Identification and Characterization of Asporin

A NOVEL MEMBER OF THE LEUCINE-RICH REPEAT PROTEIN FAMILY CLOSELY RELATED TO DECORIN AND BIGLYCAN*

Asporin, a novel member of the leucine-rich repeat family of proteins, was partially purified from human articular cartilage and meniscus. Cloning of human and mouse asporin cDNAs revealed that the protein is closely related to decorin and biglycan. It contains a putative propeptide, 4 amino-terminal cysteines, 10 leucine-rich repeats, and 2 C-terminal cysteines. In contrast to decorin and biglycan, asporin is not a proteoglycan. Instead, asporin contains a unique stretch of aspartic acid residues in its amino-terminal region. A polymorphism was identified in that the number of consecutive aspartate residues varied from 11 to 15. The 8 exons of the human asporin gene span 26 kilobases on chromosome 9q31.1–32, and the putative promoter region lacks TATA consensus sequences. The asporin mRNA is expressed in a variety of human tissues with higher levels in osteoarthritic articular cartilage, aorta, uterus, heart, and liver. The deduced amino acid sequence of asporin was confirmed by mass spectrometry of the isolated protein resulting in 84% sequence coverage. The protein contains an N-glycosylation site at Asn281 with a heterogeneous oligosaccharide structure and a potential O-glycosylation site at Ser54. The name asporin reflects the aspartate-rich amino terminus and the overall similarity to decorin.

Cartilage matrix consists of fibrillar networks, primarily of collagen II and highly negatively charged molecules of aggregates. There are also a number of noncollagenous glycoproteins that apparently contribute to the regulation of tissue assembly and properties. Among them is the family of the leucine-rich repeat (LRR) proteins, which contains several members found in the extracellular matrix. There are currently 11 known members of this family. These molecules share a common structure with a central stretch of LRRs. This LRR domain is flanked by disulfide bridged loops, with 4 cysteine residues preceding the LRR domain and 2 on its C-terminal side. Apart from chondroadherin, these proteins also contain divergent amino-terminal extensions with features unique for the different proteins. Based on amino acid sequence and gene organization the family can be divided into four distinct groups.

Decorin (1) and biglycan (2) constitute the first group (class I). These proteins have 10 LRRs and carry one and two chondroitin or dermatan sulfate chains, respectively. The glycosaminoglycan chains are linked to serine residues in the amino terminus. The molecules in this group are secreted with a propeptide.

The second group (class II) consists of fibromodulin (3), lumican (4), keratan (5), PRELP (6), and osteoadherin (7). Like the class I proteins they consist of 10 LRRs. With the exception of PRELP, they all carry polylactosamine or keratan sulfate chains linked to the LRR region and sulfated tyrosine residues in the amino-terminal extension. In contrast, the amino terminus of PRELP has a cluster of positively charged amino acid residues that mediate binding to heparan sulfate (8). Unlike all other family members, osteoadherin contains a COOH-terminal extension (7).

Epiphycan/PG-Lh/DSPG3 (9–11), mimecan/osteoglycin (12, 13), and optin/oculoglycan (14–16) form the third group (class III). These are smaller molecules with only 6 LRRs and all contain sulfated tyrosine residues in the amino-terminal extension. In addition, epiphycan carries chondroitin sulfate, other O-linked oligosaccharides, and a cluster of glutamate residues in this region (9). The amino-terminal extension of optin carries O-linked oligosaccharides (14), and contains a heparin-binding consensus sequence (17).

Chondroadherin (18) forms the fourth branch on the extracellular matrix LRR protein family tree (class IV). This protein contains 10 LRRs, but lacks both amino- and COOH-terminal extensions outside the cysteine motifs. Nyctalopin, a recently published glycosylphosphatidylinositol-anchored LRR protein may also be a member of this subfamily (19, 20).

As is evident from the summary above, the subdivision of LRR proteins into classes based on sequence does not reflect the functions of the molecules. For example, decorin, biglycan, and epiphycan are chondroitin or dermatan sulfate proteoglycans, and may as such be more functionally related than, e.g. the different class II LRR proteins. A major functional property flight; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction.

