Characterization of Epiphycan, a Small Proteoglycan with a Leucine-rich Repeat Core Protein*

(Received for publication, April 3, 1997, and in revised form, May 5, 1997)

H. Jan Johnson‡, Lawrence Rosenberg§, Haing U. Choi§, Sonya Garza‡, Magnus Höök‡, and Peter J. Neame¶

From the ‡Center for Extracellular Matrix Biology, Albert B. Alkek Institute of Biosciences and Technology and the Department of Biochemistry and Biophysics, Texas A & M University, Houston, Texas 77090, the ¶Orthopedic Research Laboratories, Montefiore Medical Center, Bronx, New York 10467, and the §Shriners Hospital for Children and the Department of Biochemistry and Molecular Biology, University of South Florida, Tampa, Florida 33612

The epiphysis of developing bones is a cartilaginous structure that is eventually replaced by bone during skeletal maturation. We have separated a dermalan sulfate proteoglycan, epiphycan, from decorin and biglycan by using dissociative extraction of bovine fetal epiphyseal cartilage, followed by sequential ion-exchange, gel permeation, hydrophobic, and Zn²⁺ chelate chromatographic steps. Epiphycan is a member of the small leucine-rich proteoglycan family, contains seven leucine-rich repeats (LRRs), is related to osteoglycin (osteointuctive factor) (Bentz, H., Nathan, R. M., Rosen, D. M., Armstrong, R. M., Thompson, A. Y., Segarini, P. R., Mathews, M. C., Dasch, J., Piez, K. A., and Seyedin, S. M. (1989) J. Biol. Chem. 264, 20805–20810), and appears to be the bovine equivalent of the chick proteoglycan PG-Lb (Shinomura, T., and Kimata, K. (1992) J. Biol. Chem. 267, 1265–1270). The intact proteoglycan had a median size of −133 kDa. The core protein was 46 kDa by electrophoretic analysis, had a calculated size of 34,271 Da, and had two approximately equimolar N termini (APTLS... and ETYDAT...) separated by 11 amino acids. There were at least three O-linked oligosaccharides in the N-terminal region of the protein, based on blank cycles in Edman degradation and corresponding serine or threonine residues in the translated cDNA sequence. The glycosaminoglycans ranged in size from 23 to 34 kDa were more heterogeneous than those in other dermalan sulfate small leucine-rich proteoglycans and were found in the acidic N-terminal region of the protein core, N-terminal to the LRRs. A four-cysteine cluster was present at the N terminus of the LRRs, and a disulfide-bonded cysteine pair was present at the C terminus of the protein core. The seventh LRR and an N-linked oligosaccharide were between the two C-terminal cysteines. An additional potential N-glycosylation site near the C terminus did not appear to be substituted at a significant level.

Cartilage contains a variety of proteoglycans. Particularly abundant are aggrecan and the small leucine-rich proteoglycans (SLRPs)¹ (1) fibromodulin and decorin. Several other proteoglycans in cartilage have also been identified, including versican, perlecan, and the leucine-rich proteoglycans lumican and biglycan. The relative abundance of these proteoglycans varies during development and by location within the tissue. It is likely that this variation has a role in the differentiation and maintenance of tissue structure. The exact roles of the SLRPs are unclear at present, but it is thought that fibromodulin and decorin are involved in the process of collagen fibrillogenesis (2, 3) and may play a crucial role in optimizing the diameter of collagen fibrils that will eventually be replaced during remodeling of the cartilage during calcification. It is possible that the SLRPs also have a role in regulating growth factors, e.g. transforming growth factor-β, which binds to decorin (4, 5).

A hallmark of the SLRPs are the cysteine clusters that flank the leucine-rich repeats. These cysteine form disulfide bonds and perhaps provide a structure that differs from leucine-rich repeat (LRR)-containing proteins that do not contain this feature. The LRR motif was first identified by Patthy (6) and is characterized by an LXXLXXNXXL sequence, where X is any amino acid and L is often a leucine, but may be any amino acid with a hydrophobic aliphatic side chain (Ile, Val, and Met). The LRR motif is conserved throughout evolution, and the increasing number of members of this family includes a range of proteins with diverse functions and distributions (reviewed by Kobe (7)). The three-dimensional structure of the porcine ribonuclease inhibitor, a member of the leucine-rich protein family, has been determined (8). In this protein, the 15 individual LRRs adopt a stacked β-sheet/α-helix hairpin structure, resulting in an overall horseshoe shape and indicating that the LRR is likely to be primarily a folding motif and not an indicator of function.

