

# Isolation and Sequence of a Novel Human Chondrocyte Protein Related to Mammalian Members of the Chitinase Protein Family\*

(Received for publication, April 3, 1996, and in revised form, May 20, 1996)

Bo Hu, Kien Trinh, William F. Figueira, and Paul A. Price‡

From the Department of Biology, University of California, San Diego, La Jolla, California 92093

**We describe the isolation of a novel protein from the conditioned medium of human articular cartilage chondrocytes in primary culture. This 39-kDa protein has the N-terminal sequence YKL, which we have termed YKL-39. The 1434-nucleotide sequence of the YKL-39 cDNA predicts a 385-residue initial translation product and a 364-residue mature YKL-39. The amino acid sequence of YKL-39 is most closely related to YKL-40, followed by macrophage chitotriosidase, oviductal glycoprotein, and macrophage YM-1. All five proteins share significant sequence identity with bacterial chitinases and have the probable structure of an  $(\alpha\beta)_8$  barrel. YKL-39 lacks the active site glutamate, which is essential for the activity of chitinases, and as expected has no chitinase activity. The highest level of YKL-39 mRNA expression is seen in chondrocytes, followed by synoviocytes, lung, and heart. YKL-39 accounts for 4% of the protein in chondrocyte-conditioned medium, prostromelysin accounts for 17%, and YKL-40 accounts for 33%. In contrast to YKL-40, YKL-39 is not a glycoprotein and does not bind to heparin.**

In this study we report the discovery of a new member of the mammalian protein family related in sequence to bacterial chitinases. This protein family has an  $(\alpha\beta)_8$  barrel structure (1, 2) and includes a protein secreted from human macrophages that does have chitinase activity, which is termed chitotriosidase (3), and three proteins with no presently known enzymatic activity, YKL-40 (4), YM-1 (GenBank™/EBI accession number M94584), and oviductal glycoprotein (6). Because the new member of this family, which we have termed YKL-39, is more closely related in size and sequence to YKL-40 than to other members of this family, it is useful to summarize research on YKL-40 as a background for the present investigation.

YKL-40 is a 40-kDa glycoprotein that was first discovered as a heparin-binding protein secreted from bovine breast tissue during the massive tissue involution that follows the cessation of lactation (7). YKL-40 was subsequently discovered as a heparin-binding protein in the conditioned medium of human synoviocytes (8), chondrocytes (4, 9), and the MG-63 osteosarcoma cell line (10). YKL-40 has also been discovered as a heparin-binding protein expressed by porcine vascular smooth muscle cells undergoing a differentiation transition (11) and as a protein expressed selectively by murine mammary tumors

initiated by *neu/ras* oncogenes (12). The present studies were initiated to further examine the expression of YKL-40 by human articular cartilage chondrocytes in culture. In the course of these studies YKL-39 was found as a protein that copurified with YKL-40.

We report here the discovery, purification, characterization, and sequence of human YKL-39.

## MATERIALS AND METHODS

**Isolation and Culture of Chondrocytes and Synoviocytes**—Chondrocytes and synoviocytes were obtained from Martin Lotz, director of the University of California, San Diego osteoarthritis cell culture facility and were isolated and cultured essentially as described (13). Cartilage from the femoral condyles and tibial plateaus of the knee joints was obtained at autopsy from donors without known history of joint disease or from healthy organ donors from the University of California, San Diego tissue bank. Cartilage slices were cut into pieces (2–3 mm<sup>3</sup>), washed with DMEM,<sup>1</sup> and treated for 15 min with trypsin (10% v/v) in a 37 °C water bath. The tissues were transferred to DMEM containing 5% fetal calf serum, penicillin-streptomycin-Fungizone, and 2 mg/ml clostridial collagenase type IV (Sigma) and digested overnight on a shaker until the tissue fragments were dissolved. The cells were washed three times with DMEM and cultured in T175 flasks containing 30 ml of DMEM plus 10% fetal calf serum until confluent. All experiments reported here used chondrocytes in primary culture or at passage 1 following a 1:3 subculture. To harvest conditioned culture medium, chondrocyte cultures were grown to confluence in T175 flasks, washed twice with 30 ml of phosphate-buffered saline, and cultured in 30 ml of serum-free DMEM for 1 week.

Synovial tissues were obtained from knee joints and washed with DMEM, minced, and treated with trypsin (10% v/v) for 15 min in a 37 °C water bath. The tissue fragments were then transferred to DMEM containing 5% fetal calf serum, penicillin-streptomycin-Fungizone, and 2 mg/ml clostridial collagenase type IV (Sigma) and digested on a shaker until dissolution of the fragments (about 3 h). The cells were washed three times with DMEM and cultured in T175 flasks. After 24 h nonadherent cells were removed, and the adherent synovial cells were further cultured until confluent and then harvested for RNA isolation.

**Purification of YKL-39**—To fractionate conditioned medium proteins by size (see Fig. 1), 200 ml of medium obtained after 1 week of culture in serum-free conditions was concentrated to 5 ml by ultrafiltration using a 10-kDa MWCO membrane and applied to a 2 × 150-cm Sephacryl S-300 HR column equilibrated with 150 mM NH<sub>4</sub>HCO<sub>3</sub> at room temperature. To remove YKL-40 and other heparin-binding proteins, 600 ml of conditioned medium was passed through a 2 × 15-cm heparin-Sepharose CL-6B column initially equilibrated with 20 mM sodium phosphate buffer, pH 7.4. The unbound proteins were concentrated to 5 ml by ultrafiltration and then applied to a 2 × 150-cm Sephacryl S-300 HR column equilibrated with 150 mM NH<sub>4</sub>HCO<sub>3</sub> at room temperature (see Fig. 2).

