Cloning and Deduced Amino Acid Sequence of a Novel Cartilage Protein (CILP) Identifies a Proform Including a Nucleotide Pyrophosphohydrolase*

(Received for publication, March 13, 1998, and in revised form, May 2, 1998)

Pilar Lorenzo, Peter Neame†, Yngve Sommarin‡, and Dick Heinegård§

From the Department of Cell and Molecular Biology, Section for Connective Tissue Biology, University of Lund, P.O.Box 94, S-221 00 Lund, Sweden and the §Department of Biochemistry and Molecular Biology, Shriners Hospital for Crippled Children, Tampa, Florida 33612

The cDNA cloning and expression in vitro and in eukaryotic cells of a novel protein isolated from human articular cartilage, cartilage intermediate layer protein (CILP) is described. A single 4.2-kilobase mRNA detected in human articular cartilage encodes a polypeptide of 1184 amino acids with a calculated molecular mass of 132.5 kDa. The protein has a putative signal peptide of 21 amino acids, and is a proform of two polypeptides. The amino-terminal half corresponds to CILP (molecular mass of 78.5 kDa, not including post-translational modifications) and the carboxyl-terminal half corresponds to a protein homologous to a porcine nucleotide pyrophosphohydrolase, NTPPHase (molecular mass of 51.8 kDa, not including post-translational modifications). CILP has 30 cysteines and six putative N-glycosylation sites. The human homolog of porcine NTPPHase described here contains 10 cysteine residues and two putative N-glycosylation sites. In the precursor protein the NTPPHase region is immediately preceded by a tetrapeptide conforming to a furin proteinase cleavage consensus sequence. Expression of the full-length cDNA in a cell-free translation system and in COS-7 or EBNA cells indicates that the precursor protein is synthesized as a single polypeptide chain that is processed, possibly by a furin-like protease, into two polypeptides upon or preceding secretion.

Articular cartilage is a tissue that varies in composition with site, e.g. from the surface to the deep zone and also from the pericellular environment close to the cells to the more distant interterritorial matrix (1, 2). The major constituents of the matrix are collagens (reviewed in Ref. 3) and large aggregating proteoglycans that contain large numbers of chondroitin sulfate chains (reviewed in Ref. 4).

A third set of molecules are the non-collagenous glycoproteins including several members of the family of leucine-rich repeat proteins and the thrombospondins. The leucine-rich repeat proteins include decorin, biglycan, fibromodulin, and lumican, all with the capacity to bind to collagen (for Refs. see Ref. 1). These molecules are found along surfaces of collagen fibrils in the tissue. Other members of this family that do not bind to collagen include chondroadherin (5) and PRELP (6).

The thrombospondin family includes the pentameric COMP, which is one of the more abundant cartilage matrix proteins (7, 8). Furthermore, thrombospondin-1 has been identified in articular cartilage (9).

There are also a number of other proteins that may not be part of the described families. One such protein is the cartilage intermediate layer protein (CILP), previously identified and purified from human articular cartilage (25). In most cases the structure and function of these proteins are not known. However, detailed knowledge of their structure, function, and variability in normal conditions as well as in disease is of fundamental importance to define mechanisms and events occurring in joint disease. Interestingly, in joint disease, such as osteoarthritis there are major alterations in the composition of the cartilage extracellular matrix. These changes reflect an altered homeostasis, probably a reason for the defective repair (1). Our preliminary studies have shown that in early osteoarthritis there are a set of proteins, in which synthesis and content are increased in early osteoarthrosis cartilage. One of these proteins is CILP. The present paper describes the cloning and expression in vitro in cell-free and in cells of CILP. The mRNA for CILP was found to also code for a protein homologous to an ectonucleotide pyrophosphohydrolase recently cloned and partially sequenced from porcine chondrocytes (10). Our data indicate that the precursor protein is synthesized by human chondrocytes and processed in close association to secretion.

MATERIALS AND METHODS
Tissue Extraction and Protein Purification

The experimental procedures for the isolation and purification of CILP were as described (25).

Protein Sequencing

Peptide Fragment Generation and Sequencing—Proteolytic digestion with Lys-C was performed at enzyme: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.

