4032 bp (Fig. 1) that corresponded to the short version of human and mouse collagen XVIII α1. The sequence is deposited in the NCBI gene bank under the accession number AF083440.

The deduced protein sequence from the cDNA of chick collagen XVIII α1 chain is 1344 amino acids long, with the first 26 amino acid residues comprising the signal peptide. The chick sequence is slightly longer than the human (1336) and mouse (1315) short variants (Ref. 21; Fig. 1). Between chick and human, the overall identity is 61%, and the homology is 71%, lower than those between human and mouse of 75 and 95%, respectively. Like the human counterpart, chick collagen XVIII consists of 11 non-collagenous (NC) domains, which are separated by 10 collagenous (COL) domains. The N-terminal COL-10 is 71 amino acid residues long with 2 imperfections in the GXY repeats, whereas the corresponding COL-10 in human collagen XVIII is 25 residues (21). Six cysteine residues are present in chick collagen XVIII, two of which are located in NC11, and four are located in the endostatin domain of NC1. The presence of cysteine residues was conserved between chick and human. A potential N-linked glycosylation site located in COL8 was conserved between chick and human. Another potential N-linked glycosylation site, present at the beginning of the NC11, was not conserved. There are eight potential GAG glycosylation sites in chick collagen XVIII, three of which are in fact GAG attachment sites (see below).

In human and mouse tissues, the mRNAs of collagen XVIII α1 exist in a short and a long splice variant, with the long version occurring prominently in liver (1, 9, 21). Attempts to obtain the long version by rapid amplification of cDNA 5' ends using poly(A)+ RNAs from chick E6–7 liver and kidney tissues failed (data not shown). The absence of the long collagen XVIII splice variant in chick was confirmed by Northern blot analysis: poly(A)+ RNAs from E6–7 chick liver, kidney, heart, and brain probed with a cRNA probe synthesized from the 5' clone P127TP21 (see Fig. 1) showed a single band of 5.5 kb (Fig. 2). The finding was consistent with the previous Northern blots probed with the 3' p10d cRNA probe (see Fig. 1), that also showed a single band at 5.5 kb (10).

Expression and Characterization of Recombinant Chick Collagen XVIII—Using P537TP64, P127TP21, INT27ANTI22, and p10d as templates and nucleotide sequences from the 5' and 3' end of the collagen XVIII-coding sequence as primers, we obtained a continuous cDNA for collagen XVIII by PCR and cloned the cDNA into an expression vector. The expression construct was stably transfected into EBNA T293 cells, and the expressed protein was isolated from the cell culture supernatant using Q-Sepharose. The fraction eluted with 0.5 M NaCl contained non-glycosylated collagen XVIII core protein with an apparent molecular mass of 187 kDa. The GAG-free core protein accounted for less than 5% of total expressed collagen XVIII. The high salt fraction eluted from the ion exchanger consisted of 11 non-collagenous (NC) domains, which are separated by 10 collagenous (COL) domains. The N-terminal COL-10 is 71 amino acid residues long with 2 imperfections in the GXY repeats, whereas the corresponding COL-10 in human collagen XVIII is 25 residues (21). Six cysteine residues are present in chick collagen XVIII, two of which are located in NC11, and four are located in the endostatin domain of NC1. The presence of cysteine residues was conserved between chick and human. A potential N-linked glycosylation site located in COL8 was conserved between chick and human. Another potential N-linked glycosylation site, present at the beginning of the NC11, was not conserved. There are eight potential GAG glycosylation sites in chick collagen XVIII, three of which are in fact GAG attachment sites (see below).

In human and mouse tissues, the mRNAs of collagen XVIII α1 exist in a short and a long splice variant, with the long version occurring prominently in liver (1, 9, 21). Attempts to obtain the long version by rapid amplification of cDNA 5' ends using poly(A)+ RNAs from chick E6–7 liver and kidney tissues failed (data not shown). The absence of the long collagen XVIII splice variant in chick was confirmed by Northern blot analysis: poly(A)+ RNAs from E6–7 chick liver, kidney, heart, and brain probed with a cRNA probe synthesized from the 5' clone P127TP21 (see Fig. 1) showed a single band of 5.5 kb (Fig. 2). The finding was consistent with the previous Northern blots probed with the 3' p10d cRNA probe (see Fig. 1), that also showed a single band at 5.5 kb (10).

