Structural features of cartilage matrix protein deduced from cDNA

(molecular cloning/epidermal growth factor homology/developmental gene regulation)

W. SCOTT ARGRAVES*^{†‡}, FERENC DEÁK*[§], KENNETH J. SPARKS*, IBOLYA KISS*[§],
AND PAUL F. GOETINCK*^{†¶}

*Department of Animal Genetics, The University of Connecticut, Storrs, CT 06268; and [†]La Jolla Cancer Research Foundation, 10901 North Torrey Pines Road, La Jolla, CA 92037

Communicated by Clifford Grobstein, October 9, 1986 (received for review September 10, 1986)

ABSTRACT cDNAs encoding the M_r 54,000 chicken cartilage matrix protein (CMP) were selected from a cartilage cDNA expression library by immunological means. Antibodies elicited against insert-encoded protein purified from one of the clones reacted specifically with chicken CMP in immunoblots of total cartilage extract, providing positive identification of the cDNA clones isolated. The cDNAs detect a 3.4-kilobase transcript that was present in sternal cartilage and in cartilaginous but not in precartilaginous embryonic limb tissues. The cDNAs code for 416 amino acids of the chicken CMP, including its COOH terminus. There are two striking features in the deduced CMP amino acid sequence: first, it contains a region with significant homologies to repeat sequences in the precursor for epidermal growth factor; and second, it is made up of two large homologous repeat sequences. These results provide the first detailed structural information on the CMP and establish it as a developmentally regulated marker of cartilage differentiation.

The importance of the extracellular matrix in developmental processes has been established from studies on morphogenesis (1, 2), tissue interactions (3, 4), and cell migration (5). The specific mode of action of the extracellular matrix components is now known to be mediated through specialized molecular domains that permit these macromolecules to interact with other matrix components or with the cell surface. In cartilage, such macromolecular interactions have been documented for the interaction between proteoglycan monomer and link protein, and the interaction between these two molecules and hyaluronic acid in the formation of link protein stabilized proteoglycan aggregates (6). To be structurally integrated in the cartilage matrix, these proteoglycan aggregates must interact, directly or indirectly and in an as yet unknown manner, with other cartilage matrix macromolecules. Some of these are: type II collagen (7), several minor collagenous proteins (8), a number of low molecular weight (8–10) and low buoyant density proteoglycans (11), chondronectin (12), and cartilage matrix protein (13). The highly structured extracellular matrix that results from these macromolecular interactions is important for the growth and morphogenesis of cartilaginous rudiments.

Recent reports on the protein chemistry (14) or the molecular cloning (15, 16) of link protein have provided structural information that suggest a molecular basis for the interaction of link protein with hyaluronic acid. As the structural details of the various components of the cartilaginous matrix become known, it will be possible to elucidate the molecular basis of their interactions and to begin to understand their role in growth and morphogenesis.

Cartilage matrix protein (CMP) is an abundant noncollagenous extracellular matrix protein of which relatively little

is known. The protein was first identified in bovine tracheal cartilage (13). The molecular weight of the intact bovine protein is 148,000; upon reduction, the protein yields subunits of M_r 54,000. CMP has been shown to interact with cartilage proteoglycan (17–19).

We describe here the identification of CMP in chicken cartilage and the isolation and sequencing of chicken CMP cDNAs. This study provides detailed structural information on CMP and establishes the presence of sequences related to the epidermal growth factor (EGF) in this matrix protein.