**SDS-PAGE**—SDS-polyacrylamide gel electrophoresis was performed under reducing conditions using 4–20% gradient gels (Novex, San Diego, CA) and stained with Coomassie Brilliant Blue or with the periodic acid-Schiff reaction (glycoprotein detection kit, Sigma). Conditioned medium was obtained from primary chondrocytes cultured in the ab-

\* This work was supported in part by a grant from NovaDx, Inc. and by United States Public Health Service Grant AG07996. 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) U49835.

‡ To whom correspondence should be addressed: Dept. of Biology, 0322, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0322. Tel.: 619-534-2120; Fax: 619-534-1492.

<sup>1</sup> The abbreviations used are: DMEM, Dulbecco's modified Eagle's medium; YKL-39, the 39-kDa human chondrocyte protein with the N terminus YKL; YKL-40, the 40-kDa human chondrocyte protein with the N terminus YKL (also called HC gp-39; see Ref. 4); PAGE, polyacrylamide gel electrophoresis.

sence of serum for 7 days and was dialyzed against milli Q water before concentration for electrophoresis (see Fig. 1, inset, lane 2).

**N-Terminal Protein Sequence Analysis**—Purified proteins were transferred to polyvinylidene difluoride membranes using a ProSpin device (Applied Biosystems, Foster City, CA) and sequenced using a Perkin-Elmer/Applied Biosystems Division model 494 sequenator equipped with online high performance liquid chromatography.

**Sequencing of the YKL-39 cDNA**—As explained under "Results," clone 118809 isolated by the Washington University-Merck EST Project from a human lung cDNA library proved to be a cDNA clone of YKL-39. We purchased this clone from Genome Systems, Inc. (St. Louis, MO). *Sma*I and *Xho*I digestion of this clone showed that it contains a 1.5-kilobase insert. We sequenced the 5' and 3' ends of the insert DNA using a T7 primer and a M13 reverse primer for the pBluescript SK(-) phagemid and a version 2.0 DNA sequencing kit (U. S. Biochemical Corp.). Synthetic primers were synthesized from the DNA sequences and used to extend the sequence in increments until the sequence of both strands was determined.

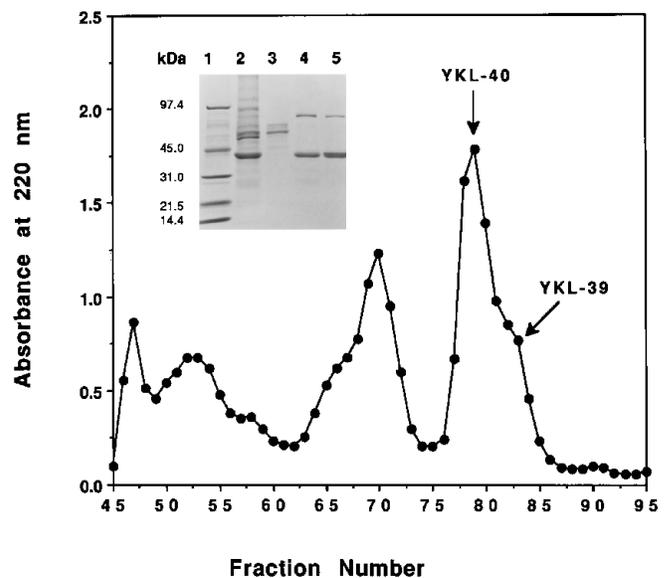
**Northern Blot Analysis**—Total RNA was isolated from chondrocytes, synoviocytes, MG-63 osteosarcoma cells, normal dermal fibroblasts, and Hep G2 liver cells with an RNA STAT-60™ kit (Tel-Tek "B," Inc., Friendswood, TX). Thirty  $\mu$ g of total RNA from each cell line was fractionated on 1% formaldehyde-agarose gel in 4-morpholinepropane-sulfonic acid buffer and transferred onto a Hybond-N membrane (Amersham Corp.). A multiple tissue Northern blot containing total RNA from human heart, brain, kidney, liver, lung, pancreas, and spleen was purchased from BioChain Institute, Inc. (San Leandro, CA). Following prehybridization for 3 h in 50% formamide,  $5 \times$  SSC,  $5 \times$  Denhardt's solution, and 100  $\mu$ g/ml denatured salmon sperm DNA at 42 °C, blots were hybridized with random primed  $^{32}$ P-labeled cDNA probes for 16–20 h under identical conditions. Filters were washed three times for 1 h each with  $0.1 \times$  SSC containing 0.1% SDS at 65 °C and then exposed to x-ray film. The blots were first hybridized with the *Eco*RI fragment of clone 118809 (nucleotides 388–1256, see Fig. 4) and after exposure to x-ray film were then stripped by boiling in H<sub>2</sub>O containing 0.5% SDS and rehybridized with cDNA probes for YKL-40 (nucleotides 135–916, see Ref. 4) and glyceraldehyde-3-phosphate dehydrogenase.

**Determination of Chitinase Activity**—The chitinase activity of YKL-39 was tested with the fluorogenic substrates 4-methylumbelliferyl  $\beta$ -D-N,N'-diacetylchitobiose and 4-methylumbelliferyl  $\beta$ -D-N,N,N'-triacetylchitotriose (Sigma), and the *p*-nitrophenyl substrates *p*-nitrophenyl  $\beta$ -D-N,N'-diacetylchitobiose and *p*-nitrophenyl  $\beta$ -D-N,N,N'-triacetylchitotriose (Sigma), as described by Renkema *et al.* (14). Chitinase from *Serratia marcescens* (Sigma) was used as a positive control.