* This work was supported by the Swedish Medical Research Council, Konung Gustaf V’s 80-årsfond, Greta och Johan Kock’s stiftelse, Axel och Margaret Ax:son Johnsons stiftelse, Alfred Österlunds Stiftelse, Konung Gustaf V’s 80-årsfond, och Margaret Ax:son Johnsons stiftelse, Alfred Österlunds Stiftelse, Konung Gustaf V’s 80-årsfond, och Margaret Ax:son Johnsons stiftelse, Alfred Österlunds Stiftelse, Konung Gustaf V’s 80-årsfond, and the Medical Faculty, Lund University. The costs of publication of this paper have been 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) AF053408. ¶ Present address: Wieslab AB, Ideon, S-223 70 Lund, Sweden.

† To whom correspondence should be addressed: Dept. of Cell and Molecular Biology, Section for Connective Tissue Biology, University of Lund, P.O.Box 94, S-221 00 Lund, Sweden. Tel.: 46-46-222-85-71; Fax: 46-46-211-34-17; E-mail: Dick.Heinegard@med kem.lu.se.

‡ P. Lorenzo, M. Bayliss, and D. Heinegård, unpublished data.
sequencer with on-line analysis of phenylthiohydantoin-derivatives on an Applied Biosystems 120A micro bore HPLC.

General—All the molecular biological procedures, including agarose gel electrophoresis, restriction enzyme digestion, ligation, bacterial transformation, and DNA sequencing, were performed by standard methods (13).

RNA Extraction—Human 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 was extracted essentially as described by Adams et al. (14) but omitting the cesium trifluoroacetate ultracentrifugation step. Instead, the RNA was precipitated twice with 1 volume of 100% isopropanol, then further purified with a RNA easy kit (QIAGEN Inc., Chatsworth, CA) according to the manufacturers’ protocol. The mRNA was purified from this preparation using the Oligotex mRNA kit (QIAGEN Inc.) by following the manufacturers’ protocol.

cDNA Probe Generation and Library Screening—Total RNA isolated from human articular cartilage was used as the template for the synthesis of the first-strand cDNA. Total RNA (20 μg) was reverse-transcribed using 400 units of Moloney murine leukemia virus reverse transcriptase (Life Technologies Inc) in a 50-μl reaction mixture containing 100 ng of oligo(dT), and 300 ng of random hexamer primers, 1 mM dNTP, for 1 h at 37 °C (15).

PCR amplification was performed with degenerate primers (forward and reverse, 1 μM each of the gene-specific primers (5'-RACE, 5 μM of cDNA, 0.5 mM dNTP mixture, 1 unit of Taq polymerase, 5 mM dNTP mixture, 5 mM MgCl2, and reverse, 1 μM each), 5 μl of cDNA, 0.2 mM dNTP mixture, 1 unit of Taq polymerase, 5 μl of 10 × Taq polymerase buffer, 1.5 mM MgCl2, in a 50-μl reaction mixture. After an initial denaturation step at 94 °C for 3 min, 35 cycles of amplification were performed in a Hybrid OmniGene thermocycler at the following conditions: 94 °C for 30 s, 48 °C for 30 s, 72 °C for 2 min. Ten percent of the products from the first amplification were re-amplified for an additional 25 cycles under the same cycling conditions. The PCR products were analyzed by electrophoresis on a 1% agarose gel (NuSieve GTG, FMC Corp.). The main product (800 bp) was isolated, and subcloned into pCR-Script Amp SK(+) (Stratagene, La Jolla, CA) and sequenced in both directions by using the standard double-strandedideoxy chain termination method.

The deduced amino acid sequence of the 800-bp cDNA fragment contained some of the peptide internal sequences. The fragment was used as a probe to screen a ZAP II cDNA library made from human chondrocytes, kindly provided by Dr. Michael Bayliss (16). Approximately 1 × 106 plaque forming recombinants were screened. One positive clone was found (named 92C-1). The plasmid pHuScript II KS(+) containing the insert was rescued from the ZAP vector by the use of in vitro excision as described in the ZAP-cDNA® synthesis kit (Stratagene, La Jolla, CA). The plasmid was then digested with XbaI/XhoI and the cDNA size (2.3 kb) was determined by electrophoresis on a 1% agarose gel. Restriction fragments were generated with BamHI and subcloned into a pCR-Script Amp SK(+) (+). Sequencing was performed as described above, compression artificats were resolved using the dttkit (U. S. Biological Corp., Cleveland, OH). The sequencing primers used included T3 and T7 universal primers and internal sequencing primers.