METHODS

Construction of pUC8–pUC9 cDNA Library. RNA was extracted from 14-day-old embryonic chicken sterna by a guanidine hydrochloride procedure (20). To aid in the extraction, sterna were mechanically disrupted in guanidine buffer by using a Polytron homogenizer. Poly(A) RNA was purified twice by oligo(dT)–cellulose chromatography and then separated by centrifugation on 5–20% sucrose gradients. Single-stranded cDNA was synthesized by using avian myeloblastosis virus reverse transcriptase (Life Sciences, St. Petersburg, FL), poly(A)-containing sternal RNA (sedimenting under 27 S in size), and oligo(dT) primer. To increase the chances of obtaining full-length transcripts, methylmercury(II) hydroxide (21) and RNasin (Promega Biotec, Madison, WI) were included. The second strand was synthesized by using the Klenow fragment of DNA polymerase I, followed by the reverse transcriptase. Synthetic *Sal*I and *Eco*RI linkers (New England Biolabs) were blunt-end-ligated to the double-stranded cDNA. To half of the double-stranded cDNA, *Sal*I linkers were added before and *Eco*RI linkers were added after nuclease S1 treatment. After *Sal*I and *Eco*RI digestion, the cDNA was cloned into the vector pUC8. The other half of the cDNA was supplied with the same linkers in reverse order and cloned in pUC9. Prior to insertion, the restricted cDNA was size-fractionated by chromatography on a Sepharose CL-4B column. cDNA strands longer than 500 bp were selected for insertion into plasmids. The vector plasmids pUC8 and pUC9 (P-L Biochemicals) were prepared by the method of Helfman *et al.* (22). cDNA was ligated to the vector at a weight ratio for vector to cDNA of 20:1.

E. coli DH-1 strain bacteria were transformed with the recombinant plasmids by the method of Hanahan (23). Transformants were then spread onto nitrocellulose filter disks, placed on agar/ampicillin plates, at a density of ≈ 1000

Abbreviations: CMP, cartilage matrix protein; EGF, epidermal growth factor.

[‡]Present address: La Jolla Cancer Research Foundation, 10901 North Torrey Pines Road, La Jolla, CA 92037.

[§]Present address: Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, H6701 Szeged, P.O. Box 521, Hungary.

[¶]To whom reprint requests should be addressed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

FIG. 1. Blot-hybridization analysis of transcripts complementary to pCMP4 insert cDNA. RNA was extracted from day 14 embryonic chicken sterna (lanes 1–2) and calvaria (lane 3) and from days 4, 5, 6, and 7 chicken limb buds (lanes 4–7). The quantities and type of RNA applied per lane were: 1, 8.0 μ g of total RNA; 2 and 3, 3.0 μ g of poly(A); 4–7, 15 μ g of total RNA. Size markers used were chicken 27S (4.6 kb) and 18S (1.5 kb) rRNA.

colonies per filter. Subsequent storage of filter-bound colonies was as described by Hanahan and Meselson (24).

Immunological Screening of the cDNA Library. Immunological screening of recombinants was done by the method of Helfman *et al.* (22). A rabbit antiserum raised against a proteoglycan-enriched fraction (A1D1-CPG1400V₀) (25) was used for the screening. The antiserum was absorbed with boiled lysate of DH-1 bacteria transformed with the plasmid pUC9 diluted 1:50 and used for screening. For the detection phase of the screening, ¹²⁵I-labeled goat anti-rabbit IgG (New England Nuclear) was used, and the immunoreactive colonies were selected following autoradiography.

Insert-encoded polypeptides were analyzed from immunoreactive bacteria extracted with NaDodSO₄/polyacrylamide gel electrophoresis buffer. The extracts were electrophoresed on NaDodSO₄/polyacrylamide gels, transferred to nitrocellulose filters (26), and assayed immunologically. Protein extracts of DH-1 bacteria transformed with the plasmid pUC9 were used as controls.

Preparation of Antibodies to Insert-Encoded Polypeptide.

FIG. 2. Indirect immunofluorescence of cartilaginous nodules that differentiated from day 4 limb bud cells cultured at high density and maintained for 6 days *in vitro*. (a) Cultures treated with anti-insert-encoded polypeptide antiserum. (b) Cultures treated with normal rabbit serum. (c and d) Phase-contrast photomicrographs for a and b, respectively.

FIG. 3. Comparison of immunological reaction of different antisera with total cartilage extract separated by NaDodSO₄/polyacrylamide gel electrophoresis and electrophoretically transferred to nitrocellulose. The antisera were anti-A1D1-CPG1400V₀ (lanes 1 and 3), anti-insert-encoded polypeptide (lanes 2 and 4), and anti-bovine CMP (lane 5). Samples were electrophoresed in the presence (lanes 1, 2, and 5) or absence (lanes 3 and 4) of dithiothreitol.