Expression and Characterization of Recombinant Chick Collagen XVIII—Using P537TP64, P127TP21, INT27ANTI22, and p10d as templates and nucleotide sequences from the 5' and 3' end of the collagen XVIII-coding sequence as primers, we obtained a continuous cDNA for collagen XVIII by PCR and cloned the cDNA into an expression vector. The expression construct was stably transfected into EBNA T293 cells, and the expressed protein was isolated from the cell culture supernatant using Q-Sepharose. The fraction eluted with 0.5 M NaCl contained non-glycosylated collagen XVIII core protein with an apparent molecular mass of 187 kDa. The GAG-free core protein accounted for less than 5% of total expressed collagen XVIII. The high salt fraction eluted from the ion exchanger consisted of 11 non-collagenous (NC) domains, which are separated by 10 collagenous (COL) domains. The N-terminal COL-10 is 71 amino acid residues long with 2 imperfections in the GXY repeats, whereas the corresponding COL-10 in human collagen XVIII is 25 residues (21). Six cysteine residues are present in chick collagen XVIII, two of which are located in NC11, and four are located in the endostatin domain of NC1. The presence of cysteine residues was conserved between chick and human. A potential N-linked glycosylation site located in COL8 was conserved between chick and human. Another potential N-linked glycosylation site, present at the beginning of the NC11, was not conserved. There are eight potential GAG glycosylation sites in chick collagen XVIII, three of which are in fact GAG attachment sites (see below).

In human and mouse tissues, the mRNAs of collagen XVIII α1 exist in a short and a long splice variant, with the long version occurring prominently in liver (1, 9, 21). Attempts to obtain the long version by rapid amplification of cDNA 5' ends using poly(A)+ RNAs from chick E6–7 liver and kidney tissues failed (data not shown). The absence of the long collagen XVIII splice variant in chick was confirmed by Northern blot analysis: poly(A)+ RNAs from E6–7 chick liver, kidney, heart, and brain probed with a cRNA probe synthesized from the 5' clone P127TP21 (see Fig. 1) showed a single band of 5.5 kb (Fig. 2). The finding was consistent with the previous Northern blots probed with the 3' p10d cRNA probe (see Fig. 1), that also showed a single band at 5.5 kb (10).
represents the homotrimer of the collagen XVIII/H9251 collagen XVIII from vitreous body (Fig. 3, lane 2), with a molecular mass of about 350 kDa. After treatment with collagenase, the protein was degraded to a band with a molecular mass around 35 kDa (lane 4). When vitreous body samples (lanes 1 and 2) were probed with 6C4; lanes 5 and 6 were probed with an anti-tenascin antibody) were treated with collagenase (lanes 2 and 6), the band for collagen XVIII shifted to 35 kDa (compare lane 1 with lane 2), and the band of tenascin did not shift relative to the band of the untreated sample (compare lane 5 with lane 6), showing that collagenase was free of nonspecific proteases.

Collagen XVIII expressed in EBNA cells was proteolytically very sensitive. The degradation fragments on the gels ranged from 51 to 75 kDa (Fig. 3, lanes 3 and 4, lane 5). Based on their reactivity with the 6C4 monoclonal antibody, they were from the C-terminal part of the molecule. To prevent degradation, we added a mixture of commercially available protease inhibitors including 3 μM aprotinin, 50 μM cathepsin B inhibitor II (Calbiochem), 100 μM E-64, 50 μM leupeptin, and 100 μM pepstatin into cell culture medium and the solutions used for Q-Sepharose chromatography. The presence of inhibitors resulted in slightly different degradation fragments; however, the average intensity and size of the bands were roughly the same as those from collagen XVIII expressed without protease inhibitors present (data not shown). Cathepsin B inhibitor II and E64, which have been reported to inhibit the release of endostatin at the C terminus (27), was not effective in preventing degradation of the whole molecule. In addition, the results showed that the yield of collagen XVIII increased with longer incubation periods; however, at the same time, the degradation of collagen XVIII increased even more. As a compromise, we collected the culture supernatant for purification in 24-h intervals.

**GAGs and Their Attachment Sites of the Recombinant Collagen XVIII**—Similar to the highly glycosylated collagen XVIII from vitreous, amnion, kidney, and meninges (10), recombinant collagen XVIII appeared in Western blots as a smear with an apparent molecular mass of 350 kDa (Fig. 5A, lane 1). Treatment with heparitinase caused a slight decrease in size to 320 kDa and the appearance of a faint band in the size of the core protein at 187 kDa (Fig. 5A, lane 2). When treated with chon-
droitinase ABC, the size of protein shifted to 240 kDa and the core protein band of 187 kDa became more prominent (Fig. 5A, lane 4). When treated with a mixture of heparitinase and chondroitinase, the recombinant protein appeared as two sharp, overlapping bands with molecular mass round 187 kDa, showing that the recombinant collagen XVIII expressed in 293EBNA cells is a hybrid HSPG/chondroitin sulfate proteoglycan (Fig. 5A, lane 3). An identical GAG composition of HS/CS was also detected with recombinant collagen XVIII expressed in chick meningeal cells (data not shown). To compare GAG composition of recombinant with that of the natural collagen XVIII, we treated vitreous body collagen XVIII with the same enzymes. The results confirmed previous experiments (10, 19), showing that collagen XVIII from all chick tissues so far tested showed a smear with an average molecular mass of 280 kDa (Fig. 8A, lanes 2, 4) but not by chondroitinase alone (Fig. 8B, lane 4). The collagen XVIII core protein band is indicated by an arrowhead.