The 2.3-kb clone was used to generate a 314-bp PCR probe with primers designed toward the 3'-end of the insert. The probe was used for rescoring the human chondrocyte ZAP II cDNA library as described. One 1.6-kb clone (named 92C-2) was isolated, subcloned, and sequenced. Sequence analyses were carried out using the PGCgene Program package (Intelligenetics). Similarity searches were performed by the use of E-mail service at the National Center for Biotechnology Information using the BLAST program.

Isolation of cDNA Ends—The Marathon cDNA amplification kit (CLONTECH) was used to obtain a library of adaptor-ligated double-stranded cDNA from human articular cartilage. One microgram of poly(A)+ RNA was used as template for the first strand synthesis as described by CLONTECH in the manual for the Marathon cDNA amplification kit, 5’- and 3’-RACE. The 5’- and 3’-ends were PCR amplified using 3 μl of the library as the template with the primer supplied by CLONTECH (AP1) and 0.2 μM each of the gene-specific primers (5’-RACE, 5’-GTACAATGGGGGCGCAATTCATTGAC-3’ from 835 to 859 bp and 3’-RACE, 5’-TCAAGTGCAGGGCCATCT-3’ from 2517 to 2541 bp (Fig. 2), 0.2 mM dNTP mixture, 1 unit of Taq polymerase, 5 μl of 10 × reaction buffer, and 1.25 units of cloned Pfu DNA Polymerase (Stratagene) in a 50-μl reaction volume. After an initial denaturation step at 94 °C for 1 min, 30 cycles of amplification were performed using the following conditions: 94 °C for 45 s, 65 °C for 45 s, 72 °C for 4 min. The PCR products were analyzed on a 1% agarose gel, isolated, and subcloned into pCR-Script Amp SK(+). Sequencing of both strands was performed by AmpliTaQ DNA Polymerase, FS (Perkin-Elmer) using fluorescent dye termination sequencing chemistry and analysis of the reaction products on an Applied Biosystems 373A DNA sequencer. The sequencing primers used included T3 and T7 universal primers and internal sequencing primers.

Isolation of Genomic Clones—Three genomic clones were subcloned in the bacteriophage P1 were obtained from Genome Systems, Inc. Screening Services (St. Louis, MO). One clone containing the gene for CILP was used for isolation of a Xhol/EcoRI fragment containing the 3’-end. The fragment, GXRI, was subcloned into pBluescript II KS(+) and used for the construction of the expression vectors (Fig. 1).

Northern Blotting—Ten micrograms of total RNA isolated from human articular cartilage was electroblotted on a 1% formaldehyde agarose gel, transferred to a nitrocellulose filter (NitroPure, Micron Separation), and hybridized with randomly-primed [32P]dCTP probe of 1585-bp BsmI/EcoRI fragment from the 92C-1 clone. After hybridization, the membrane was washed twice at 42 °C for 15 min with 2 × SSC, 0.1% SDS for 15 min, then with 1 × SSC, 0.1% SDS, and finally with 0.1 × SSC, 0.1% SDS, and exposed to x-ray film (X-Omat AR, Kodak).

Expression Constructs

A full-length cDNA for expression studies in three vectors was constructed as schematically depicted in Fig. 1.

In pBluescript II KS(+) (Stratagene), clone 92C-2 was digested with XhoI, blunted with T4 DNA polymerase, extracted with phenol, and then digested with AvrII. The released fragment (1081 bp) was separated by agarose electrophoresis isolated and ligated into the clone 92C-1, previously digested with SmaI/EcoRI. The insert was then re-amplified and re-cloned into a XhoI/EcoRI fragment of the human CILP gene containing the 3’-end. The use of this genomic fragment (606 bp) was for convenience. It did not contain any intron and its sequence was identical to the 3PM clone except for the poly(A) tail. The 5’-end was amplified by PCR using the 5’ Marathon Clone (5PM) as a template to generate an XhoI/XhoI primer at the position 97 with a ClaI site introduced and a reverse primer at the position 843 at the Sou I site. The generated 746-bp PCR fragment was ligated in the recombinant clone.

To make a construct in the pSVL SV40 Late Promoter Expression Vector (Pharmacia Biotech), the pBluescript II KS(+) bearing the full-length CILP cDNA was digested with SalI/Sau3AI. The generated fragment was ligated into the pSVL SV40 previously digested with XhoI/Sau3AI. The 716-bp fragment ClaSI/AvrII was then added. In pCEP 4 (Inviogene) the full-length CILP cDNA was ligated into the NorI/KpnI site. All three vectors containing full-length CILP cDNA were checked by standard sequencing with forward and reverse primers that covered the ligation sites.