Received for publication, December 4, 2000, and in revised form, December 27, 2000
Published, JBC Papers in Press, January 10, 2001, DOI 10.1074/jbc.M010932200

Pilar Lorenzo$$, Anders Aspberg$$, Patrik Önnerfjord§, Michael T. Bayliss¶, Peter J. Neame¶, and Dick Heinegård**

From the Department of Cell and Molecular Biology, Section for Connective Tissue Biology, Lund University, BMC plan C12, SE-221 84 Lund, Sweden, the Royal Veterinary College, Royal College Street, London NW1 0TU, United Kingdom, and the Shriner’s Hospital for Children, Tampa, Florida 33612

* This work was supported by the Medical Faculty of Lund University, the Swedish Medical Research Council, IngaBritt and Arne Lundbergs Research Foundation, Greta and Johan Kocks Stiftelse, Konung Gustaf V:s 80-arfsfond, Reumatikerförbundet and Alfred Lundbergs Research Foundation, Greta and Johan Kocks Stiftelser, Konung Gustaf V:s 80-arfsfond, Reumatikerförbundet and Alfred Osterlunds Stiftelse. The costs of publication of this article were underwritten by the GenBank©/EBI Data Bank with accession number(s) AF316824 and AF316825.

$$ Contributed equally to the results of this work.

§ Contributed equally to the results of this work.

¶ To whom correspondence should be addressed. E-mail: dick.heinegard@medkem.lu.se; Tel.: 46-46-222-8571; Fax: 46-46-211-3412.

** The abbreviations used are: LRR, leucine-rich repeat; GdnHCl, guanidinium hydrochloride; HPLC, high-pressure liquid chromatography; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction.
Asporin, A Novel LRR Protein

that is shared between most of the class I, II, and IV LRR proteins is a capacity to bind to collagen via the LRR domain. This is a high affinity binding with \( K_d \) in the nanomolar range. The different NH\(_2\)-terminal extensions offer a variety of opportunities for interactions with other matrix constituents, including other fibers of collagen, thereby providing cross-linking and stabilization of the fibrillar network.

Several of these molecules appear to have roles in modulating the assembly of collagen fibrils as is indicated by experiments in vitro (21–24) as well as by gene inactivation studies (25–27). Invariably, these studies show altered collagen fiber dimensions when the abundance of the LRR protein is changed.

The present work started with a study of altered biosynthesis of proteins in early human osteoarthritics. We found a number of proteins to be up-regulated, one being a component with an apparent size of 39 kDa. This component appeared structurally related to fibromodulin since it cofractionated in a variety of separation procedures. We now define the primary structure including a putative polymorphism, oligosaccharide side chain substituents, and tissue expression of the protein. It represents a novel member of the LRR protein family belonging to the decorin/biglycan group (class I). The protein is named asporin based on the presence of a polyaspartate stretch in the amino-terminal region and the similarity with decorin.

**EXPERIMENTAL PROCEDURES**

**Tissue Extraction**—Normal human knee cartilage (18.5 g of tissue wet weight, donor age 32 to 50 years) and menisci (3.5 g of tissue wet weight) were obtained at surgery. The tissues were dissected clean, sliced into fine pieces, and disrupted using a high speed homogenizer (Polytron, Kinematica GmbH) in 12 volumes (v/w) of 4 M GdnHCl, 0.05 M sodium acetate, pH 5.8, containing protease inhibitors (5 mM benzamidine hydrochloride, 0.1 mM 6-aminohexanoic acid). After extraction for 24 h at 4 °C the remaining insoluble material was removed by centrifugation at 20,000 × g at 4 °C for 30 min.

**Protein Purification**—Proteins in the cartilage extract were separated from proteoglycans by CsCl density gradient centrifugation with a starting density of 1.5 g/ml under dissociative conditions in 4 M GdnHCl as described elsewhere (28). The gradient tube was divided into 4 equal fractions using a Beckman tube slicer, and the top fraction (D4) was used for subsequent purification. The D4 fraction was then concentrated by ultrafiltration (PM-10 membrane, Amicon), followed by diaflow against 4 M GdnHCl, 20 mM Tris-HCl, pH 8, and applied to a Superose 6 column (2.2 × 100 cm) in 2.5-ml aliquots. Fractions of 2.5 ml were collected, monitored for protein content by measuring their absorbance at 280 nm, and analyzed by SDS-PAGE after ethanol precipitation, as previously described (29).

The proteins from the extract were separated into two peaks, a larger containing proteins of high molecular weight (fractions 25 to 45) and a smaller containing the smaller proteins. The latter fractions (46 to 65) were pooled and concentrated by ultrafiltration followed by diaflow against 7 M urea, 20 mM Tris-HCl, pH 8. The pooled material was then loaded onto a 30-ml bed volume column of DEAE-cellulose (1.6 × 15 cm, DE52, Whatman) equilibrated in the urea buffer. After sample loading, the column was washed with 5 bed volumes of the equilibration buffer, and eluted with a 800-ml linear gradient (27 bed volumes) of 0 to 1 M NaCl in the equilibration buffer at a flow rate of 20 ml/h. Fractions of 10 ml were collected, monitored for protein content by measuring their absorbance at 280 nm, and analyzed by SDS-PAGE.