By using a novel purification protocol, we have isolated milligram quantities of an LRR-containing proteoglycan from bovine fetal epiphysis and determined its primary structure. We have named this proteoglycan epiphycan based on its isolation from the epiphysis and show it to be a mammalian homolog of the avian proteoglycan PG-Lb, isolated from developing chick limb (9). The amino acid sequence is very similar to a recently published mouse cDNA-derived sequence (10). Unlike other cartilage-derived LRR-containing proteoglycans and glycoproteins, epiphycan contains only seven LRRs instead of the more usual 10 or 11. We show that all of the glycosylation, with the

* This work was supported by National Institutes of Health Grants AR35322 (to P. J. N.); AR42919 (to M. H.), and AR21498 (to L. R.) and by a grant from the Shriners of North America (to P. J. N.). 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) U77127. To whom correspondence should be addressed: Shriners Hospital for Children, 12502 N. Pine Dr., Tampa, FL 33612. Tel.: 813-975-7146; Fax: 813-978-9442; E-mail: pneame@com1.med.usf.edu.

¶ This paper is available on line at http://www.jbc.org

¹ The abbreviations used are: SLRPs, small leucine-rich proteoglycans; LRR, leucine-rich repeat; PG, proteoglycan; GdnHCl, guanidine hydrochloride; PCR, polymerase chain reaction; PAGE, polyacrylamide gel electrophoresis; HPLC, high performance liquid chromatography; GAG, glycosaminoglycan; CHAPS, 3-(3-cholamidopropyl)dimethylammonio)-1-propanesulfonic acid.
exception of one N-linked oligosaccharide, occurs in the N-terminal domain.

EXPERIMENTAL PROCEDURES

Materials—Guanidine hydrochloride (GdnHCl) was from Research Plus Laboratories. Phenylmethylsulfonyl fluoride, iodoacetamide, sodium citrate, and sodium chloride were from Sigma. Pepstatin A was from Calbiochem. Sepharose CL-4B, octyl-Sepharose, chelating Sepha-rose Fast Flow, DEAE-Sepharose, Q-Sepharose, Sepharose PD10, Su-perose 6, and Superdex 75 10/30 columns were from Pharmacia Biotech Inc. Radioisotopes (125I iodine and Na125I) were supplied by DuPont or by Amersham Life Science, Inc. Toy polymerase, dNTPs, and restriction enzymes were supplied by Life Technologies, Inc. Chondroitinase ABC was supplied by Seikagaku. PCR primers were purchased from Life Technologies, Inc., National Biosciences (Plymouth, MN), or the Texas A & M Core Facility (College Station, TX); were synthesized in-house (Shriners Hospital for Children, Tampa, FL); or were gifts from Michelle Deere. 32P-Labeled molecular mass markers were purchased from Amersham Corp. The pCR1 plasmid (TA cloning) was purchased from Invitrogen. Wizard™ PCR prep was purchased from Promega. Sequencing was performed with Sequenase Version 2.0 (U. S. Biochemical Corp.) or by the Automated DNA Sequencing Facility at the Interdisciplinary Center for Biotechnology Research at the University of Florida (Gainesville, FL). ProSpin™ centrifugal filters were purchased from Applied Biosystems/Perkin-Elmer (Foster City, CA). The plBluescript plasmid was obtained from Stratagene. Trypsin (sequencing-grade) and endoprotease Lys-C were purchased from Boehringer Mannheim. Zeta-probe GT membrane was purchased from Bio-Rad.

Isolation of Epiphycan from Bovine Fetal Epiphyseal Cartilage—Dermatan sulfate proteoglycans were isolated from bovine fetal epiphyseal cartilage, as described previously for skin and bovine articular cartilage, by ion-exchange followed by gel permeation chromatography (11, 12).