## RESULTS

**Purification of Cousin**—To harvest proteins secreted from chondrocytes, primary cultures of human articular cartilage chondrocytes were grown to confluence in media containing 10% fetal calf serum and then switched to serum-free medium. After 1 week of culture, conditioned medium was concentrated by ultrafiltration and fractionated by molecular weight using Sephacryl S-300 HR. As seen in Fig. 1, the major protein component in conditioned medium emerges in an asymmetric peak centered at fraction 79. Fraction 79 gave a band<sup>2</sup> at 40 kDa upon SDS-PAGE (Fig. 1, inset) and proved to be YKL-40 by the criteria of N-terminal protein sequencing (YKLVXYT~~SW~~-) and by radioimmunoassay (9). The next most abundant protein component in chondrocyte-conditioned medium emerges in an asymmetric peak centered at fraction 70. N-terminal protein sequencing of fraction 70 identified a major sequence identical to human prostromelysin and a minor sequence identical to bovine serum albumin. (The presence of some bovine albumin in conditioned medium is not unusual in such experiments, since albumin is the most abundant protein in fetal calf serum and so is the most likely serum constituent to contaminate serum-free conditioned medium.)

<sup>2</sup> When purified YKL-40 is dried in preparation for loading onto a gel, it gives variable amounts of dimeric YKL-40 (data not shown). This accounts for the 80-kDa band seen in Fig. 1 (inset), a band that was not seen in the SDS-PAGE of these fractions when the drying step was omitted.



**FIG. 1. Separation of proteins in chondrocyte-conditioned medium by filtration over a Sephacryl S-300 HR column.** Human articular cartilage chondrocytes were grown to confluence and then changed to serum-free medium. After 7 days of culture in serum-free conditions, 200 ml of medium was removed, concentrated to 5 ml by ultrafiltration, and applied to a  $2 \times 150$ -cm Sephacryl S-300 HR column equilibrated with 150 mM NH<sub>4</sub>HCO<sub>3</sub> at room temperature. Fraction volume, 4.8 ml (see "Materials and Methods" for details). Inset, SDS-polyacrylamide gel electrophoresis of fractions from the Sephacryl S-300 HR column. Proteins were electrophoresed on a 4–20% gradient gel and stained with Coomassie Brilliant Blue. Lane 1, molecular mass standards; lane 2, 0.5 ml of chondrocyte-conditioned medium; lane 3, 10  $\mu$ g of fraction 70; lane 4, 10  $\mu$ g of fraction 79; lane 5, 10  $\mu$ g of fraction 83.

Several experiments were carried out to ascertain the cause of the asymmetry in the YKL-40 peak centered at fraction 79 in Fig. 1. When the level of YKL-40 was determined by radioimmunoassay (9), there was a good correlation between  $A_{220}$  and immunoreactivity for fractions 77–80 but less immunoreactivity than expected based on absorbance for fractions 81–84 (data not shown). This result indicated the possible presence of an  $A_{220}$ -absorbing protein constituent not recognized by the radioimmunoassay for YKL-40. SDS-PAGE of fraction 83 revealed a doublet at 39–40 kDa (Fig. 1, inset). N-terminal protein sequencing of fraction 83 revealed the apparent presence of two sequences, and therefore two proteins, each in approximately equimolar amounts. One sequence is that of YKL-40 (YKLVXYT~~SW~~SQYR-) and the other is very similar to YKL-40 (YKLVXYFTNWSQDR-).

To further evaluate the possible existence of a putative YKL-40-related protein, another portion of the serum-free conditioned medium whose fractionation is shown in Fig. 1 was separated by SDS-PAGE (Fig. 1, inset) and transferred to a polyvinylidene difluoride membrane. Subsequent N-terminal protein sequencing of the intense protein band centered at 40 kDa revealed a mixture of two sequences, the major (90%) of which is identical to YKL-40 and the minor (10%) of which is identical to the putative YKL-40-related protein. Interestingly, the N-terminal sequencing of YKL-40 purified from conditioned medium by heparin affinity chromatography, our customary procedure for purification of the protein (9), yielded only the expected sequence of YKL-40 with no trace of the putative YKL-40-related protein.

To see whether the putative YKL-40-related protein might be among the conditioned medium proteins that did not bind to the heparin column, the unbound protein fraction was concentrated by ultrafiltration and fractionated by molecular weight

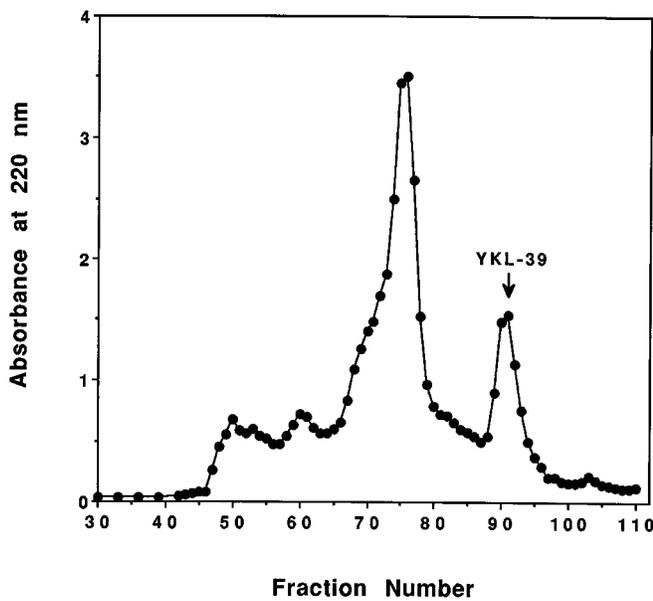


FIG. 2. Purification of YKL-39 by filtration over Sephacryl S-300 HR of the conditioned medium proteins that do not bind heparin. 600 ml of 7-day conditioned medium (Fig. 1) were first passed over a  $2 \times 30$ -cm column of heparin-Sepharose CL-6B to remove YKL-40 and other heparin-binding proteins. The unbound fraction was concentrated to 5 ml and applied to a  $2 \times 150$ -cm Sephacryl S-300 HR column equilibrated with 150 mM  $\text{NH}_4\text{HCO}_3$  at room temperature. Fraction volume, 4.2 ml (see "Materials and Methods" for details).