CILP Expression

In Vitro Cell Free Transcription/Translation Assay—The plasmid plasmid pHuScript II KS(+) with the 5’-end fragment cloned into pHerpes (a gift from P. Lorenzo, P. Åman, Y. Sommarin, and D. Heinegård, unpublished data) was constructed as schematically depicted with TNT® T7 Polymerase (TNT® T7 Coupled Reticulocyte Lysate System, Promega). Supercoiled plasmid (500 ng) was used according to the manufacturer’s protocol. The 25-μl reaction mixture contained 12.5 μl of TNT rabbit reticulocyte lysate, TNT reaction buffer, T3 RNA polymerase, RNasin® 20 Units, amino acid mixture without methionine, supplemented with [35S]methionine as the radioactive precursor (1000 Ci/mmol). After 90 min incubation at 30 °C, labeled polypeptides were separated from free amino acids by ethanol precipitation (10 volumes of ethanol containing 50 mM sodium acetate). Precipitates were re-suspended in electrophoresis sample buffer (2% SDS, 0.125 M Tris-HCl, pH 6.5, 0.002% bromophenol blue, and 20% glycerol) without or with 10% 2-mercaptoethanol, boiled at 100 °C for 4 min, and electrophoresed on a 4–16% gradient SDS-polyacrylamide gels according to Laemmli (12) and visualized by fluorography (17). Control reaction with the pBluescript II KS(+) plasmid only was run simultaneously.

Cell Transfections

Transient Transfection—The monkey kidney-derived cell line, COS-7, was grown in F-12/Dulbecco’s medium, supplemented with 10% fetal calf serum, streptomycin (0.1 mg/ml), and penicillin (100 units/ml) at 37 °C in a CO2 incubator. Semiconfluent cells were transfected by using the LipofectAMINE method, as described by the manufacturer (Life Technologies, Inc.), with either 6 μg of pSVL SV40 vector containing the expression construct or 6 μg of pSVL SV40 vector only. After
Advancing the understanding of CILP precursor structure

72 h post-transfection the cells were labeled with [35S]methionine (1000 Ci/mmol) in a methionine-free medium. After 8 h labeling the medium was removed, the cells were washed with phosphate-buffered saline, and lysed with RIPA buffer (0.5% Nonidet P-40, 0.5% Tween 20, 0.5% deoxycholic acid, 0.15 M NaCl, 10 mM KCl, 1 mM EDTA, and 20 mM Tris-HCl, pH 7). Medium and cell extract were immunoprecipitated with polyclonal rabbit antiserum raised against CILP (25). The precipitated products were electrophoresed on a 4–16% gradient SDS-polyacrylamide gel and visualized by fluorography. For control, the cells were transfected with the respective vector lacking the cDNA insert.

**Stable Transfection**—Human embryonic kidney cells, 293-EBNA (In-vitrogen), were grown in Dulbecco’s modified Eagle’s medium supplemented with 10% fetal calf serum, streptomycin (0.1 mg/ml), penicillin (100 units/ml), and 250 μg/ml Geneticin. When the cells were semi-confluent they were stably transfected by using the LipofectAMINE method with either 5 μg of the pCEP4 vector containing the expression construct or 5 μg of the pCEP4 vector only. The cells were selected with 300 μg/ml hygromycin. After 1 week of transfection the expression of CILP was analyzed by immunoblotting the medium.

**Antibodies**—A synthetic peptide (EDRTFLVGNLEIRERRLFNC), apart from an extra COOH-terminal cysteine residue with the SPDP cross-linker (Saveen Biotech AB, Malmo, Sweden), and used to generate a rabbit polyclonal antiserum.

**RESULTS**

**Protein Sequencing**—We were unable to determine the NH2-terminal of CILP, suggesting that it may be blocked. Additionally, the NH2-terminal may be rather heterogeneous as a result of proteolytic cleavage.

After Lys-C digestion of intact protein, peptides were separated by reversed phase HPLC, peaks were collected by hand and 14 were analyzed, only 11 peaks gave readable sequences. 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 cysteines. Eleven of the sequenced peptides could be aligned against the translated cDNA sequence (Table I and Fig. 2). It should be noted that all the peptides sequenced were located within the first 680 amino acids.

One peptide (3306, Table I) contained a blank cycle which on subsequent analysis of the translated cDNA sequence proved to be the asparagine of a carbohydrate N-linkage site. This site is therefore likely to contain an N-linked oligosaccharide.