Octyl-Sepharose Chromatography—The DEAE-bound and eluted proteoglycan-containing fractions from the gel permeation chromatography step were applied to an octyl-Sepharose column that had been equilibrated with 2 mM GdnHCl and 0.15 M sodium acetate, pH 6.3, at 25 °C and allowed to bind for 2 h. The column was washed with 3 volumes of 2 M GdnHCl, and the proteoglycans were eluted with a linear gradient of 2–6 M GdnHCl. Fractions were analyzed for uronate, revealing the presence of two major peaks. Fractions containing both epiphycan and decorin were pooled; concentrated to 5 mg/ml with an Amicon YM-2 filter, and dialyzed against 4 x 105 M GdnHCl, 5 mM EDTA, and 50 mM Tris, pH 7.5.

Zn2+ Chelate Chromatography—Chelating Sepharose (50 ml) was charged with 500 ml of ZnCl2 (2 mg/ml) and then washed with water. The gel was packed into a 60-ml water-jacketed column on top of 10 ml of uncharged chelating Sepharose and equilibrated with 500 ml of Chelating equilibration buffer (15 M NaCl and 50 ml Tris-HCl, pH 7.5). The column was then washed with 20 ml of 2 M GdnHCl, and the proteoglycans were eluted with a linear gradient of 2–6 M GdnHCl. Fractions were analyzed for uronate, revealing the presence of two major peaks. Fractions containing both epiphycan and decorin were pooled; concentrated to 5 mg/ml with an Amicon YM-2 filter, and dialyzed against 4 x 105 M GdnHCl, 5 mM EDTA, and 50 mM Tris, pH 7.5.

The void volume peak from the Superdex 75 column was identified as being from the N terminus to Edman degradation. It was dialyzed against trypsin digestion buffer (50 mM Tris-HCl, pH 8) and further digested with trypsin. The products further separated on a Superdex 75 column, followed by reversed-phase HPLC. In this case, fractions were also assayed for sulfated GAGs by the dimethyl methane blue assay (14).

Peptide Nomenclature—Peptides are numbered in order from the N terminus to the C terminus. The letter preceding the number indicates which enzyme released the peptide from the parent protein. Thus, K-1-L-1 is the N-terminal peptide (T-1) derived from subdigestion of the N-terminal endoglycan Lys-C-derived peptide (K-1) with trypsin.

Identification of the Epiphycan-encoding cDNA Sequence—Initial protein sequence data enabled us to design two degenerate oligonucleotide primers (forward primer, HTAAYTTYAYSHGVTYYA; and reverse primer, CVCVABCKRTRTGGSWNT) that, based on sequence similarity to chick PG-Lb, would be expected to give a product of 230 base pairs. Reverse transcription-PCR of bovine chondrocyte RNA, with an annealing temperature of 40 °C, gave a band of the expected size, which was excised from an agarose gel. The PCR product was reamplified using similar primers modified with EcoRI and BamHI clamps at the 5’-ends and cloned into pBluescript™. Sequence analysis of the cloned insert indicated that the PCR product corresponded to the determined protein sequence.

A bovine cDNA AZAP library (Stratagene) was used to produce a cDNA library from bovine epiphyseal cartilage mRNA prepared by the method of Smale and Sasse (15). PCR primers were designed based on the sequence obtained from the previously obtained PCR product and the DNA sequence of the AZAP library arms. The entire coding region for bovine epiphycan was PCR-amplified and sequenced in both directions, either by sequence analysis of cloned PCR products or by direct sequencing of the PCR products using the originating PCR primers.

Isolation and Analysis of Radioiodinated Proteoglycans—The SLRPs epiphycan, decorin, and biglycan (50 µg each) were radioiodinated with 0.5 mCi of [125I]iodine by the chloramine-T method (16). The labeled proteoglycan was separated from unincorporated Na125I on a PD-10 column, followed by a 1-ml mono-Q-Sepharose column.

Molecular Mass Estimation—Labeled proteoglycans were further purified on a 0.5-ml DEAE-Sepharose column. The molecular masses of the intact proteoglycans were estimated by gel permeation chromatography on a Superose 6 10/30 column that was eluted at a flow rate of 0.4 ml/min in 4 M GdnHCl, 50 mM sodium acetate, pH 5.8, and 0.05% CHAPS and that had been calibrated with 13C-labeled molecular mass markers.