using the same Sephacryl S-300 HR column. As can be seen in Fig. 2, there is a protein constituent at the approximate elution volume of YKL-40 centered at effluent fraction 91. N-terminal protein sequencing of fraction 91 revealed the presence of a single sequence, YKLVXYFTNWSQDRQEPGKFTPENI-, the sequence predicted for the YKL-40-related protein, with no trace of the YKL-40 sequence itself. SDS gel electrophoresis revealed this protein to be 39 kDa in apparent mass and therefore slightly smaller than YKL-40 (Fig. 3a), and staining of the gel for carbohydrate showed that the YKL-40-related protein is not a glycoprotein, whereas, as previous studies have shown (4, 7), YKL-40 itself is (Fig. 3b). Since YKL-40 and the YKL-40-related protein have the same N-terminal sequence YKL but differ in apparent mass, we have given the YKL-40-related protein the provisional name YKL-39 to denote its 39-kDa apparent mass and so distinguish it from the 40-kDa YKL-40. The other major protein component seen in the chromatogram shown in Fig. 2 emerges in an asymmetric peak centered at fraction 76. N-terminal protein sequencing of fraction 76 revealed it to be a mixture of human prostromelysin and bovine albumin.

Based on the recovery of  $A_{220}$  absorbance, a direct measure of total protein, we calculate from the data in Figs. 1 and 2 that 10 confluent T175 flasks of articular cartilage chondrocytes accumulate 18 mg of total protein (assuming a 1 mg/ml protein concentration has an  $A_{220} = 10.0$ ) into conditioned medium after 1 week of culture in serum-free conditions. Of this total, YKL-40 accounts for 6 mg (33% of total protein) and YKL-39 accounts for 0.75 mg (4%). The five fractions in Fig. 1 that contain almost all of the prostromelysin (fractions 69–73; based on SDS-PAGE) account for 3 mg of the conditioned medium protein (17%). Because these fractions are contaminated with albumin, the actual amount of prostromelysin in conditioned medium must be somewhat lower.

**Determining the cDNA Sequence of YKL-39**—Before initiating efforts to clone YKL-39 we investigated the possibility that clones for YKL-39 may have already been partially sequenced

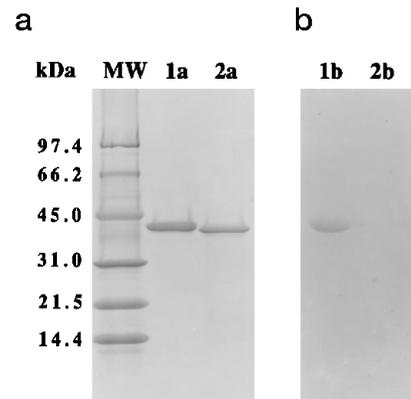


FIG. 3. SDS-PAGE comparison of YKL-39 and YKL-40. Proteins were electrophoresed on a 4–20% gradient gel and stained with Coomassie Brilliant Blue (a) and the Sigma glycoprotein detection kit (b). Lane MW, molecular mass standards; lanes 1a and 1b, 5 and 10  $\mu\text{g}$  of YKL-40; lanes 2a and 2b, 5 and 10  $\mu\text{g}$  of YKL-39.

as part of the Washington University-Merck EST Project. One approach was to design a predicted cDNA sequence from the N-terminal 25-residue sequence of YKL-39 using where possible the preferred codon usage in humans. When we used this to screen the EST data base with the Blast N program (15), only one sequence gave a significant match, clone 257753 (GenBank<sup>TM</sup>/EBI accession number N40107). The other approach was to use the known cDNA sequence of YKL-40 itself to identify sequences closely related to, but not identical with, YKL-40. This approach yielded an additional clone, clone 118809 (GenBank<sup>TM</sup>/EBI accession number T91693). The partial sequences for clones 257753 and 118809 that were available in the EST data base indicated a region of overlap identity, and we therefore concluded that they were clones of the same gene. We then obtained clone 118809 and determined its complete cDNA sequence.

Fig. 4 shows the complete nucleotide sequence of clone 118809 and the deduced amino acid sequence of YKL-39. The coding region of YKL-39 is terminated by a TGA triplet at nucleotide 1191 and is followed by 242 nucleotides of 3'-untranslated region with a potential polyadenylation signal (AATAAAA) at nucleotides 1391–1396. The ATG, found at nucleotides 36–38, was considered to be the initiation codon according to the rules for translation initiation described previously (16). The open reading frame codes for a 385-residue-long protein containing a 21-residue transmembrane signal peptide with a potential signal peptidase cleavage site at amino acid residue 21 (17). The predicted protein, after removal of the signal peptide, has a length of 364 amino acids and a calculated molecular mass of 40,825 Da. The predicted N-terminal sequence of YKL-39 is identical to the N-terminal 25-residue sequence determined by protein sequencing of the purified protein.