Total cellular protein from one expressing recombinant was extracted from 2 liters of the bacteria grown in liquid broth culture. Protein was precipitated from the cleared lysate prepared as described by Clewell and Helinski (27) by the addition of ammonium sulfate to a final concentration of 70% and was collected by centrifugation, solubilized, dialyzed against 20 mM Tris, pH 7.5/1.0 mM EDTA, and chromatographed on DEAE-cellulose and Sephacryl S-200. Elution was monitored immunologically. The final purification step was an electrophoretic one on NaDodSO₄/polyacrylamide gel electrophoresis. Insert-encoded protein was electroeluted from gel slices and dialyzed against 0.05 M Tris·HCl, pH 7.5/0.15 M NaCl, and $\approx 10 \mu$ g in 1.0 ml was mixed with an equal volume of Freund's complete adjuvant. The mixture was emulsified, and 0.1-ml aliquots were injected subcutaneously in areas along the back of a female New Zealand rabbit. A booster injection of the same dose of antigen in Freund's incomplete adjuvant was administered 3 weeks later. The rabbit was bled 2 weeks following the secondary immunization, and the serum was analyzed for antigen reactivity by immunoblot assay.

DNA Sequence Analysis. Nucleotide sequences were determined by sequencing both strands of the 0.9-kb insert from pCMP4 and the 1.3-kb insert of pCMP6 by the dideoxy chain-termination method of Sanger *et al.* (28). Overlapping deletions of the cDNA inserts in M13 phage vector were prepared by enzymatic digestion with BAL-31 as described (24). Some regions were sequenced by using oligonucleotide primers that were synthesized on the basis of preceding sequences.

RNA Blot Analysis. RNA was electrophoresed in the presence of formamide on 0.8% agarose/2.2 M formaldehyde/0.02 M sodium acetate gels (29), blot-transferred to nitrocellulose filters, and hybridized to nick-translated insert DNA (30). After hybridization, the filters were washed with 2 $\times$ NaCl/Cit (1 $\times$ NaCl/Cit = 0.15 M NaCl/0.015 M sodium citrate) at room temperature and consecutively with 0.2 $\times$ and 0.1 $\times$ NaCl/Cit at 65°C. The size of the hybridizing transcripts was determined relative to chicken 18S and 27S ribosomal RNA.

Extraction of Cartilage Macromolecules. Fourteen-day-old chicken embryonic sterna were stripped of their perichondrium, minced with a razor blade, and stirred in 15 vol of 4 M guanidinium hydrochloride/0.05 M sodium acetate, pH 5.8, containing 0.1 M 6-aminocaproic acid, 0.005 M benzamidine, and 0.01 M EDTA as protease inhibitors (31) for 24 hr at 4°C. The mixture was clarified by centrifugation