The fractions containing asporin were pooled, concentrated by ultrafiltration followed by diaflow against 7 M urea, 10 mM HCOOH, pH 4.0, and chromatographed on a 20-ml bed volume of Q-Sepharose Fast Flow (1.6 × 8.5 cm, Amersham Pharmacia Biotech) anion exchange column equilibrated in urea buffer. The column was washed with 5 bed volumes and the bound proteins were step eluted at a flow rate of 20 ml/h with the equilibration buffer containing 1 m NaCl. Fractions of 2 ml were collected, monitored for protein content by measuring their absorbance at 280 nm, and analyzed by SDS-PAGE.

The fractions containing asporin were pooled and equilibrated by diaflow to 7 M urea, 10 mM HCOOH, pH 4.0, and applied to a Mono Q HR 5/5 column (Amersham Pharmacia Biotech). The bound proteins were eluted with a 15-ml linear gradient (15 bed volumes) from 0 to 1 M NaCl at a flow rate of 30 ml/h. Fractions of 2 ml were collected, monitored for molecular weight, and protein, and analyzed by SDS-PAGE.

Initial characterization of the meniscus extract showed a low fraction of high molecular weight proteoglycan. Therefore, the sample was taken directly to chromatography, omitting cesium chloride gradient centrifugation. Forty milliliters of meniscus extract were concentrated by ultrafiltration (PM-10 membrane, Amicon) against 7 M urea, 20 mM Tris-HCl, pH 8.0, by diaflow and directly chromatographed over Q-Sepharose Fast Flow followed by chromatography over Mono Q, as described above except that the pH was kept at 8.0.

After SDS-PAGE analysis, the fractions from the Mono Q chromatography containing asporin were pooled and concentrated by ultrafiltration, followed by diaflow against 4 M GdnHCl, 50 mM sodium acetate, pH 5.8, 24 h at 4 °C the remaining insoluble material was washed with 5 bed volumes of the equilibration buffer, and applied to a Mono S chromatography column (5/500, Pharmacia) equilibrated and eluted at 0.2 ml/min with a linear gradient of 0 to 1 M NaCl, pH 3.5. Fractions of 0.5 ml were collected, monitored for protein content by measuring their absorbance at 280 nm. Protein patterns were analyzed by SDS-PAGE.

**Protein Sequencing**—Proteolytic digestion with Lys-C (Roche Molecular Biochemicals) was performed at enzyme to substrate ratios of 1:50 according to the manufacturer’s instructions. Peptides were separated by reversed phase HPLC on a Vydac C18 column (2.1 × 30 mm), eluted with a gradient of acetonitrile (0–70% over 45 min) in 0.1% trifluoroacetic acid at a flow rate of 0.2 ml/min. The effluent was monitored at 220 nm. Peptides were sequenced on an Applied Biosystems 477A automated sequencer with on-line analysis of phenylthiohydantoin derivatives on an Applied Biosystem HPLC system.

**cDNA Cloning**—All the molecular biological procedures, including agarose gel electrophoresis, restriction enzyme digestion, ligation, bacterial transformation, and DNA sequencing, were performed according to standard methods (30).

The amino acid sequences obtained from endoproteinase Lys-C-digested asporin were used to search the GenBankTM data base with the TBLASTN 2.1 program (31). The EST sequences identified from this search were aligned and assembled. The resulting full-length sequence was used for designing primers h39k-S (5′-CTTTCTACATAAGACC-3′) and h39k-AS (5′-AAATGACATTACCAATTAC-3′).

Human osteoarthritic articular cartilage was obtained at surgery after total hip replacement, kept in phosphate-buffered saline during dissection, shaved and frozen in liquid nitrogen. Total RNA and mRNA were purified as described previously (32). First strand cDNA was prepared with oligo(dT)\(_12\)\(_20\) and reverse transcriptase with Superscript (Life Technologies). After digestion of the mRNA with RNase H, the asporin cDNA was obtained using the polymerase chain reaction (PCR) with primers h39k-S and h39k-AS and Pfu DNA polymerase. After initial denaturation step at 95 °C for 1 min, the DNA was amplified for 30 cycles of 45 s at 95 °C, 45 s at 54 °C, and 2 min 40 s at 72 °C. The resulting 1.2-kilobase product was isolated from agarose gel, purified using a Qiaquick PCR purification kit (Qiagen), and ligated into the pCR-Script SK(+) vector (Stratagene). The PCR product and several of the resulting pCR-Script clones were sequenced using the BigDye kit (ABI) and run on a ABI 310 DNA sequencer. In addition to primers T3, T7, h39k-S, and h39k-AS two internal primers were used: h39k-IntS (5′-ATGAAAAAATATGTTAGAGAAAATAC-3′) and h39k-IntAS (5′-AGGTTTGCACATTCC-3′). The resulting sequence tracings were assembled using the SeqMan II module of the LaserGene software (DNASTar Inc.).