Isolation and Analysis of Core Proteins—The 125I-labeled SLRPs epiphycan, decorin, and biglycan (400,000 cpm each) were digested with chondroitinase ABC (10 units/ml) for 24 h at 37 °C in 0.1 M Tris, 30 mM sodium acetate, 0.2% bovine serum albumin, 10 mM EDTA, 10 mM N-ethylmaleimide, 5 mM phenylmethylsulfonyl fluoride, and 1 mg/ml pepstatin A. The 125I-labeled core proteins were applied either to a Superose 6 column as described above or onto a 5–15% SDS-polyacrylamide gel to determine the molecular mass.

Isolation and Analysis of Glycosaminoglycan Chains—Proteoglycans (50 µg) in 3 M sodium acetate, pH 6.0, were precipitated with 2 volumes of ethanol for 1 h at 20 °C. The precipitated material was collected by centrifugation for 15 min, washed with 70% ethanol, and dried in a Speedvac™. The proteoglycans were resuspended in 100 µl of 0.5 M NaOH containing 5 mCi of Na125I, at a final concentration of 1 × 107 cpm/ml, and allowed to react for 24 h at 45 °C. The samples were placed on ice; 1 M acetic acid was added dropwise until gas was no longer released; and the samples were then dried in a Speedvac™. The tritiated samples were washed twice in 10% methanol and dried again. Finally, the samples were resuspended in 500 µl of phosphate-buffered saline, 0.1% bovine serum albumin, and 0.1% dextran sulfate and applied to a 1-ml DEAE-Sepharose column equilibrated with phosphate-buffered saline containing 0.1% bovine serum albumin. The column was washed with equilibration buffer, and the GAGs were eluted with 2.5 column volumes of phosphate-buffered saline containing 1 M NaCl and 0.1% bovine serum albumin. Fractions were analyzed for radioactivity and analyzed by gel permeation chromatography on a Superose 6 10/30 column as described above. Molecular size estimates for GAG chains are based on the data of Wasteson (17).

RESULTS

Isolation of Small Proteoglycans from Epiphyseal Tissue—Small proteoglycans were isolated from bovine fetal epiphyseal cartilage as described previously for skin and bovine articular cartilage.
Characterization of Epiphycan

---

Fig. 1. Gel permeation chromatography of SLRP's (decorin, biglycan, and epiphycan) from bovine fetal epiphyseal cartilage on Sepharose CL-4B in 4 M GdnHCl. Fractions (14 ml) were collected at a flow rate of 5 ml/h and monitored for protein (A$_{280}$) and uronate (A$_{280}$) using the m-phenylphenol reaction (upper panel) and by SDS-PAGE (lower panel). The first peak that eluted in the void volume had an anomalous color in the m-phenylphenol reaction. As shown in the toluidine blue-stained 4–20% gradient slab gel, the second m-phenylphenol-positive peak eluted at $K_{av}$ = 0.4–0.7 and contained the dermatan sulfate proteoglycans, which were pooled as shown by the horizontal bar at the top of the chromatogram.

---

Cartilage (11, 12) by dissociative extraction, equilibrium density gradient centrifugation, DEAE-Sephacel chromatography, and gel permeation chromatography on Sepharose CL-4B. The resultant material (Fig. 1) contained a mixture of small proteoglycans.

The proteoglycan-containing fractions from the gel permeation chromatography were applied to an octyl-Sepharose column. The proteoglycans were eluted with a linear gradient of increasing GdnHCl concentration. Fractions were analyzed for uronate, which revealed the presence of two major peaks (Fig. 2). The material in these peaks was analyzed by SDS-PAGE and peptide mapping. The results indicated that the first peak contained biglycan, and the second peak contained decorin and a second somewhat larger proteoglycan. This unknown proteoglycan gave rise to unique tryptic peptides, which, when sequenced, showed homology to avian PG-Lb. This proteoglycan was named epiphycan. Fractions containing either biglycan or a mixture of epiphycan and decorin were pooled separately.

The pooled mixture of epiphycan and decorin was passed over chelating Sepharose charged with Zn$^{2+}$. Decorin bound to the Zn$^{2+}$-charged column, whereas epiphycan was not retained. The column was eluted with a linear pH gradient from pH 8 to 4. Epiphycan was completely separated from decorin in this chromatographic step as determined by monitoring for uronate, protein, and pH and by SDS-PAGE (Fig. 3). The fractions containing epiphycan were pooled as shown and assessed for homogeneity by SDS-PAGE.