GTT GGG GTC ATC AAT TAC GCC AGT GCA GTC AAG AAC GAG TTC TCC CTC AAG ACT CAC CAA ACC AAA GCC GAG CTC CTG CAG GCG	84
Val Gly Val Ile Asn Tyr Ala Ser Ala Val Lys Asn Glu Phe Ser Leu Lys Thr His Gln Thr Lys Ala Glu Leu Leu Gln Ala	28
CTC GAG AGG ATA GAG CCA CTC TCC ACA GGG ACC ATG ACC GGC TTG GCC ATC CAG TTT GCC ATC AGC CGG GCC TTC AGT GAC ACA	168
Val Gln Arg Ile Glu Pro Leu Ser Thr Gly Thr Met Thr Gly Leu Ala Ile Gln Phe Ala Ile Ser Arg Ala Phe Ser Asp Thr	56
GAA GGG GCC AGG CTG AGA TCT CCC AAT ATT AAT AAG GTG GCA ATT GTT GTG ACC GAT GGA CGT CCC GAC GAG GGA GTG CAG GAC	252
Glu Gly Ala Arg Leu Arg Ser Pro Asn Ile Asn Lys Val Ala Ile Val Val Thr Asp Gly Arg Pro Gln Asp Gly Val Gln Asp	84
GTG TCA GCG AGG GCC AGG CAG GCT GGC ATT GAG ATC TTC GCC ATC GGG GTT GGC CGG GTG GAC ATG CAC ACG CTA CGG CAA ATC	336
Val Ser Ala Arg Ala Arg Gln Ala Gly Ile Glu Ile Phe Ala Ile Gly Val Gly Arg Val Asp Met His Thr Leu Arg Gln Ile	112
GCA AGC GAG CCC CTG GAC CAC GTG GAC TAC GTG GAG AGC TAC AGT GTC ATA GAG AAG CTG ACC CAC AAG TTT CAA GAA GCG	420
Ala Ser Glu Pro Leu Asp Asp His Val Asp Tyr Val Glu Ser Tyr Ser Val Ile Glu Lys Leu Thr His Lys Phe Gln Glu Ala	140
TTC TGT GTG GTG TCA GAC CTG TGT GCC ACT GGC GAC CAC GAG TGT GAG CAG ATC TGT ATC AGC ACC CCA GGA TCA TAC AAG TGT	504
Phe Cys Val Val Ser Asp Leu Cys Ala Thr Gly Asp His Asp Cys Glu Gln Ile Cys Ile Ser Thr Pro Gly Ser Tyr Lys Cys	168
GCT TGT AAA GAG GGC TTC ACG CTG AAC AAT GAT GGG AAG ACC TGC AGT GCT TGC AGT GGT GGG TCA GGA TCT GCC CTG GAT CTC	588
Ala Cys Lys Glu Gly Phe Thr Leu Asn Asn Asp Gly Lys Thr Cys Ser Ala Cys Ser Gly Gly Ser Gly Ser Ala Leu Asp Leu	196
GTT TTC CTG ATC GAT GGC TCC AAG AGT GTT CGG CCT GAG AAC TTT GAG CTG GTG AAG AAA TTC ATC AAC CAA ATT GTG GAA TCC	672
Val Phe Leu Ile Asp Gly Ser Lys Ser Val Arg Pro Glu Asn Phe Glu Leu Val Lys Phe Ile Asn Gln Ile Val Glu Ser	224
CTG GAA GTG TCA GAG AAG CAG GCC CAG GTT GGC GTT GGT CAG TAC TCC AGC TCT GTC AGG CAG GAG TTT CCT CTG GGA CAG TTC	756
Leu Glu Val Ser Glu Lys Gln Ala Gln Val Gly Leu Val Gln Tyr Ser Ser Ser Val Arg Gln Glu Phe Pro Leu Gly Gln Phe	252
AAG AAC AAG AAG GAC ATC AAA GCA GCA GTC AAG AAA ATG GCC TAC ATG GAG AAA GGC ACC ATG ACA GGC CAG GCT CTG AAG TAC	840
Lys Asn Lys Lys Asp Ile Lys Ala Ala Val Lys Met Ala Tyr Met Glu Lys Gly Thr Met Thr Gly Gln Ala Leu Lys Tyr	280
CTC GTG GAC AGT TCC TTT TCC ATT GCT AAT GGG GCT AGG CCT GGG GTT CCC AAG GTG GGC ATA GTC TTC ACA GAT GGA CGG TCA	924
Leu Val Asp Ser Ser Phe Ser Ile Ala Asn Gly Ala Arg Pro Gly Val Pro Lys Val Gly Ile Val Phe Thr Asp Gly Arg Ser	308
CAA GAT TAC ATC ACT GAT GCT GCT AAG AAA GCC AAA GAT TTA GGC TTT AGG ATG TTT GCT GTG GGA GTT GGT AAC GCA GTG GAG	1008
Gln Asp Tyr Ile Thr Asp Ala Ala Lys Lys Ala Lys Asp Leu Gly Phe Arg Met Phe Ala Val Gly Val Gly Asn Ala Val Glu	336
GAT GAG CTG AGG GAA ATT GCT TCC GAA CCA GTA GCT GAG CAC TAC TTC TAC ACA GCT GAC TTC AGA ACC ATC AGC AAC ATT GGG	1092
Asp Glu Leu Arg Glu Ile Ala Ser Glu Pro Val Ala Glu His Tyr Phe Tyr Thr Ala Asp Phe Arg Thr Ile Ser Asn Ile Gly	364
AAG AAG CTG CAA ATG AAA ATC TGC GTT GAG GAA GAT CCA TGC GAA TGT AAA TCT ATA GTG AAG TTC CAG ACA AAA GTT GAA GAA	1176
Lys Lys Leu Gln Met Lys Ile Cys Val Glu Glu Asp Pro Cys Glu Cys Lys Ser Ile Val Lys Phe Gln Thr Lys Val Glu Glu	392
CTC ATC AAT ACA CTG CAA CAG AAA TTG GAA GCT GGT GCA AAA AGG ATT GAA GCC CTG GAG AAT AAG ATC ATC TAG GAC CCC AGA	1260
Leu Ile Asn Thr Leu Leu Gln Lys Leu Glu Ala Val Ala Lys Arg Ile Glu Ala Leu Glu Asn Lys Ile Ile Stp	416
ACA GTA CTA TTT CCG TCT TTC CTA ATG TAA CAA AAC TAG AAT TC	1304