A first draft full-length mouse asporin sequence was assembled from sequences obtained through a BLASTN search of the mouse EST section of GenBank with the human asporin sequence. Using this draft sequence the primers m39k-S (5′-ACCTTGTAACCGCGCCAG-3′), m39k-AS (5′-TTATATATTATGACCATCTAG-3′), m39k-IntS (5′-GACCTCACAATAATACATAC-3′), and m39k-IntAS (5′-TTGGTATAATGCAAAGATCC-3′) were designed. Mouse aorta first strand cDNA was prepared and asporin cDNA amplified from this by reverse transcriptase PCR using primers m39k-S and m39k-AS, as described above. The PCR product was cloned into pCR-Script and sequenced using all four m39k primers, as well as T3 and T7 primers. The human and mouse asporin cDNA sequences were deposited in GenBank with the accession numbers AF316824 and AF316825, respectively.

**Messenger RNA Expression Analysis**—For Northern blot analysis 10 μg of total RNA isolated from human osteoarthritic articular cartilage were electrophoresed on 1% formaldehyde-agarse gel, and transferred to a nitrocellulose filter (NitroPure, Micron Separation). Membranes of Multiple Tissue Northern blot and Human RNA Master Blot were from CLONTECH. The membranes were hybridized with a 463-base pair DNA fragment (nucleotides 392–845 of the human sequence, Fig. 3).
labeled with \(\alpha\)-PtdCTP by using the Random Primed DNA labeling kit (Roche Molecular Biochemicals). Hybridization and washing of the membranes were according to the manufacturer’s instructions. The membranes were allowed to expose x-ray film ( Biomax MS, Kodak) or analyzed by the Bas2000 phosphoimaging system (Fuji).

**RESULTS**

**Partial Purification of Asporin from Cartilage—**Extraction of human articular cartilage with 4 M GdnHCl followed by cesium chloride gradient centrifugation separated the matrix proteins from the bulk of the large proteoglycans in the cartilage. Fractionation of the extract by gel filtration on Superose 6 resulted in two pools, one containing large proteins and the other with proteins of lower molecular masses (5-7 kDa). The proteins in this latter pool were fractionated by DEAE ion exchange chromatography, where asporin was observed in the fractions also containing fibromodulin. These fractions were then chromatographed on a Q-Sepharose anion exchange column in 7 M urea, 20 mM Tris-HCl, pH 8.0. Asporin was recovered in a few fractions identified by SDS-PAGE. These fractions were pooled and further fractionated by gel filtration on two tandemly arranged columns of Superdex 200 and Superose 6 (Fig. 2). Asporin eluted in a few fractions together with a minor proportion of the fibromodulin.

**Peptide Sequencing of Asporin—**After Lys-C digestion of the intact protein excised from an SDS-polyacrylamide gel, peptides were separated by reversed phase HPLC. Peaks were collected and analyzed. Some peaks gave two sequences, but by analysis of the relative yields of the amino acids at each cycle, it was possible to determine both sequences with a high degree of confidence. As the protein was not reduced and carboxymethylated, no peptides were isolated that contained cysteine.

**Mass Spectrometry—**Mass spectrometric studies were performed using a Bruker Scout 384 Reflex III matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometer. The instrument was tuned in the positive ion mode with delayed extraction and an acceleration voltage of 26 kV, which also extracted the peptides out of the gel. After a minimum 1-h digestion at room temperature, peptides were purified from buffer using miniaturized C-18 reversed phase tips (Ziptips<sup>TM</sup>, Millipore). Purified peptides were eluted directly onto the sample target using acetonitrile, 0.1% trifluoroacetic acid (1:1). Various matrices were used to increase the sequence coverage. When using water-soluble matrices such as 2,4,6-thihydroxyacetophenone and 2,5-dihydroxybenzoic acid, a low content of large proteoglycans (data not shown). An aliquot of the mixtures before and after digestion was diluted with sample buffer (2% SDS, 0.125 M Tris-HCl, pH 6.8, 0.002% bromphenol blue, and 20% glycerol), boiled at 100 °C for 4 min and electrophoresed on the gradient polyacrylamide gel. Proteins were visualized by staining with Coomassie Brilliant Blue R-250 (Serva).