Recombinant collagen XVIII has eight potential GAG glycosylation sites, five of which are located in NC-11. Other GAG attachment sites are listed in COL-10, NC-9, and NC-8. The sequences of potential GAG attachment sites are listed in Fig. 6. Sites 2, 3, and 4 appear as a group of three SG residues separated by 3 and 7 residues, respectively. Sites 3, 7, and 8 are conserved between chick, Xenopus (22), mouse, and human (Fig. 1). To reveal which site is in fact glycosylated, we used site-directed mutagenesis and expressed the mutated protein in EBNA cells. TM278, a triple mutant in which the serine residues in sites 2, 7, and 8 were substituted by alanines, appeared as a non-glycosylated protein of 187 kDa (Fig. 7, lane 1), showing that the glycosylation sites are located at these 3 sites. A double mutant DM78, in which the serine residues of sites 2 and 8 were substituted by alanines, showed a smear with an average molecular mass of 290 kDa (Fig. 7, lane 3). The remaining GAG (site 7) in the DM28 mutant collagen XVIII was completely digested with the mixture of heparitinase and chondroitinase ABC (Fig. 8B, lane 3) and partially digested with either heparitinase (Fig. 8B, lane 2) or chondroitinase (Fig. 8B, lane 4), suggesting that the GAG attached at site 7 is a chondroitin sulfate or heparan sulfate GAG. The abundance of the core protein after heparitinase or chondroitinase ABC treatment (Fig. 8B, lanes 2 and 4) showed the major proportion of GAGs attached to site 7 is chondroitin sulfate. The double mutant DM27, in which the serine residues of sites 2 and 7 were substituted with alanine, showed a smear with an average molecular mass of 150 kDa (Fig. 8A, lanes 2 and 4) and could only be completely digested with both enzymes (Fig. 8C, lane 3), suggesting that the GAG attached on site 8 is either chondroitin sulfate or heparan sulfate. Based on the gel pattern of lanes 2 and 4 in Fig. 8C, the dominant GAG at site 8 is chondroitin sulfate.

The observation that EBNA cells secrete more collagen XVIII core protein when transfected with doubly or triply mutated as compared with singly or non-mutated collagen XVIII cDNAs indicates that not all of the three attachment sites were occupied with GAG side chains in every collagen XVIII molecule. This would also explain that core protein bands appeared in all of the mutant samples in Fig. 7. The fact that the collagen XVIII core protein appeared after heparitinase (Fig. 5A, lane 2) and chondroitinase ABC treatment (Fig. 5A, lane 4) suggests that some of molecules contain only chondroitin sulfate GAG or only heparan sulfate GAG.

Potential glycosylation sites 2, 3, and 4 are present as three SGs following each other in short sequence. SG residues of site 2 and 3 are separated by three residues (DFG; Fig. 6), and the SG residues of site 3 and 4 are separated by seven residues (AGDRHHHP; Fig. 6). The results presented above showed that the serine at site 2 was glycosylated, and the GAG chain attached to this site was heparan sulfate. To investigate whether the serine residues in sites 3 and 4 has an effect on glycosylation of serine in site 2, we substituted the serine residues in sites 3, 7, and 8 to alanines (TM378) and serine residues in sites 4, 7, and 8 to alanines (TM478). TM478 ap-
The protein band is indicated by an arrowhead.

A mixture of heparitinase and chondroitinase (lane 3) and a broad band with a molecular mass around 320 kDa (lane 4). Untreated DM78 appeared as a broad band with a molecular mass around 240 kDa (lane 1). The protein was reduced to 187 kDa, the size of the core protein, after treatment with heparitinase (lane 2) or the enzyme mixture (lane 3) but was not affected by treatment with chondroitinase (lane 4). Untreated DM28 and DM27 appeared as broad bands with a molecular mass around 320 kDa (B and C, lanes 1). The sizes of the proteins were only reduced to a sharp band of 187 kDa and slightly and led to a minor increase in core protein (B and C, lanes 2). Treatment with heparitinase shifted the molecular weight only slightly and to a minor increase in core protein (B and C, lanes 2). Chondroitinase digestion led to a more prominent shift and a major increase of core protein (B and C, lanes 4). The collagen XVIII core protein band is indicated by an arrowhead.