Molecular Mass Determinations—The molecular mass of epiphycan was estimated by gel permeation chromatography of a 125I-labeled sample on a Superose 6 10/30 column. The elution profile of the epiphycan proteoglycan (Fig. 4) was compared with those of decorin and biglycan isolated from the same tissue and with 14C-labeled molecular mass standards. The elution profile of 125I-epiphycan showed one major peak with an elution position at $K_{av}$ = 0.31, which corresponds to a molecular mass of ~130,000 Da. Epiphyseal 125I-decorin and 125I-biglycan each eluted as single peaks at $K_{av}$ = 0.34 (120 kDa) and 0.29 (150 kDa), respectively.

The molecular masses of the 125I-labeled core proteins were determined by SDS-PAGE after digestion with chondroitinase ABC. Epiphycan, decorin, and biglycan core proteins were of similar size and migrated into the resolving gel to a position equivalent to a protein of 46 kDa (Fig. 5).

The size of the GAG chains was estimated by gel permeation chromatography after reductive β-elimination in the presence of 3H-labeled sodium borohydride. The radiolabeled GAG chains were purified on DEAE-Sepharose and analyzed on a Superose 6 column (Fig. 6). The epiphycan 3H-labeled GAGs eluted as a heterogeneous peak at $K_{av}$ = 0.46–0.57, corresponding to a molecular mass of ~23,000–34,000 Da. Tritium-labeled GAG chains derived from decorin and biglycan each eluted as one major peak at $K_{av}$ = 0.57 (23 kDa) and 0.54 (25 kDa), respectively.

Peptide Mapping—An endoprotease Lys-C digestion of unreduced epiphycan followed by gel permeation chromatography and reversed-phase separation of the products resulted in the isolation of peptides that were subsequently sequenced (Table I). A search of the GenBank™ Data Bank confirmed that all of these peptides covered most of the epiphycan core protein (Fig. 7).

cDNA Analysis—Alignment of the sequences of the tryptic peptides (Table I) generated from the decorin/epiphycan mixture (Fig. 2) with the protein sequence of PG-Lb enabled two degenerate PCR primers to be designed (Fig. 7). Reverse transcription-PCR amplification of a pool of fetal cartilage mRNA resulted in the expected 230-base pair product. The product was gel-purified and reamplified with primers that had restriction site-containing clamps attached at the 5'-end. This product was cloned into pBluescript and sequenced, confirming that the product derived from DNA coding for epiphycan. Similar products were obtained by PCR amplification of a human cDNA
The cloned PCR product was used to define specific primers that allowed the entire coding sequence to be amplified from a cDNA library generated from reverse-transcribed fetal epiphyseal cartilage mRNA. From this sequence, nested primers were used in a second round of PCR amplification to generate an 800-base pair PCR product, which was then excised and sequenced in both directions. The entire coding region of bovine epiphycan mRNA and the deduced amino acid sequence were determined (Fig. 7). The coding sequence of bovine epiphycan has 966 base pairs, corresponding to a translated protein of 321 amino acids. A signal peptide of 19 amino acids precedes the mature protein, which has a calculated size of...
Characterization of Epiphycan

34,721 Da and a pI of 4.52. The shorter form has a calculated size of 33,531 Da and a pI of 4.54.

Post-translational Modifications—Edman degradation of the intact protein indicated that epiphycan had two N-terminal sequences (APTLES... and ETYDAT...) in approximately equimolar amounts. The first N-terminal sequence (APTLES...) derived from removal of the signal peptide. The two sequences could be identified separately by virtue of the fact that a repeat of the second sequence beginning ETY... could be found after 11 cycles of sequencing. The second N-terminus (ETYDAT...) may derive from the action of a protease or may be a result of cleavage by exopeptidases. Similar N-terminal processing occurs in biglycan (19) and in decorin (20) and, in these cases, appears to have a role in control of GAG chain length, either by altering the rate of intracellular transport of the proteoglycan or by altering the rate of synthesis of the GAG chain (21).