**Sample Preparation for Mass Spectrometry—**Coomassie-stained bands on SDS-PAGE gels were excised and washed extensively using 40% acetonitrile in 25 mM NH<sub>4</sub>HCO<sub>3</sub>, pH 7.8. After washing, the gel pieces were dried in a SpeedVac and subsequently reduced and alkylated using 10 mM dithiothreitol and 55 mM iodoacetamide at 56 °C (30 min) and at 20 °C (30 min), respectively. Samples were then washed and dried before digestion overnight at 37 °C using 10–20 μl of sequencing grade endoproteases such as trypsin (Promega) or Glu-C (Roche Molecular Biochemicals). An aliquot of the mixtures before and after digestion was diluted with sample buffer (2% SDS, 0.125 M Tris-HCl, pH 6.8, 0.002% bromphenol blue, and 20% glycerol), boiled at 100 °C for 4 min and electrophoresed on the gradient polyacrylamide gel. Proteins were visualized by staining with Coomassie Brilliant Blue R-250 (Serva).

A, elution profile after Mono Q anion exchange chromatography. B, SDS-PAGE analysis of some of the fractions after reduction on a 4–16% gradient gel. The gel was stained for proteins with Coomassie Brilliant Blue. Molecular mass markers are indicated on the left and the position of asporin by an arrow.
Asporin, A Novel LRR Protein

**TABLE I**

| Peptide | TBLASTN hit | Alignment |
|---------|-------------|-----------|
| 1 | None (see peptide 2) | KsRRLYLS |
| 2 | KsRRLYLSHTL6SE1P : :: :::::::: : : | AK000136 118 KLRRLYSHNL6SE1P 165 |
| 3 | KxTDIEAGSLANPPRRE1 : :: :::::::: : : | AK000136 493 KITDIENGSLANPRVRE1 549 |
| 4 | KxLPTTLLXL1den : :: :::::::: : : | AK000136 376 KGLPTTLLXHLDYN 420 |
| 5 | KxVFQNNSTSIQE | Fibromodulin 342 KYYVFQNNSTSIQE 386 |
| 6 | KxTDIEAGSLANPR : :: :::::::: : : | AK000136 493 KITDIENGSLANPR 537 |
| 7 | KxLTSLYGLLNN : :: :::::::: : : | AK000136 34 KGLTSLYGLLNN 72 |
| 8 | KELQLGLGNN | KELQLGLGNN 492 |
| 9 | None | KsLAEL |

**Determination of the Asporin Nucleotide and Amino Acid Sequence**—TBLASTN searches with the 9 peptide sequences obtained (Table I) showed that six of these were contained within an EST clone (GenBank™ accession number AK000136). One peptide was derived from fibromodulin. Two peptides were too short to produce BLAST hits, but the sequences of these are present in the AK000136 sequence (Fig. 3).

AK000136 is an EST sequence deposited in GenBank™ as a putative extracellular matrix protein. The deduced AK000136 sequence contains several leucine-rich repeats and the two COOH-terminal cysteine residues typical of the extracellular matrix LRR-repeat protein family. When AK000136 was used as the query in further BLAST searches, a number of other EST sequences were identified. Assembly of these sequences produced a longer open reading frame that included a signal peptide and the amino-terminal 4-cysteine motif of the extracellular matrix LRR proteins. The cDNA of the novel LRR protein was cloned through reverse transcriptase PCR from cellular matrix LRR proteins. The cDNA of the novel LRR protein was cloned through reverse transcriptase PCR from human femoral head osteoarthritic cartilage, using primers corresponding to the 5′- and 3′-untranslated regions of the assembled consensus sequence. The mouse homologue was similarly identified through BLAST searches of the mouse EST database with the human sequence and cloned from mouse femoral head osteoarthritic cartilage, using primers corresponding to the 5′- and 3′-untranslated regions of the assembled consensus sequence. The mouse homologue was also identified through BLAST searches of the mouse EST database with the human sequence and cloned from mouse femoral head osteoarthritic cartilage, using primers corresponding to the 5′- and 3′-untranslated regions of the assembled consensus sequence.