FIG. 4. Nucleotide sequence and predicted amino acid sequence of that portion of chicken CMP encoded by a 1.3-kb stretch of cDNA. The region of the CMP that has homology with the mouse epidermal EGF precursor (amino acids 366–401) is underlined. Double basic amino acids are boxed.

at $12,000 \times g$ for 30 min, and the supernatant was dialyzed against the Tris/NaCl buffer at 4°C.

Cell Culture and Immunocytochemistry. Dissociated limb buds from stage 23–24 embryos (32) were plated at a density of 5×10^6 cells per 35-mm culture plate and cultured as described earlier (33). Under these conditions the mesodermal cells undergo chondrogenesis *in vitro*. Cultures were fixed in 2% paraformaldehyde, washed in phosphate-buffered saline, and incubated either with normal rabbit serum or with the antiserum directed against the insert-encoded polypeptide. The cultures were then incubated with fluorescein-conjugated goat anti-rabbit IgG and examined with a Nikon Labophot equipped with an epiilluminator and a 590-nm filter.

RESULTS

Screening of Expressing cDNA Library. Approximately 50,000 transformant colonies were screened by the double-antibody technique with anti-A1D1-CPG-1400V₀ antiserum. Four individual clones that reacted with the antiserum were isolated. These were later designated pCMP1 to pCMP4. Of

CMP	C A T G D H D C E Q I C I S T P G S Y K	167
EGF	C A T Q N H G C T L G C E N T P G S Y H	385
CMP	C A C K E G F T L N N D G K T C	183
EGF	C T C P T G F V L L P D G K Q C	401

FIG. 5. Comparison of amino acid sequences from chicken CMP (amino acids 148–183) and mouse EGF precursor (amino acids 366–401). Single letter amino acid symbols are used. Identical residues are boxed.

these, pCMP4 contained the largest cDNA insert. To isolate additional related clones not detected by the original screening, the cDNA library was rescreened by using ^{32}P -labeled pCMP4 insert DNA as a probe. An additional nine positive clones were identified. Two of these clones had inserts larger than the pCMP4 insert. One of these clones (pCMP5) was found to be immunoreactive, whereas the second (pCMP6) was not. It was assumed that clone pCMP6 did not express because its insert was not in the proper reading frame. The reading frame of the insert was shifted by nuclease S1 treatment of the *Sal* I-cleaved plasmid. Upon recircularization, the pCMP6 plasmid derivatives were used to transform *E. coli*. The resulting transformants were screened immunologically, and two clones, pCMP6D1 and pCMP6D2, were selected. The insert-encoded polypeptides from the six recombinants had a molecular weight ranging from 29,000 to 49,000.

Characterization of Inserts from the Immunoreactive Recombinants. Insert cDNA from pCMP4 cross-hybridized with filter-bound insert DNA from all six clones (data not shown). This suggests that the inserts, which are related to each other on the basis of the immunological properties of their translation products, are derived from the same mRNA species. The sizes of the inserts ranged from 0.8 to 1.3 kb. Partial restriction map analysis of five of the clone insert cDNAs also revealed identity in overlapping regions. After alignment of the restriction maps, it was clear that the inserts differ only in the distance they extend in the 5' direction away from the *Eco*RI border.

Tissue Specificity and Developmental Regulation. To obtain information about the mRNA complementary to the cDNAs, insert cDNA from one of the clones (pCMP4) was ^{32}P -labeled

FIG. 6. Comparison of the amino acid sequence of two homologous regions found in chicken CMP. Single letter amino acid symbols are used. Identical residues are boxed. Chemically similar residues are indicated by dashed boxes. Asterisks indicate spaces inserted to optimize homology.

by nick translation and then hybridized to sternal and calvarial mRNA in a blot hybridization experiment. The probe detected a 3.4-kb transcript abundant in sternal RNA but absent in calvarial RNA (Fig. 1, lanes 1–3). A second, lower molecular weight sternal transcript was faintly detectable as well.