A disulfide bond was unequivocally assigned between the two C-terminal cysteines. A peptide with two N termini (TPQ... and DMY...) K-9-K-12) (Table 1) eluted from a Superdex 75 column at a position consistent with a size of 4–8 kDa. These two peptides have predicted molecular masses of 1859 and 5756 Da, respectively, and would not be expected to coelute on gel permeation chromatography. Reduction of this peptide allowed the isolation of a peptide (K-9R) that corresponded to one-half of the disulfide-bonded pair. Thus, Cys-278

![Figure 4](image_url) 125I-Labeled proteoglycans purified from epiphyseal cartilage and analyzed by gel permeation chromatography on Superose 6. ●, radiolabeled epiphycan; □, epiphyseal decorin; △, epiphyseal biglycan. Epiphycan showed one major peak with an elution position at $K_{av} = 0.31$, corresponding to a size of ~130 kDa.

![Figure 5](image_url) 125I-Labeled proteoglycan core proteins purified from epiphyseal cartilage and analyzed by SDS-PAGE. Radiolabeled proteoglycans were digested with chondroitinase ABC and applied to a 5–15% SDS-polyacrylamide gel. Molecular mass marker positions (in kilodaltons) are shown on the left. Lane 1, epiphycan; lane 2, decorin; lane 3, biglycan. The high molecular mass bands in the biglycan sample are likely aggrecan core protein.

![Figure 6](image_url) GAG chain size determination by gel permeation chromatography. GAG chains, radiolabeled with tritium by treatment with 3H-labeled sodium borohydride, were purified on DEAE-Sepharose and then applied to a Superose 6 column. Epiphycan GAG chains have an elution profile at $K_{av} = 0.46–0.57$, corresponding to a size of ~23–34 kDa. Decorin GAG chains have an elution profile at $K_{av} = 0.57$, corresponding to a size of ~23 kDa, whereas biglycan GAGs have an elution profile at $K_{av} = 0.54$, corresponding to a size of ~25 kDa.
and Cys-311 are linked by a disulfide bond. We have not been able to confirm the presence of a disulfide-bonded loop at the N terminus.

Consensus sequences for N-linked oligosaccharides were present at positions 282 and 301. A peptide was found in which the N-terminal residue could not be identified (XLTYIRK, pep-
We have previously isolated decorin and biglycan from fetal skin and from bovine articular cartilage using a protocol involving dissociative extraction, ion-exchange chromatography, and hydrophobic chromatography on octyl-Sepharose 4B (11, 12). When small proteoglycans were isolated from fetal bovine epiphysis using the same procedure, we found three proteoglycans in roughly equal amounts. The third proteoglycan, which was named epiphycan based on its tissue source, was found to be the mammalian homolog of the avian proteoglycan PG-Lb. Epiphycan coeluted with decorin on octyl-Sepharose, but the two proteoglycans could be separated by metal chelate chromatography on Zn^{2+}-charged columns.

The ability to prepare significant amounts of proteoglycan (~0.5 mg from 50 g of epiphysis) has enabled us to chemically characterize the proteoglycan. The peptide sequence data allowed degenerate oligonucleotide primers to be designed. The PCR product was sequenced, and this was used to initiate the determination of the cDNA sequence of bovine and mouse (10) epiphycan.

N-terminal sequence analysis of intact epiphycan indicated the presence of two N termini. One of these (APTLSIN . . . ) is generated by removal of the signal peptide. The other N terminus (ETYDAT . . . ) did not conform to a signal peptide cleavage site and may derive from the action of a protease cleaving at NYNS-ETYD or from the action of exopeptidases. A similar two-step processing pathway has been noted for biglycan (19, 22).

The sizes of the two alternative core proteins calculated from the deduced amino acid sequence (33,531 and 34,721 Da) were smaller than those of decorin (36,421 Da) and biglycan (37,113). There are one (epiphycan), two (biglycan), or three (decorin) N-linked oligosaccharides attached to these proteins. As determined by SDS-PAGE (Fig. 5), there is little difference between the sizes of the core proteins of these three SLRPs after digestion with chondroitin ABC. If N-linked oligosaccharides were the only substituents on the epiphycan core protein, then the difference between epiphycan and decorin or biglycan would be substantial (at least 6 kDa), implying that, in epiphycan, there are additional post-translational modifications that would increase the apparent core protein size to the same range as decorin and biglycan. These modifications are presumably O-glycosylations; based on Edman degradation, there appear to be at least two O-linked oligosaccharides and one N-linked glycosaminoglycan in epiphycan.