The human and mouse asporin sequences are shown in Fig. 3. The predicted amino acid sequences of the two proteins are 90% identical. The four amino-terminal cysteines show the C-X₉₋₁-C-X₃₋₁-C pattern typical of decorin and biglycan (37), which clearly identifies asporin as a member of the class I branch of the LRR proteins. Indeed, like decorin and biglycan, asporin contains a highly conserved putative propeptide sequence (amino acid residues 15–32). The putative propeptide cleavage site conforms to the bone morphogenetic protein-1 cleavage site in biglycan (38). Alignment of the LRRs of asporin to decorin and biglycan reveal a striking conservation in amino acid sequence as well as repeat length (Fig. 4). Construction of a phylogenetic tree of the extracellular matrix LRR proteins using Clustal W confirmed that asporin belongs to the type I group, i.e. the decorin and biglycan branch (Fig. 5). Unlike decorin and biglycan, asporin contains no consensus glycosaminoglycan attachment sites (Ser-Gly) in its amino terminus. There is, however, one conserved consensus site for N-linked glycosylation (Asn268 and Asn275 in the human and mouse asporin sequences, respectively). In contrast to all previously identified extracellular matrix LRR proteins, asporin has a stretch of 13 aspartic acid residues in its amino-terminal region. Interestingly, we found that the number of consecutive aspartic acid residues is variable. When performing direct sequencing of the human asporin PCR product, the sequence trace ended abruptly after the first 13 Asp residues in the human sequence (Asp30). Sequencing subcloned cDNA revealed that some clones contained an additional Asp codon at this position. Indeed, several clones with varying numbers of Asp residues (11–15) were identified in the human EST data base (not shown). In addition, the genomic sequence of human asporin (see below) coded for 15 contiguous Asp residues. Since the first-strand cDNA used in cloning the human asporin cDNA was prepared from tissue pooled from several individuals, we believe that this represents a polymorphism. We found no corresponding variation in the Asp stretch of the mouse protein, which comprises 7 Asp and 1 Asn residues.

A UniGene search of GenBank™ with AK000136 yielded clusters Hs.10760, Rn.43324, and Mm.132637 for the human, rat, and mouse homologues, respectively. The human asporin gene is located on chromosome 9q31.1–32, within the interval D9S1842-D9S196. This interval also contains the genes for the LRR proteins osteoadherin/osteomodulin (OMD) and mimecan/osteglycin (OGN). A full-length asporin cDNA was assembled from our sequence and a number of overlapping EST sequences to obtain the 5′- and 3′-untranslated regions. BLAST searches
of the high throughput genomic sequence division of GenBank™ identified a contig from chromosome 9 that contained the full asporin sequence (GenBank™ accession number AL137848). The first exon is also present in the overlapping contig AL157827. As shown in Fig. 6, the asporin gene spans over 26 kilobases and consists of 8 exons.
Asporin follows the gt-ag rule and the introns show the same codon phases as the corresponding introns in decorin and biglycan (Table II). Indeed, the introns are positioned in the exact corresponding locations as in decorin and biglycan (Fig. 4). It is presently unknown whether any additional alternatively spliced untranslated exons are present in the 5’-end of the gene, as is the case in decorin. Like biglycan no consensus TATA box is found 5’ of the first exon of the asporin gene. A number of transcription factor binding sites (including AP-1) were, however, identified immediately upstream of the asporin exon 1 (not shown).

Asporin mRNA Expression—Northern blot analysis demonstrated that the asporin gene codes for a single message of 2.56 kilobases (Fig. 7). Using a commercial human tissue RNA filter (Multiple Tissue Northern) we found that the highest amount of message was present in the heart tissue, followed by the liver whereas the message was almost undetectable in the other tissues. As articular cartilage is not included on the commercial membrane it was not possible to directly compare asporin expression in cartilage with that in other tissues. However, the Northern blot analysis of human osteoarthritic cartilage total RNA showed a strong hybridization signal. Considering that 10 mg of total cartilage RNA (less than 1 mg of mRNA) was loaded, as compared with 2 mg of poly(A)1 RNA from the other tissues, and that the autoradiograms were exposed equally long, the expression of asporin may well be higher in articular cartilage than in the other tissues investigated.

A broader screening for the presence and relative abundance of asporin was done by hybridization of a normalized mRNA dot blot (Human RNA Master Blot) which covers adult and fetal tissues (Fig. 8). The asporin cDNA probe hybridized with a wide range of human tissues with the highest signal levels in aorta and uterus. Moderate expression levels were found in small intestine, heart, liver, bladder, ovary, stomach, and in the adrenal, thyroid, and mammary glands. Low asporin expression was observed in trachea, bone marrow, and lung. There was a notable lack of signal in the central nervous system as well as in spleen and thymus. A similar asporin expression pattern was observed in fetal tissues.

Characterization of the Asporin Protein—Peptide mapping using MALDI-TOF mass spectrometry was used to verify the previously obtained protein sequence. After digestion with endoproteinases the peptide masses obtained were compared with the peptides expected from the novel protein. The identified peptides of asporin are listed in Table III. The identified peptides cover 84% of the mature protein sequence. The only major peptide missing in the sequence is the absolute amino-terminal peptide (amino acids 33–62) containing the consecutive Asp residue. This is probably due to the extreme acidity of this peptide, which makes ionization and thus mass spectrometry very difficult.