Eight potential GAG glycosylation sites were located in the chick collagen XVIII. It is well established that collagen XVIII is a — and 5′ sensitive to degradation. We also have preliminary data showing that collagen XVIII from vitreous body and the meninges is highly expressed in liver. Northern blots with both 3′ and 5′ probes and 5′ rapid amplification of cDNA 5′ ends did not reveal a long version of collagen XVIII in chick tissues. We speculate that the chick has only the short version of the protein.

GAG Glycosylation of Collagen XVIII—Eight potential GAG attachment sites were located in the chick collagen XVIII.

DISCUSSION

Cloning of Chick Collagen XVIII by PCR—For cloning of the entire collagen XVIII we amplified the missing segments from a random-primed cDNA library by extending a previously isolated 3′ sequence (p10d) (10). The T3 or T7 promoter sequence, part of the vector, was used as the forward primer. The only requirement for the PCR extension was to design a specific and a nested reverse primer from the known collagen XVIII cDNA. The method turned out to be a very reliable alternative to the more elaborate 5′ rapid amplification of cDNA 5′ ends. Furthermore, it not only was successfully used to obtain the 5′ half of collagen XVIII but also allowed us to extend the extremely long sequence of laminin α1 from the very 5′ end in the 3′ direction. Thus, the method was applicable to extend DNA sequences from any location of the molecule in both directions.

Chick Collagen XVIII—The protein sequence of the chick collagen XVIII showed a high homology to human and mouse collagen XVIII with all 10 collagenous and all 11 non-collagenous domains preserved. The most conserved regions were in the C-terminal NC1 domain, with 90% homology in the endostatin region. Recombinant chick collagen XVIII was more sensitive to proteolytic degradation than the endogenous one in vitreous body. The reason, we speculate, is that triple helix formation is incomplete, as shown by presence of the collagen XVIII monomers in non-boiled non-reduced samples (Fig. 3). The dominant fragment was from the C-terminal part of the protein, suggesting that collagen XVIII has unique features at its C terminus, NC1, part of the molecule making it very sensitive to degradation. We also have preliminary data showing that collagen XVIII from vitreous body and the meninges also has its C-terminal part clipped off, and a naturally existing endostatin-like peptide was identified in vitreous body. 2 Collagen XVIII in human and mouse exists in a short and a long version (1, 9, 21), whereby the long version is highly expressed in liver. Northern blots with both 3′ and 5′ probes and 5′ rapid amplification of cDNA 5′ ends did not reveal a long version of collagen XVIII in chick tissues. We speculate that the chick has only the short version of the protein.

S. Dong and W. Halfter, unpublished.
but not the relatively distant serine at site 4. The diagrams show the location of the mutations (stars) and the glycosylated site 2 in TM378, TM478, and DM78 (circles). DM78 (A, lane 4) showed a smear of 235 kDa. The GAG on site 2 in DM78 was shown to be HS (see Figs. 7 and 8). TM478 (A, lane 3), derived from DM78 by substitution of serine on site 4 with alanine, showed a banding pattern similar to DM78. TM378 (A, lane 2), derived from DM78 by substitution of serine on site 3 by alanine, appeared as a smear of 290 kDa. The GAGs on site 2 in TM378 (A, lane 2) can be partially digested by heparitinase (B, lane 1) and chondroitinase (B, lane 2) and entirely digested by a mixture of the two enzymes (B, lane 3); compare with the banding pattern of the triple mutant TM278 with all GAG attachment sites removed, as shown in A, lane 1). The collagen XVIII core protein band is indicated by an arrowhead.

Three of these sites carry GAGs, confirming that collagen XVIII is a proteoglycan (10). All three attachment sites and their adjacent amino acids were conserved in human, mouse, and Xenopus collagen XVIII (22), consistent with the notion that collagen XVIIIIs from other species are also proteoglycans (21). The chick SG consensus sites 1–6 were reduced in human and mouse to 1 single site, suggesting an evolutionary pressure to maintain at least one SG site at this position of the protein. The recombinant chick collagen XVIII expressed in either human 293-EBNA cells or chick meningeal cells turned out to be a hybrid chondroitin sulfate proteoglycan/HSPG, in contrast to the endogenous collagen XVIII from chick vitreous body, amnion, kidney, and meninges, which has only heparan sulfate side chains. Obviously, the in vitro expression of the protein does not entirely recapitulate the normal posttranslational modifications that occur in the chick embryo, and the unusual posttranslational modification seems irrelevant to the species from which the cells derived.