**Reverse Transcriptase-PCR**—Total RNA isolated from human articular cartilage was used as the template for synthesis of the first-strand cDNA with both oligo(dT) and random hexamers as primers. The cDNA was used for PCR amplification with degenerate primers based on the amino acid sequences LHVQPD, EAMETN, and PYMVMNP. A 830-bp PCR fragment was amplified, isolated, and ligated into pCR-Script Amp SK (+). The fragment covered nucleotides between 926 and 1756 bp (Fig. 2). The translated sequence contained several of the determined amino acid sequences (Table I, Fig. 2).

**Nucleotide and Deduced Amino Acid Sequences**—Two overlapping clones (92C-1 and 92C-2, Fig. 1) were isolated by screening with the 830-bp PCR fragment approximately 1 × 106 clones from a human articular chondrocyte AZAP-cDNA library. These clones encode most of the cDNA sequence for CILP (Fig. 1). However, the first methionine did not conform to the Kozak consensus and no signal peptide and no poly(A)-tail was detected. No full-length clone was found after extensive rescreening of the cDNA library. The 5′- and 3′-ends were generated by the RACE technique. The 5′-end obtained was a fragment of 988 bp (5PM, Fig. 1) that contained the first ATG codon and a putative signal peptide. The 3′-end obtained was a cDNA fragment of 730 bp (3PM, Fig. 1) that contained a poly(A) tail. An overview of the cloning strategy is given in Fig. 1. All cDNA fragments were sequenced in full. The resulting nucleotide and deduced amino acid sequences are shown in Fig. 2. The identity of the clones with the isolated protein was verified by the partial
amino acid sequences from the protein, as indicated in Fig. 2. The full-length cDNA encoded a total of 4175 bp with an open reading frame starting with the first ATG codon at the position 130 and ending at position 3681 with a TAA codon. The sequence flanking the first ATG codon conforms to the consensus sequence for initiation of translation proposed by Kozak (18), (A/G)XXAUG.

The open reading frame coded for a protein of 1184 amino acids, with a calculated molecular mass of 132.5 kDa. Prediction of the eukaryotic secretory signal sequence according to Von Heijne (19) indicated a putative cleavage site between amino acids 21 and 22. The protein contains 40 cysteine residues which are likely to be involved in disulfide bond formation. Many of these are located toward the NH2-terminal part of the molecule. There are eight consensus sites for N-glycosylation, and the calculated isoelectric point is 8.31.

FIG. 2. Nucleotide and translated protein sequences. The putative signal peptide is boxed. The arrows indicate the putative cleavage site. The underlined amino acids represent peptide sequenced (Table I). The cysteines are circled, and the putative N-glycosylation sites are marked with a star. The stop codon is marked and the polyadenylation signals are underlined. The putative furin cleavage site is indicated by the shadowed area. The sequence homologous to porcine NTPPhase is shown in italics beginning at residue 683. The Type 1 thrombospondin repeat sequence consensus is identified at residues 131–157, with conserved residues underlined.

Amino acid sequences from the protein, as indicated in Fig. 2. The full-length cDNA encoded a total of 4175 bp with an open reading frame starting with the first ATG codon at the position 130 and ending at position 3681 with a TAA codon. The sequence flanking the first ATG codon conforms to the consensus sequence for initiation of translation proposed by Kozak (18), (A/G)XXAUG.

The open reading frame coded for a protein of 1184 amino acids, with a calculated molecular mass of 132.5 kDa. Prediction of the eukaryotic secretory signal sequence according to Von Heijne (19) indicated a putative cleavage site between amino acids 21 and 22. The protein contains 40 cysteine residues which are likely to be involved in disulfide bond formation. Many of these are located toward the NH2-terminal part of the molecule. There are eight consensus sites for N-glycosylation, and the calculated isoelectric point is 8.31.

A 3′-noncoding region of 493 bp downstream of the stop codon contained three consensus sequences for polyadenylation, 5′-AATAAA-3′, at positions 3740–3745, 4003–4008, and 4011–4016 followed by a short poly(A) stretch located 17 bp after the last polyadenylation signal.