The protein has one potential N-glycosylation site at Asn281. Treatment of asporin with N-glycosidase F confirmed that the
Asporin, A Novel LRR Protein

Exonic sequence is in upper case, intronic sequence in lower case, and splice donor/acceptors in boldface. Codon phase refers to position of intron before (0), after the first nucleotide (1) or after the second nucleotide (II) of the codon.

| Exon | Exon length (bp) | 5’ Splice donor | Intron length (bp) | 3’ Splice acceptor | Codon phase |
|------|-----------------|----------------|-------------------|-------------------|-------------|
| 1    | 197             | .CTGACAGgtatg tgtt | >7200<sup>a</sup> | ctggttagCTCTAGA . | .           |
| 2    | 303             | .GATTAGAgtagaatt | 3843              | ttctcaagGTGIGAC . | .I          |
| 3    | 112             | .ACTTTAAGttagaaat | 4096 | tgatcaagGCTCCTGA . | .O          |
| 4    | 213             | .GTTTTTGTGttagttt | 1320 | cctgacagAAATGAG . | .I          |
| 5    | 110             | .CTTAAAGGtttgtat | 4352 | ctattagGCTTACC . | .I          |
| 6    | 93              | .TACAAAGGttaaagctt | 704  | atttttagGCTGGGC . | .II         |
| 7    | 138             | .CCTCAGAGGtaaaaca | 2147 | tacaacagATAATCT . | .II         |

<sup>a</sup> The genomic sequence is from an unfinished hgtgs clone (GenBank™ accession number AL137848.4). All exons except number 1 are in the same contig of this clone. The size of intron 1 thus awaits completion of the sequence.

<sup>b</sup> Includes 3’-untranslated sequence.

This notion is further supported by the detection of an oxidized methionine in these peptides, matching the presence of a methionine residue in the amino-terminal peptide sequence (amino acids 33–59). Assuming that we have 13 consecutive Asp residues in the sequence, a corresponding glycosylmass of 1329 and 1038 is obtained for the peptide with and without sialic acid, respectively.

Another approach to verify that the protein contains the proposed sequence is to measure its intact mass. Since this could not be obtained directly from the partially purified asporin sample, we investigated the possibilities of extracting the protein directly from the gel avoiding the problems of contamination, e.g. by fibromodulin. An intact mass of 43,200 ± 500 Da was derived for the nontreated protein after extraction from the gel. The large mass deviation was due to difficulties in assigning the peak maximum caused by the heterogeneity of glycosylation together with the fact that formylation reactions usually occur upon extraction of the gel with formic acid. This results in peak broadening with a shift toward higher masses.

The theoretical mass of the mature protein with 13 consecutive Asp residues in the amino-terminal is 39,609 Da. However, by adding a mass for an N-linked oligosaccharide of ~2,000 Da and for an O-linked oligosaccharide of ~1,000 Da to the theoretical mass, the observed mass range indicates that the suggested amino-terminal is present in the tissue.

**DISCUSSION**

Asporin is a new member of the LRR protein family most closely related to decorin and biglycan. The four amino-terminal cysteines show the C-X<sub>2</sub>-C-X<sub>2</sub>-C-X<sub>2</sub>-C pattern typical of the class I LRR proteins. Furthermore, like decorin and biglycan, asporin contains a putative propeptide with a conserved cleavage site corresponding to the recognition sequence for bone morphogenetic protein-1, i.e. the enzyme shown to cleave probiglycan (38). Moreover, the sequence and length of the LRR repeats of asporin are more similar to those of decorin and biglycan than to other members of the ECM LRR-repea proteins. This is also evident from the evolutionary tree of the LRR proteins. Finally, like the decorin (40) and biglycan (41) genes, the human asporin gene is divided into 8 exons. The introns are inserted in the coding sequence at exactly the corresponding positions to those of decorin and biglycan.

The eight exons of the human asporin gene span 26 kilobases on chromosome 9q31.1–32. It is not yet clear if the asporin gene also contains an additional alternative spliced exon 1, as does the decorin gene (40). Like in the biglycan gene (41), no TATA box was found in the 5’-flanking region of exon 1 of asporin. We did, however, locate a number of recognition sites for transcrip-
tion factors in the 400 base pairs upstream of the deduced transcription start.

The extracellular matrix LRR protein genes appear to be organized in clusters of four. Decorin, lumican, keratocan, and epiphycan (class I, II, II, and III, respectively) map to chromosome 12q23. Asporin, osteoadherin, and mimecan (class I, II, and III, respectively) are found on chromosome 9q32. Interestingly, **ECM2**, a gene encoding a LRR protein containing an amino-terminal von Willebrand factor repeat, has been located between asporin and osteoadherin (42). Fibromodulin, PRELP, and opticin (class II, II, and III, respectively) locate to chromosome 1q32. Biglycan (class I) is unique in not being part of such a cluster but rather found in isolation on chromosome X. It appears that several duplications have occurred during evolution, resulting in the clustered organization of the LRR genes. The biglycan gene may then have relocated to chromosome X. Alternatively, four more LRR protein genes remain to be identified, one on chromosome 1 and three on chromosome X.