Previous studies show that the GAGs are connected to the core proteins via SG consensus sequences. By comparing peptide sequences close to GAG attachment (16, 17), it was found that the presence of acidic amino acids before or after the SG sites enhances the chance of a protein being glycosylated (16). Furthermore, multiple SGs in short sequence increase the chance of a site becoming connected with a heparan sulfate side chain (17). Recent studies, however, showed that the clear-cut identification of an SG consensus sequence as a glycosylation site or the prediction of the type of glycosylation at a specific site is impossible. Some peptide domains distant to the SG sites were shown also to be important in the glycosylation (23, 24). Furthermore, the fact that endogenous and recombinant chick collagen XVIII is different in its GAG glycosylation suggests that additional regulatory factors such as the cell type in which the protein is expressed and cell culture conditions are also important for glycosylation.

All three GAG attachment sites identified in collagen XVIII fulfill the requirements postulated for GAG glycosylation, as they have acidic residues ahead of the SG consensus sequences, and in one case (site 2), three SGs follow in short distance to each other. Indeed, site 2 is connected to the HS GAG, and in one case (site 3), the mutation of the SG at site 3 converts the HS into a CS side chain. However, sites 1, 3, and 4 also fulfill the requirements for GAG attachment sites, yet they are not glycosylated. Additional factors that turned out to be important for glycosylation were (a) the cell type expressing the proteins and (b) the growth condition of the cells. Collagen XVIII from vitreous body, amnion, the meninges, and kidney is a HSPG and has no CS side chains, in contrast to the recombinant protein in which 2 of the 3 GAG side chains are substituted for CS. It could mean that cells from the ciliary body, amnion, kidney, and the meninges, all of which express collagen XVIII (10, 26), are capable of producing the fully heparan sulfate-glycosylated protein, whereas the 293-EBNA cells cannot. This is probably not the case, since chick meningeal cells and 293-EBNA cells produce the same incorrect GAGs in vitro. The different glycosylation in vivo and in vitro could mean that growth conditions of cells in a living organism promote the glycosylation in a way that has not been reproduced in vitro. It was remarkable that the presence of fetal calf serum in the culture medium had a strong influence on the properties of the GAG chains, and it is conceivable that growth factors or signaling molecules are required to promote the correct glycosylation. The most reliable predictor for GAG glycosylation in collagen XVIII was the conservation of the SG consensus sites in different species. All three sites with GAGs were conserved.
between human, mouse, *Xenopus*, and chick, and an obvious ID for a GAG attachment site is the conservation of the site throughout different species.

REFERENCES
1. Oh, S. P., Kamagata, Y., Muragaki, Y., Timmons, S., Ooshima, A., and Olsen, B. R. (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 4229–4233
2. Rehn, M., and Pihlajaniemi, T. (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 4232–4238
3. O’Reilly, M. S., Boehm, T., Shing, Y., Fukai, N., Vazios, G., Lane, W. S., Flynn, E., Burkhead, J. R., Olsen, B. R., and Folkman, J. (1997) *Cell* 88, 277–285
4. Sasaki, T., Fukai, N., Mann, K., Goering, W., Olsen, B. A., and Timpl, R. (1998) *EMBO J.* 17, 4249–4256
5. Fukai, N., Eklund, L., Marneros, A. G., Oh, S. P., Keene, D. R., Tamarkin, L., Niemela, M., Lives, M., Li, E., Pihlajaniemi, M., and Olsen, B. (2002) *EMBO J.* 21, 1535–1544
6. Sertie, A. L., Sass, V., Camargo, A. A., Zatz, M., Brabe, C., and Passos-Bueno, M. R. (2000) *Hum. Mol. Genet.* 9, 2051–2058
7. Hynes, R. O., and Zhao, Q. (2000) *J. Cell Biol.* 150, 89–95
8. Ackley, B. D., Crew, J. R., Elamaa, H., Pihlajaniemi, T., Kuo, C. J., and Kramer, J. M. (2001) *J. Cell Biol.* 152, 1219–1222
9. Saarela, J., Rehn, M., Oikarinen, A., Autio-Harmainen, H., and Pihlajaniemi, T. (1998) *Am. J. Pathol.* 153, 611–626
10. Halfter, W., Dong, S., Schurer, B., and Cole, G. J. (1998) *J. Biol. Chem.* 273, 25484–25492