The putative NXS site of *N*-glycosylation in YKL-40, asparagine residue 60 and serine residue 62 (4), is not found in YKL-39. There is a single potential recognition site for *N*-glycosylation in YKL-39, the NWS sequence at residues 30 to 32. This site does not appear to be a functional site of *N*-glycosylation in YKL-39 secreted from chondrocytes, however, since no carbohydrate could be detected in purified YKL-39 (Fig. 3b) and N-terminal sequencing of the mature YKL-39 revealed the expected repetitive yield of asparagine at residue 9. Although the predicted number of amino acid residues in the mature form of YKL-39 is slightly larger than for YKL-40 (364 versus 362 residues), the presence of carbohydrate in YKL-40 but not in YKL-39 is probably sufficient to account for the fact that YKL-40 appears to be slightly larger than YKL-39 based

1	AGAAGAAGCTGGCCAAGGATATGGGACCAACCCATGGACCAGAAGTCTCTCTGGGCA	59
1	MetAspGlnLysSerLeuTrpAla	8
60	GGTGTAGTGGTCTTGCTGCTCTCCAGGGAGGATCTGCCTACAACCTGGTTGCTACTTT	119
9	GlyValValValLeuLeuLeuLeuGlnGlyGlySerAlaTyrLysLeuValCysTyrPhe	28
	↑	
120	ACCAACTGGTCCCAGGACCGGCAGGAACCCAGGAAATTCACCCCTGAGAATATTGACCCC	179
29	ThrAsnTrpSerGlnAspArgGlnGluProGlyLysPheThrProGluAsnIleAspPro	48
180	TTCCATGCTCTCATCTCATCTTTCATTCCGCAGCATCGAAAACACAAAGTTATCATC	239
49	PheLeuCysSerHisLeuIleTyrSerPheAlaSerIleGluAsnAsnLysValIleIle	68
240	AGGCACAGAGTGAAGTGAAGTGCCTACCAGACCTCAACAGCTCAAAACCAAGAAATCCC	299
69	LysAspLysSerGluValMetLeuTyrGlnThrIleAsnSerLeuLysThrLysAsnPro	88
300	AACTGAAAATTCCTTGTCCATTGGAGGGTACCTGTTTGGTTCCAAAGGGTCCACCCCT	359
89	LysLeuLysIleLeuLeuSerIleGlyGlyTyrLeuPheGlySerLysGlyPheHisPro	108
360	ATGGTGGATTCTTACATCACGCTTGAATTCATTAACCTCCATAATCCTGTTTCTGAGG	419
109	MetValAspSerSerThrSerArgLeuGluPheIleAsnSerIleIleLeuPheLeuArg	128
420	AACCAATACTTGTAGGACTGGATGAAGTGGATCTACCCAGATCAGAAAGAAAACACT	479
129	AsnHisAsnPheAspGlyLeuAspValSerTrpIleTyrProAspGlnLysGluAsnThr	148
480	CATTTCACTGTCTGATTGATGAGTTAGCAGAGCCCTTCAGAAAGGACTTCACAAAATCC	539
149	HisPheThrValLeuIleHisGluLeuAlaGluAlaPheGlnLysAspPheThrLysSer	168
540	ACCAAGGAAGGCTTCTCTGACTGCGGGCGIATCTCAGGGAGGCAATGATTGATAAC	599
169	ThrLysGluArgLeuLeuLeuThrAlaGlyValSerAlaGlyArgGlnMetIleAspAsn	188
600	AGCTATCAAGTTGAGAAACTGGCAAAGATCTGGATTTCATCAACCTCCTGTCCCTTGAC	659
189	SerTyrGlnValGluLysLeuAlaLysAspLeuAspPheIleAsnLeuLeuSerPheAsp	208
660	TTCCATGGGCTTGGGAAAAGCCCTTATCACTGGCCCAACGCCCTCTGAGCAAGGGG	719
209	PheHisGlySerTrpGluLysProLeuIleThrGlyHisAsnSerProLeuSerLysGly	228
720	TGGCAGGACAGAGGGCCAAGCTCCTACTACAATGTGGAATATGCTGGGGTACTGGATA	779
229	TrpGlnAspArgGlyProSerSerTyrTyrAsnValGluTyrAlaValGlyTyrTrpIle	248
780	CATAAGGGAATGCCATCAGAGAAGGTGTCATGGGCATCCCCACATATGGGCATCCTTC	839
249	HisLysGlyMetProSerGluLysValValMetGlyIleProThrTyrGlyHisSerPhe	268
840	ACACTGGCCCTTCGCAAAAACCCGCTGGGGCCCTCGCCTCGGCCCTGGAGCTGCTGGA	899
269	ThrLeuAlaSerAlaGluThrThrValGlyAlaProAlaSerGlyProGlyAlaAlaGly	288
900	CCCATCACAGATCTTCAGGCTTCCTGGCCTATTATGAGATCTGCCAGTTCCGAAAGGA	959
289	ProIleThrGluSerSerGlyPheLeuAlaTyrTyrGluIleCysGlnPheLeuLysGly	308
960	GCCAAAGATCACGGCCTCCAGGATCAGCAGGTTCCTACGCACTCAAGGGGAACCACTGG	1019
309	AlaLysIleThrArgLeuGlnAspGlnGlnValProTyrAlaValLysGlyAsnGlnTrp	328
1020	GTGGGCTATGATGATGTAAGAGTATGGAGCAAGGTTCAAGTCTTAAAGAATTTAAAC	1079
329	ValGlyTyrAspValLysSerMetGluThrLysValGlnPheLeuLysAsnLeuAsn	348
1080	CTGGGAGGACCATGATCTGGTCTTATGACATGGATGACTTCACTGGCAATTCCTGCAAC	1139
349	LeuGlyGlyAlaMetIleTrpSerIleAspMetAspAspPheThrGlyLysSerCysAsn	368
1140	CAGGGCCCTTACCCTCTTGTCCAAGCAGTCAAGAGAAGCCTTGGCTCCTGTGAAGGATT	1199
369	GlnGlyProTyrProLeuValGlnAlaValLysArgSerLeuGlySerLeu***	385
1200	AACCTACAGAGAAGCAGGCAAGATGACCTTGCTGCCGTCGCTCTCTCCAGGAAT	1259
1260	TCTCATGTGGGATTCCTCCAGGCTGGCCCTTTGGATCTCTCTTCCAAGCCTTCTCCTG	1319
1320	ACTTCCCTCTTAGATCATAGATGGACCTGGTTTTGTTTTCTCCGACGCTGTGACTGTTG	1379
1380	CCCTGAAGTACAAATAAAAAATTCATTTTGTCCAGTAAAAAATAAAAAAATAAAAA	1435