The asporin amino-terminal extension is unusual in containing an extended stretch of aspartate residues. Messenger RNAs very similar to asporin were recently identified in the zebrafish and the cichlid *Oreochromis* (43). These proteins (referred to as biglycan-3 by the authors) belong to the class I LRR proteins based on the amino-terminal cysteine spacing and the amino acid sequence of the LRRs. Stretches of aspartic acid residues in the amino-terminal extensions clearly identify these proteins as fish homologues of asporin. Unlike in human asporin, the mouse and the two fish polyaspartate sequences are inter-

---

**Fig. 8. Expression of asporin mRNA in different human tissues.** A Human RNA Master Blot was hybridized with the 32P-labeled asporin cDNA probe. A, autoradiography of the hybridization signal. B, diagram showing the type and position of poly(A)⁺ RNAs. C, PhosphorImager quantification of the hybridization signal.
structures. Whether the variability results from different tissue compartments or whether all variants are present at a given location in the tissue is not known. The fact that the protein analyzed was extracted from a pool of tissue from several donors can of course also contribute to the observed variation. Additional data on structure-function relationships in relation to oligosaccharide variability may provide important information on the role of such substituents.

The protein contains an additional putative glycosylation site (Ser54 in the human sequence) that appears to be substituted with an O-glycosidically linked oligosaccharide. O-Linked glycosylation has previously been described in the amino-terminal extension peptides of epiphycan and opticin. It remains to be elucidated whether oligosaccharide substituents in this region may modify the properties of this structure.

The role of the propeptides of class I LRR proteins is unclear. It has been implied that this sequence affects the glycosaminoglycan structure of decorin (44) and biglycan (45). The presence of a conserved propeptide in asporin, which does not contain any glycosaminoglycan attachment consensus sequence, suggests that the propeptides may have other primary functions.

Asporin mRNA is expressed in a number of different tissues, including articular cartilage. Beside cartilage, the highest expression levels were found in other tissues with high content including articular cartilage. Beside cartilage, the highest expression levels were found in aorta and uterus, suggesting expression by smooth muscle cells. Indeed, intermediate expression levels were detected in other tissues with high content (large abundance) of smooth muscle cells.

The functional implications of the protein are not clear. However, in attempts to separate the protein from fibromodulin in a collagen co-precipitation assay, both proteins appeared to bind to collagen. This would be in analogy with properties of other LRR proteins of the type containing 10–11 repeats where most members have been shown to bind tightly to collagen with derivatives.
equilibrium dissociation constants in the nanomolar range.

Extended stretches of aspartic acid residues like in asporin are unusual. Osteopontin, however, a prominent component of the mineralized extracellular matrix of bone and teeth, has a polyaspartic acid sequence in the center of the core protein (46). This protein binds to hydroxyapatite and may have a role in bone mineralization (47). Whether asporin also has a role in mineral deposition is not clear. In this context the increased synthesis of the protein in early osteoarthritis is of interest, particularly in view of the frequently altered deposition of calcium phosphate in this disorder (reviewed in Ref. 48). A particularly in view of the frequently altered deposition of hydroxyapatite in early osteoarthritis is of interest, particularly in view of the frequently altered deposition of calcium phosphate in this disorder (reviewed in Ref. 48). A particularly

The other two members of the family decorin and biglycan have been shown to bind growth factors, particularly transforming growth factor-β (52). Whether also asporin has this capacity remains to be demonstrated.
46. Oldberg, Å., Franzén, A., and Heinegård, D. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 8819–8823
47. Hunter, G. K., Kyle, C. L., and Goldberg, H. A. (1994) Biochem. J. 300, 723–728
48. Jimenez, S., Williams, C., and Karasik, D. (1998) in Osteoarthritis (Brant, K., Doherty, M., and Lohmander, L., eds) pp. 31–49, Oxford University Press, Oxford
49. Oldberg, Å., Franzén, A., and Heinegård, D. (1988) J. Biol. Chem. 263, 19430–19432
50. Hunter, G. K., and Goldberg, H. A. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 8562–8565
51. Hunter, G. K., and Goldberg, H. A. (1994) Biochem. J. 302, 175–179
52. Hildebrand, A., Romaris, M., Rasmussen, L. M., Heinegård, D., Twardzik, D. R., Border, W. A., and Ruoslahti, E. (1994) Biochem. J. 302, 527–534