Similarity to other Proteins—Nucleotide and amino acid similarity searches were made against the GenBank data bases at The National Center for Biotechnology Information, using the BLAST network service (20). Toward the NH2 terminus of the molecule there is a stretch of 27 amino acids, from 131 to 157, shown in Fig. 2, which is related to the Type 1 thrombospondin repeat consensus proposed by Kozak (18), (A/G)XXAUG. The open reading frame coded for a protein of 1184 amino acids, with a calculated molecular mass of 132.5 kDa. Prediction of the eukaryotic secretory signal sequence according to Von Heijne (19) indicated a putative cleavage site between amino acids 21 and 22. The protein contains 40 cysteine residues which are likely to be involved in disulfide bond formation. Many of these are located toward the NH2-terminal part of the molecule. There are eight consensus sites for N-glycosylation, and the calculated isoelectric point is 8.31.

A 3′-noncoding region of 493 bp downstream of the stop codon contained three consensus sequences for polyadenylation, 5′-AATAAA-3′, at positions 3740–3745, 4003–4008, and 4011–4016 followed by a short poly(A) stretch located 17 bp after the last polyadenylation signal.

Similarity to other Proteins—Nucleotide and amino acid similarity searches were made against the GenBank data bases at The National Center for Biotechnology Information, using the BLAST network service (20). Toward the NH2 terminus of the molecule there is a stretch of 27 amino acids, from 131 to 157, shown in Fig. 2, which is related to the Type 1 repeats of thrombospondin (21). The COOH terminus of the molecule showed very high similarity (90%) to a porcine ectonucleotide pyrophosphohydrolase, NTPPhase (10). Based on the biochemical characterization of CILP, which has an apparent molecular mass of 92 kDa (25), and on the NH2-terminal sequence reported for a catalytically active proteolytic fragment of the porcine NTPPhase (22), it appears that CILP is synthesized as a proform that also contains a protein homologous to a porcine enzyme with NTPPhase activity. In fact the precursor protein has a consensus sequence for precursor cleavage catalyzed by furin (23), i.e. RX(K/R)R, from positions 700 to 703 (RNKR).
The processing of the precursor protein was studied by expressing the cDNA construct transiently and stably, in two different cell lines. When the protein was expressed in vitro by using a cell-free transcription-translation system in the presence of [35S]methionine, a single labeled peptide with an apparent molecular mass of 123 kDa was detected (Fig. 4A). The processing of the precursor protein was studied by expressing the cDNA construct transiently and stably, in two different cell lines.

With COS-7 cells, analysis by immunoprecipitation of cell extracts and medium with polyclonal antibodies against CILP showed that the precursor protein was secreted to the medium (Fig. 5B). The polyclonal antibody against CILP recognized a major component with a molecular mass of 92 kDa (N1, Fig. 5A) or against the synthetic peptide EDRTFLVGNLEIRERRLFNC (Fig. 5B), corresponding to the NH$_2$-terminal of NTPPHase. A, pBluescript II KS(+) with the expression construct was in vitro transcribed-translated by the T7 RNA polymerase and visualized by fluorography. Lane 1, control plasmid only. Lane 2, plasmid with the expression construct. B and C, COS-7 cells were transfected with the pSVL vector containing the expression construct and grown in a serum-free medium containing [35S]methionine. After labeling, the cell extract (B) and the medium (C) were immunoprecipitated with the antibodies against CILP. The immunoprecipitated material was resolved on a 4–16% gradient SDS-polyacrylamide gel and visualized by fluorography. Lane 1, transfection with the vector only. Lane 2, transfection with the vector containing the expression construct. The molecular weight standards are indicated on the left.

The processing of the precursor protein was further analyzed by immunoblotting of the medium from 293-EBNA cells after stable transfection and expression (Fig. 5A). The polyclonal antibodies against CILP recognized a major component with a molecular mass corresponding to isolated CILP, C1 (Fig. 5A). An additional immunoreactive component of lower molecular mass (e.g. C2, 62 kDa) was present in a lower proportion (Fig. 5A). The lower molecular mass component is not a result of different glycosylation of CILP, since after digestion with N-glycanase both proteins similarly changed their electrophoretic mobility (data not shown).

It is likely then that this form is generated by alternative cleavage of the precursor protein. With antibodies generated against the synthetic peptide EDRTFLVGNLEIRERRLFNC (Fig. 5B), corresponding to the NH$_2$-terminal of NTPPHase, a molecular mass of 61 kDa (22) which is in agreement with the calculated molecular mass of 51.8 kDa of the translated sequence from base 2173 to base 3552 and assuming carbohydrate substitutions at two putative N-glycosylation sites.