Intact epiphycan is intermediary in size between decorin and biglycan. The core protein, with O- and N-linked oligosaccharides attached, is similar in size to both these proteoglycans.
The average size of the GAG chains released from epiphycan is rather larger than that of the GAG chains obtained from either decorin or biglycan. This would be consistent with the presence of one GAG chain in epiphycan. The tryptic peptide containing Ser-64 elutes over a broad range on reversed-phase HPLC (Fig. 8). This, coupled with its high molecular mass, suggests that it is substituted with a GAG chain. In contrast, Ser-95 is found on a peptide that elutes late and as a symmetrical peak on reversed-phase HPLC. This suggests that this peptide has a smaller and more homogeneous carbohydrate substituent, likely a conventional O-linked oligosaccharide.

Epiphycan is the mammalian homolog of chick PG-Lb (9) and, within the LRR-containing region, is 78% identical. Epiphycan is also related (49% identity within the LRR-containing region) to osteoglycin (formerly named osteoinductive factor), which is a proteoglycan found in the extracellular matrix of developing bone (23). The epiphycan/PG-Lb family appears to be a separate branch of the leucine-rich proteoglycans. An unrooted phylogeny diagram is shown in Fig. 9, indicating the relationship of epiphycan to chick PG-Lb, to a partial sequence of a shark analog of these proteoglycans,2 and to osteoglycin, decorin, biglycan, fibromodulin, lumican, PRELP, and chondroadherin. Each family appears to be essentially unrelated to the other families, although all have the common feature of LRRs. It is noteworthy that mammalian osteoglycin is more remotely related to mammalian epiphycan (47% identical to the LRR region of bovine epiphycan) than the partial shark-derived sequence (56% identical to the LRR region of bovine epiphycan), implying that it diverged from a common ancestor earlier than the cartilaginous fishes. It is therefore reasonable to assume that osteoglycin has a different role from epiphycan in the same way that decorin and biglycan are in the same subfamily, but have different properties and therefore, presumably, different roles in the extracellular matrix.

The seven LRRs in epiphycan are heterogeneous in length. The start of the first detectable consensus sequence is 15 residues after the fourth cysteine. In common with other SLRPs, the first LRR is atypical, starting with a hydroxylamino acid and a weak consensus motif. This is likely to be due to a dramatic change in structure at this point, corresponding to the interface with the four-cysteine cluster. The lengths of the LRR-containing sequences are 24, 24, 20, 26, 21, 31, and 32 amino acids. The last LRR appears between the two C-terminal cysteines. This dimeric repeat pattern (long-short-long-short-long) differs from that of the proteoglycans decorin, biglycan, fibromodulin, and lumican, which have a triplet repeat pattern (long-long-short) (24). This, in turn, differs from the complete regularity of the RNase inhibitor, which has 15 LRRs spaced at

2 P. J. Neame, unpublished data.
Characterization of Epiphycan

18717

P. J. Neame and C. Kay, unpublished data.

intervals of 28 or 29 amino acids (8). A similar regularity is seen in chondroadherin, where the repeats are spaced at intervals of 24 amino acids (25). The significance of these patterns will probably become clear once the 3-dimensional structure of these domains has been determined, but may reflect a mechanism for changing the overall curvature of the molecule.3

Information on structurally important features of proteins can often be obtained by comparison of the same protein in different species. Comparison of chick PG-Lb, human epiphycan, and murine PG-Lb with bovine epiphycan shows that the majority of the canonical SLRP structure (cysteine-rich region, a series of LRRs, followed by a C-terminal disulfide bond(s)) is highly conserved (Fig. 10). The majority of changes are conservative. In common with decorin and biglycan, the N terminus, in front of the first cysteine, differs considerably between species. However, a section of 30 amino acids in front of the first cysteine is quite conserved between members of this family, in front of the first cysteine, differs considerably between species. However, a section of 30 amino acids in front of the first cysteine is quite conserved between members of this family. This region contains either an O-linked oligosaccharide (in epiphycan), as shown here, or a GAG chain (postulated in PG-Lb). The conservation of this region may indicate functional importance. Osteoglycin is the closest relative to epiphycan, related to these proteoglycans. It bears a greater similarity to bovine osteoglycin than to bovine epiphycan and so may be a shark counterpart to osteoglycin.