FIG. 4. Complete nucleotide sequence of YKL-39 cDNA and deduced amino acid sequence of the protein. The underlined sequence corresponds to that determined by N-terminal protein sequencing. The cleavage site of the putative signal sequence is indicated by the arrow after residue 21, and the N terminus of mature YKL-39 begins at residue 22. The stop codon is marked by asterisks, and the polyadenylation signal sequence AATAAA is indicated by boldface.

on its elution position from Sephacryl S-300 HR (Fig. 1) and on its apparent SDS-PAGE molecular weight (Fig. 3).

**Tissue and Cell-type Specificity of YKL-39 mRNA Expression**—Five cell cultures were tested for the production of YKL-39 by Northern blot. The <sup>32</sup>P-labeled cDNA probe for YKL-39 hybridized with a single 1.5-kilobase band in the chondrocyte and synoviocyte RNA samples (Fig. 5a). This size is in good agreement with the 1434-base pair size of clone 118809. No YKL-39 mRNA could be detected in RNA from normal human fibroblasts, the HEP G2 human liver cell line, and the MG-63 human osteoblastic osteosarcoma cell line. As expected (4, 10), reprobe of this membrane demonstrated high levels of YKL-40 mRNA in chondrocytes and MG-63 cells, lower levels in synoviocytes, and undetectable levels in HEP G2 and fibroblasts (data not shown). The lack of YKL-39 mRNA expression by MG-63 cells is in agreement with the fact that the previously reported N-terminal protein sequencing of the 40-kDa protein band seen in SDS-PAGE of MG63 conditioned media identified the sequence of YKL-40 but failed to detect evidence of the sequence of YKL-39 (10).

It is possible that culture under the serum-free conditions that are needed for protein isolation from conditioned medium could significantly alter the quantitative pattern of secreted

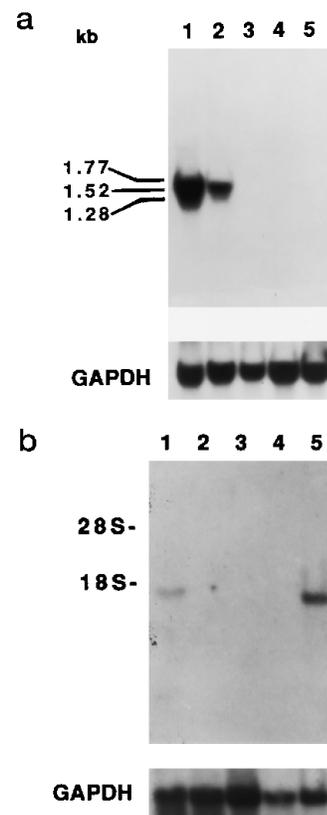


FIG. 5. Northern blot analysis of YKL-39 message levels in human cell lines and human tissues. *a*, 30  $\mu$ g of total RNA from chondrocytes (lane 1), synoviocytes (lane 2), MG-63 osteosarcoma cells (lane 3), normal dermal fibroblasts (lane 4), and Hep G2 liver cells (lane 5) were run on a 1% formaldehyde-agarose gel and blotted onto a Hybond-N membrane. *b*, a commercially obtained human multiple tissue blot containing 20  $\mu$ g of total RNA from human heart (lane 1), brain (lane 2), kidney (lane 3), liver (lane 4), lung (lane 5), pancreas (lane 6), and spleen (lane 7). The blots were first hybridized with a <sup>32</sup>P-labeled YKL-39 cDNA probe and then reprobbed with a <sup>32</sup>P-labeled glyceraldehyde-3-phosphate dehydrogenase cDNA fragment.

protein expression. To evaluate the possible effect of serum-free conditions on YKL-39 expression, confluent primary chondrocytes were cultured for 1 week in serum-free medium or in medium containing 10% fetal calf serum, and the level of YKL-39 was determined by Northern blot. The YKL-39 levels in the serum-free and serum-containing cultures proved to be identical (data not shown). We could also detect no effect of serum-free conditions on the level of YKL-40 mRNA or on the medium levels of YKL-40 antigen.

Seven tissues from adult humans were examined for the presence of YKL-39 mRNA (Fig. 5b). YKL-39 mRNA is expressed strongly in lung and is detectable in heart. No YKL-39 mRNA could be detected in brain, spleen, pancreas, or liver. A reprobe of this membrane failed to detect YKL-40 mRNA in any of these tissues. Further evidence on the tissue distribution of YKL-39 mRNA expression is provided by the frequency with which YKL-39 cDNA clones have been identified by the Washington University-Merck EST Project. Two different clones have been obtained from a human lung cDNA library (GenBank<sup>TM</sup>/EBI accession numbers T91693 and T66009), one clone has been obtained from each of two infant brain libraries (GenBank<sup>TM</sup>/EBI accession numbers H10721 and H10989), and one has been obtained from a placenta library (GenBank<sup>TM</sup>/EBI accession number N40107). This evidence suggests that YKL-39 may also be expressed in developing brain and in placenta.