The processing of the precursor protein was further analyzed by using a cell-free transcription-translation system in the presence of [35S]methionine. The in vitro synthesized polypeptide was resolved on a 4–16% gradient SDS-polyacrylamide gel and visualized by fluorography. Lane 1, control plasmid only. Lane 2, plasmid with the expression construct. B and C, COS-7 cells were transfected with the pCEP 4 vector containing the expression construct and grown in a serum-free medium containing [35S]methionine. After labeling, the cell extract (B) and the medium (C) were immunoprecipitated with the antibodies against CILP. The immunoprecipitated material was resolved on a 4–16% gradient SDS-polyacrylamide gel and visualized by fluorography. Lane 1, transfection with the vector only. Lane 2, transfection with the vector containing the expression construct. The molecular weight standards are indicated on the left.
components that are secreted to the medium with no detectable

cursor protein is fully cleaved by the 293-EBNA cells into two

for CILP. In summary, these experiments show that the pre-

cursor protein is cleaved by the 293-EBNA cells into two

components that are secreted to the medium with no detectable

precursor present.

**DISCUSSION**

Isolation and peptide mapping of CILP enabled us to define

probes and primers for characterization of the full-length

cDNA coding for a novel extracellular matrix macromolecule.

This protein acts as a precursor for two different proteins.

The cloning of a cDNA of 4175 bp that is transcribed to a

single mRNA species of the same size indicates that the mRNA

encodes a larger protein than expected for CILP. This was

verified by the complete deduced amino acid sequence. The

polyepitope consists of 1184 amino acids with a calculated

molecular mass of 132.5 kDa. Thus, the precursor protein is

made of CILP (molecular mass of 78.5 kDa, without post-

translational modifications) and a protein homologous to a

nucleotide pyrophosphohydrolase (NTPPHase) described in porcine

chondrocytes (10). The proteolytic fragment of the porcine

NTPPHase was suggested to be derived from a larger matrix

vesicle-associated protein of apparent molecular mass of 127 kDa

(22). In the present study the human homolog of the porcine

NTPPHase has 460 amino acid residues with a molecular mass

of 51.8 kDa, without post-translational modifications.

In the precursor protein described here, the two proteins

appear to be generated after proteolytic cleavage at a tetrapep-

tide conforming to a furin proteinase cleavage consensus se-

quence RX(K/R)R (RNKR, spanning amino acids 700 to 703 in

Fig. 2) with basic amino acids in positions −4, −2, and −1

to the cleavage site (23). This motif is conserved in a number of

constitutively processed precursors. Whether the presently

described precursor protein is cleaved by furin or

furin-like proteases synthesized by human chondrocytes

remains to be established. The precursor protein is synthesized

as a single polypeptide chain as shown by in vitro cell-free

expression. The precursor protein is cleaved to the two proteins in

the cell systems used for expression. The presence of small

amounts of the cleaved products in the COS-7 cell extract upon

transient expression indicates that the cleavage occurs already

intracellularly, probably immediately prior to secretion. Both

products, CILP and the homologous NTPPHase, are fully

secreted from the cells as shown in the two cell systems used for

the study of the expression of the protein. Only in the COS-7

cells a proportion of the precursor protein is secreted into the

medium, possibly a result from a limited capacity to process the

high-level of protein expressed in these cells after transfection.

More complete processing was accomplished in the EBNA

cells, where the proform of the protein was not secreted into the

medium. After processing, a product corresponding to CILP

isolated from human articular cartilage (92 kDa) and a smaller

variant (82 kDa) were detected by the polyclonal antibody

against CILP. The smaller product may be formed by an alter-

native cleavage site used by the proteases in the expression

system. Since this smaller form is not detected in explants of

human articular cartilage it either has a faster turnover or the

processing enzymes in human articular chondrocytes are more

specific than those in the EBNA cells. Alternative putative

cleavage site spanning amino acids 538–539 or 551–553 (Fig.

2) might be targeted in the cell lines used.

CILP is processed by the cells into a single polypeptide chain with a calculated molecular mass of 78.5 kDa. It is slightly

basic with an isoelectric point of 8.15. It contains 6 putative

N-glycosylation sites which is consistent with a molecular mass of 92 kDa, the apparent molecular mass of the protein purified from human articular cartilage (25) after post-translational modifications. The protein contains 30 cysteine residues mostly

distributed toward the NH₂-terminal half of the molecule. Due to

the special role of cysteine residues for protein structure, the

high content of these amino acids indicates a high degree of

intrachain disulfide cross-linking.