It remains to be seen to what extent the mammalian homolog of PG-Lb mimics the avian proteoglycan. Immunolocalization of PG-Lb in developing chick limb indicated that PG-Lb was most abundant in the region that contained flattened chondrocytes (9). This would imply that epiphycan may have a function in a region of cartilage that is not associated with calcification, perhaps acting to delay the onset of calcification or to arrange the matrix so that it is ready for the extensive remodeling that occurs during calcification. It may control collagen fibrillogenesis in a region where the type II collagen will be completely removed and replaced with a calcified, type I collagen-containing extracellular matrix.

Acknowledgment—We are grateful to Ray Boynton for preparation of the epiphyseal RNA.

REFERENCES

1. Iozzo, R. V., and Murdoch, A. D. (1996) *FASEB J.* **10**, 598–614
2. Vogel, K. G., Paulsson, M., and Heinegård, D. (1984) *Biochim. Biophys. Acta* **773**, 587–597
3. Hedbom, E., and Heinegård, D. (1993) *J. Biol. Chem.* **268**, 27307–27312
4. Yamaguchi, Y., Mann, D. M., and Ruoslahti, E. (1990) *Nature* **346**, 281–284
5. Hildebrand, A., Romaris, M., Rasmussen, L. M., Heinegård, D., Twardzik, D., Border, W. A., and Ruoslahti, E. (1994) *Biochem. J.* **302**, 527–534
6. Patthy, L. (1987) *J. Mol. Biol.* **196**, 567–577
7. Kobe, B. (1994) *Trends Biochem. Sci.* **19**, 415–421
8. Kobe, B., and Deisenhofer, J. (1993) *Nature* **366**, 751–756
9. Shimomura, T., and Kimata, K. (1992) *J. Biol. Chem.* **267**, 1265–1270
10. Kurita, K., Shimomura, T., Ujita, M., Zako, M., Kida, D., Iwata, H., and Kimata, K. (1996) *Biochem. J.* **318**, 909–914
11. Choi, H. U., Johnson, T. L., Pal, S., Tang, L.-H., Rosenberg, L., and Neame, P. J. (1989) *J. Biol. Chem.* **264**, 2876–2884
12. Rosenberg, L., Choi, H. U., Tang, L.-H., Johnson, T. L., Pal, S., Webber, C., Reiner, A., and Poole, A. R. (1985) *J. Biol. Chem.* **260**, 6304–6313
13. Neame, P. J., Choi, H. U., and Rosenberg, L. C. (1989) *J. Biol. Chem.* **264**, 8653–8661
14. Fardales, R. W., Buttle, D. J., and Barrett, A. J. (1986) *Biochim. Biophys. Acta* **883**, 173–177
15. Smale, G., and Sasse, J. (1992) *Anal. Biochem.* **203**, 352–356
16. Bolton, A. E., and Hunter, W. M. (1973) *Biochem. J.* **133**, 529–539
17. Wasteson, A. (1971) *J. Chromatogr.* **59**, 87–97
18. Deere, M., Johnson, H. J., Garza, S., Harrison, W. R., Yoon, S., Elder, F. B., Kucherlapati, R., Hoik, M., and Hecht, J. T. (1996) *Genomics* **38**, 399–404
19. Marcum, J. A., and Thompson, M. A. (1991) *Biochem. Biophys. Res. Commun.* **175**, 706–712
20. Bourdon, M., Kruusius, T., Campbell, S., Schwartz, N., and Ruoslahti, E. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 3194–3198
21. Oldberg, A., Antonsson, P., Moses, J., and Fransson, L. A. (1996) *FEBS Lett.* **386**, 29–32
22. Roughley, P. J., White, R. J., and Mert, J. S. (1996) *Biochem. J.* **318**, 779–784
23. Beintz, H., Nathan, R. M., Rosen, D. M., Armstrong, R. M., Thompson, A. Y., Segarini, P. R., Mathews, M. C., Dasch, J. R., Piez, K. A., and Seyedin, S. M. (1989) *J. Biol. Chem.* **264**, 20805–20810
24. Bengtsson, E., Neame, P. J., Heinegård, D., and Sommarin, Y. (1995) *J. Biol. Chem.* **270**, 25639–25644
25. Neame, P. J., Sommarin, Y., Boynton, R. E., and Heinegård, D. (1994) *J. Biol. Chem.* **269**, 21547–21554
26. Higgins, D. G., Bleasby, A. J., and Fuchs, R. (1991) *Comput. Appl. Biosci.* **8**, 189–191