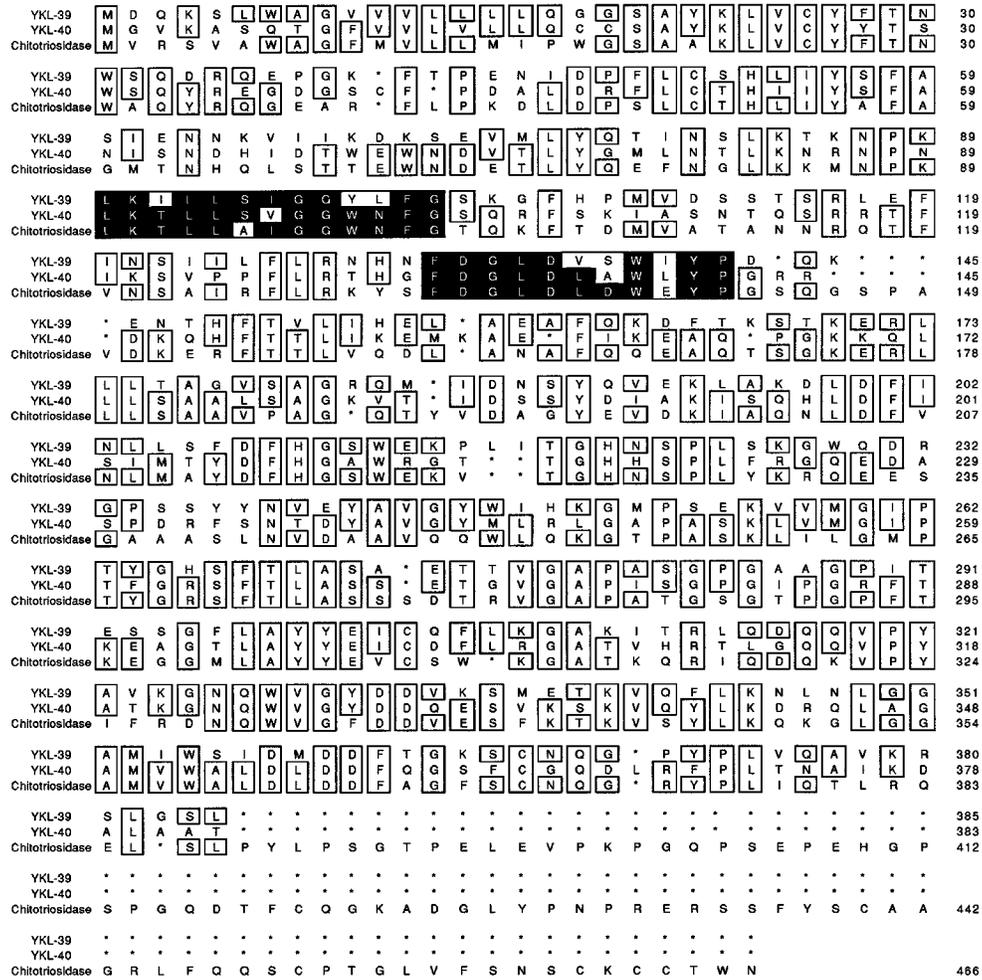


FIG. 6. Alignment of YKL-39 with mammalian members of the chitinase protein family. The sequences of human YKL-39 (Fig. 5), human YKL-40 (4), and human chitinotriosidase (3) are compared starting with the first residue of the initial translation product. Identical amino acids are boxed. The shaded residues indicate the sequence regions that correspond to the putative active site in bacterial chitinases (18, 19).

DISCUSSION

*Comparison of YKL-39 to Related Mammalian Proteins—* Pairwise comparison of YKL-39 with the four previously identified mammalian members of the protein family related in sequence to bacterial chitinases using the ALIGN program (Protein Identification Resource) shows that YKL-39 is most closely related to human YKL-40 (4) followed by human chitinotriosidase (3), human oviductal glycoprotein (6), and murine YM-1 (GenBank™/EBI accession number M94584). The regions of sequence identity encompass all of YKL-39, YKL-40, and YM-1 and the N-terminal domains of chitinotriosidase and of oviductal glycoprotein. The C-terminal domain of chitinotriosidase is related to the C-terminal domain of nematode and insect chitinases (3) and has no counterpart in YKL-40, YKL-39, or YM-1. The C-terminal domain of oviductal glycoprotein is thought to be a region of extensive protein glycosylation (6).

The amino acid sequence of YKL-39 is compared to the two proteins which are most closely related to it in sequence, YKL-40 and chitinotriosidase, in Fig. 6. As can be seen, there is extensive sequence identity among all three human proteins, particularly in the regions that are thought to be involved in substrate binding in the bacterial chitinases (18, 19) (Fig. 6, shaded residues). It is of interest to note that the glutamate residue which is known from mutagenesis studies to be essential for the activity of bacterial chitinases (20) is found in chitinotriosidase (Fig. 6, glutamate 139) but not in YKL-39, YKL-40, YM-1, or oviductal glycoprotein. This observation is consistent

with the fact that chitinotriosidase is a glycosidic bound hydrolyase, while no enzymatic activity has yet been reported for YKL-40, YM-1, or oviductal glycoprotein. Although we have not carried out extensive tests of the possible enzymatic activities of YKL-39, we did examine its possible chitinase activity (see "Materials and Methods"). YKL-39 was not active in any of the four assays tested (data not shown).