The NH₂-terminal domain of CILP containing the Type 1

thrombospondin-like repeat and 16 cysteines is associated with a strikingly hydrophilic region from amino acids 130–225. Following this, a relatively exposed region may lead to another compact, disulfide-bonded region between residues 354 and 445. Another extended region with no cysteines leads to the very basic furin-type cleavage site.

The lack of similarities between CILP and other proteins of
defined structure neither allowed us to predict a secondary

structure nor any remarkable structural feature or function.

There is a homology to the Type 1 repeats of thrombospondin in the consensus sequence for heparin binding (24), from residues

131 to 135 (WSPWS), but we have no experimental evidence that CILP actually binds to heparin.

The human homolog of the nucleotide pyrophosphohydrolase

(NTPPHase) that forms the COOH-terminal of CILP is made of

460 amino acid residues, with a calculated molecular mass of

51.8 kDa, and an isoelectric point 8.73. It has 10 cysteine

residues and two putative

N-glycosylation sites which is consistent with a molecular mass

of 92 kDa, the apparent molecular mass of the protein purified from human articular cartilage (25) after post-translational modifications. The protein contains 30 cysteine residues mostly
distributed toward the NH₂-terminal half of the molecule. Due to

the special role of cysteine residues for protein structure, the

high content of these amino acids indicates a high degree of

intrachain disulfide cross-linking.

The NH₂-terminal domain of CILP containing the Type 1

thrombospondin-like repeat and 16 cysteines is associated with a strikingly hydrophilic region from amino acids 130–225. Following this, a relatively exposed region may lead to another compact, disulfide-bonded region between residues 354 and 445. Another extended region with no cysteines leads to the very basic furin-type cleavage site.

The lack of similarities between CILP and other proteins of
defined structure neither allowed us to predict a secondary

structure nor any remarkable structural feature or function.

There is a homology to the Type 1 repeats of thrombospondin in the consensus sequence for heparin binding (24), from residues

131 to 135 (WSPWS), but we have no experimental evidence that CILP actually binds to heparin.

The human homolog of the nucleotide pyrophosphohydrolase

(NTPPHase) that forms the COOH-terminal of CILP is made of

460 amino acid residues, with a calculated molecular mass of

51.8 kDa, and an isoelectric point 8.73. It has 10 cysteine

residues and two putative

N-glycosylation sites which is consistent with a molecular mass

of 92 kDa, the apparent molecular mass of the protein purified from human articular cartilage (25) after post-translational modifications. The protein contains 30 cysteine residues mostly
distributed toward the NH₂-terminal half of the molecule. Due to

the special role of cysteine residues for protein structure, the

high content of these amino acids indicates a high degree of

intrachain disulfide cross-linking.

The NH₂-terminal domain of CILP containing the Type 1

thrombospondin-like repeat and 16 cysteines is associated with a strikingly hydrophilic region from amino acids 130–225. Following this, a relatively exposed region may lead to another compact, disulfide-bonded region between residues 354 and 445. Another extended region with no cysteines leads to the very basic furin-type cleavage site.
16. Charles, I. G., Palmer, R. M. J., Hickery, M. S., Bayliss, M. T., Chubb, A. P.,
    Hall, V. S., Moss, D. W., and Moncada, S. (1993) Proc. Natl. Acad. Sci.
    U. S. A. 90, 11419–11423
17. Chamberlain, J. P. (1979) Anal. Biochem. 98, 132–135
18. Kozak, M. (1991) J. Cell Biol. 115, 887–903
19. von Heijne, G. (1986) Nucleic Acids Res. 14, 4683–4690
20. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990)
    J. Mol. Biol. 215, 403–410
21. Lawler, J., and Hynes, R. O. (1986) J. Cell Biol. 103, 1635–1648
22. Masuda, I., Hamada, J., Haas, A. L., Ryan, L. M., and McCarty, D. J. (1995)
    J. Clin. Invest. 95, 699–704
23. Hosaka, M., Nagahama, M., Kim, W., Watanabe, T., Hatusawa, K., Ikenizu
    J., Murakami, K., and Nakayama, K. (1991) J. Biol. Chem. 266, 12127–12130
24. Guo, N., Krutzs, H. C., Negre, E., Zabrenetzky, V. S., and Roberts, D. D. (1992)
    J. Biol. Chem. 267, 19549–19555
25. Lorenzo, P., Bayliss, M. T., and Heinegård, D. (1998) J. Biol. Chem. 273,
    23463–23468