The high degree of sequence identity between YKL-40 and YKL-39, which includes continuous stretches of identity up to 10 residues in length, suggests that antibodies against one protein could cross-react with the other. We therefore tested the possible cross-reactivity of YKL-39 on the radioimmunoassay that we developed for measurement of human YKL-40 in serum and tissues (9). No cross-reactivity was observed, which indicates that the dominant epitope recognized by the antiserum used for the YKL-40 assay does not recognize YKL-39. It should be noted, however, that this result does not rule out the presence of minor cross-reacting antibodies in this antiserum.

Recent investigations have assigned bacterial chitinases and the related mammalian proteins to a gene family termed family 18 of the glycosylhydrolases (1). Based on the crystallographic structure of one member of this family, it has been further suggested that all members of this gene family have the tertiary structure of an (αβ)<sub>8</sub> barrel (also called a TIM barrel) (2). TIM barrels are the most common architecture of enzymes and are found in 10% of the enzymes whose structures are presently known (21). The substrate binding site in such enzymes is

invariably formed by the loops which connect the  $\beta$  sheet and  $\alpha$  helical segments that are located at the C-terminal end of the 8-stranded barrel. It seems probable that this region of the YKL-39 structure also has specific binding properties, forming either an active site or a specific glycan binding site.

We speculate that the high level of sequence identity between chitinases and YKL-39 in the region of the sequence that corresponds to the putative chitinase active site (Fig. 6, *shaded residues*) is best explained by the need to conserve residues involved in glycan binding and that YKL-39 is likely to bind a given glycan structure with high specificity. This hypothesis is supported by the fact that YKL-40 binds a glycan, heparin, with an affinity greater than found for the well established heparin binding of fibronectin (11). Since YKL-39 clearly has no affinity for heparin in spite of its similarity in net charge and sequence to YKL-40, it seems likely that heparin binding is attributable to a specific binding site on YKL-40, one that is conserved in all species tested (cow, human, pig). We are currently investigating the glycan binding specificities of YKL-39.

*Physiologic Processes in Which YKL-39 Could Function*—We think that the expression of YKL-39 at high levels in primary chondrocyte cultures suggests that the protein could function in tissue remodeling processes. Evidence for this hypothesis is provided by the identity of the other major proteins in chondrocyte-conditioned medium (Fig. 1). The most abundant protein in chondrocyte-conditioned medium, YKL-40, is thought to be involved in remodeling of breast tissue (7), vascular smooth muscle (11), and cartilage (4). YKL-40 is not in fact even expressed at detectable levels in normal cartilage or in cartilage explants until 2 days in culture (4) (data not shown). YKL-40 is, however, expressed at high levels in arthritic cartilage, an abnormal tissue characterized by high levels of tissue destruction and turnover. The next most abundant protein in chondro-

cyte-conditioned medium, prostromelysin, is an enzyme with broad substrate specificity which is involved with the digestion of a wide variety of extracellular matrix proteins in tissue remodeling processes. If YKL-39 is involved in the remodeling of cartilage, it is likely that YKL-39 is, like YKL-40, not expressed by normal cartilage but is induced in explant culture and expressed at high levels in arthritic cartilage.

## REFERENCES

1. Henrissat, B., and Bairoch, A. (1993) *Biochem. J.* **293**, 781–788
2. Coulson, A. F. W. (1994) *FEBS Lett.* **354**, 41–44
3. Boot, R. G., Renkema, G. H., Strijland, A., van Zonneveld, A. J., and Aerts, J. M. F. G. (1995) *J. Biol. Chem.* **270**, 26252–26256
4. Hakala, B. E., White, C., and Recklies, A. D. (1993) *J. Biol. Chem.* **268**, 25803–25810
5. Johansen, J. S., Cintin, C., Jørgensen, M., Kamby, C., and Price, P. A. (1995) *Eur. J. Cancer* **31**, 1437–1442
6. Arias, E. B., Verhage, H. G., and Jaffe, R. C. (1994) *Biol. Reprod.* **51**, 685–694
7. Rejman, J. J., and Hurley, W. L. (1988) *Biochem. Biophys. Res. Commun.* **150**, 329–334
8. Nyirkos, P., and Golds, E. E. (1990) *Biochem. J.* **268**, 265–268
9. Johansen, J. S., Jensen, H. S., and Price, P. A. (1993) *Br. J. Rheumatol.* **32**, 949–955
10. Johansen, J. S., Williamson, M. K., Rice, J. S., and Price, P. A. (1992) *J. Bone Miner. Res.* **7**, 501–512
11. Shackleton, L. M., Mann, D. M., and Millis, A. J. T. (1995) *J. Biol. Chem.* **270**, 13076–13083
12. Morrison, B. W., and Leder, P. (1994) *Oncogene* **9**, 3417–3426
13. Maier, R., Ganu, V., and Lotz, M. (1993) *J. Biol. Chem.* **268**, 21527–21532
14. Renkema, G. H., Boot, R. G., Muijsers, A. O., Donker-Koopman, W. E., and Aerts, J. M. F. G. (1995) *J. Biol. Chem.* **270**, 2198–2202
15. Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2444–2448
16. Kozak, M. (1989) *J. Cell Biol.* **108**, 229–241
17. Von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17–21
18. Watanabe, T., Oyanagi, W., Suzuki, K., Ohnishi, K., and Tanaka, H. (1992) *J. Bacteriol.* **174**, 408–414
19. Myashita, K., and Fujii, T. (1993) *Biosci. Biotechnol. Biochem.* **57**, 1691–1698
20. Watanabe, T., Kobori, K., Miyashita, K., Fujii, T., Sakai, H., Uchida, M., and Tanaka, H. (1993) *J. Biol. Chem.* **268**, 18567–18572
21. Farber, G. K. (1993) *Curr. Opin. Struct. Biol.* **3**, 409–412