RNA analysis was also done to examine the developmental expression of transcripts complementary to pCMP4. The transcript was not detectable in RNA from day 4 limb buds (Fig. 1, lanes 4–7). It was first detectable in RNA from day 5 limb buds and in increasing amounts in day 6 and day 7 limb RNA. This pattern parallels the histological appearance of cartilage in the developing chicken embryonic limb and also correlates with the expression of type II collagen (34, 35).

Since the appearance of transcripts complementary to the pCMP cDNA coincides temporally with the histological appearance of cartilage in the limb, the availability of the antiserum to the protein whose synthesis is directed from the pCMP cDNA allowed us to investigate directly the tissue specificity of the translation product. Stage 23–24 chicken limb buds were dissociated, and the cells were plated at a density that allows chondrogenesis to take place. When these cultures were analyzed immunocytochemically with the anti-insert-encoded polypeptide antiserum, immunoreactivity was found to be associated with the differentiated cartilaginous nodules (Fig. 2). No immunoreactivity was visible in the noncartilaginous internodular areas. These results indicate that the transcripts and translation products coded for by sequences in the pCMP inserts are associated with the differentiation of chondrocytes.

Immunological Identity of Insert-Encoded Protein with CMP. To confirm that the clones corresponded to CMP, an antiserum was generated against the M_r 34,000 insert-encoded polypeptide synthesized by pCMP4-containing bacteria and was used to probe cartilage proteins on immunoblots. The antiserum reacted specifically with a M_r 54,000 protein in 4 M guanidine extracts of embryonic chicken cartilage and showed no cross-reactivity with other cartilage proteins (Fig. 3). When nonreduced extracts were tested, the antiserum

reacted with a high molecular weight species that corresponded to the unreduced species with which the antiserum used for the original screening of cDNA clones had reacted.

A rabbit antiserum against bovine CMP reacted with a M_r 54,000 protein in chicken cartilage extracts that had the same mobility as the protein detected with the antiserum against the insert-encoded polypeptide (Fig. 3). Moreover, this antiserum also detected the insert-encoded polypeptides in the same assay (data not shown). These results establish that the chicken M_r 54,000 protein is immunologically similar to bovine cartilage matrix protein and that the latter protein shares at least one immunological determinant with the insert-encoded polypeptides.

Sequence Analysis. The nucleotide sequence and the deduced amino acid sequence for chicken CMP are presented in Fig. 4. The sequence, derived from two overlapping clones, extends 1.3 kilobases and contains a single open reading frame that encodes a polypeptide with a calculated M_r of 45,586 and an estimated pI of 7.75. The inferred sequence lacks an initiation codon and preceding signal sequence and, therefore, is incomplete. Since the molecular weight calculated from the sequence and the molecular weight of the CMP differ only by 8000, the cDNA sequence should represent at least 80% of the protein sequence. Following the stop codon are 48 bp of the 3' untranslated region. The *Eco*RI site that flanks this region is most likely endogenous because all of the CMP clones examined begin there and not at the poly(A) tail.

The CMP polypeptide contains a sequence of 36 amino acids (residues 148–183) that has a 55.6% identity with the repeat sequence (residues 366–401) of mouse EGF precursor (36) (Fig. 5). The identity includes the same number and spacing of cysteine residues, and the arrangement is like that found in the B-type EGF-like segment of blood coagulation factors (37). Located on the COOH-terminal side of the EGF-like region of the CMP are six potential dibasic cleavage sites (38).

An analysis of the nucleotide and amino acid sequences reveals that CMP contains a large homologous repeat sequence (Fig. 6). Amino acid sequences 1–171 and 234–381 have a homology of 39%. If allowance is made for conservative substitutions, the homology approaches 55%. A comparison at the nucleotide level for the same regions reveals a homology of 50%. In the alignment of the two homologous segments, spaces were introduced in the second repeat to optimize homology between the EGF-like sequence of the first repeat and a possible vestige of an EGF-like sequence (amino acids 371–381) in the second repeat.

No consensus N-glycosylation sequences Asn-Xaa-Ser-Thr (39) are present in the CMP amino acid sequence.

DISCUSSION

The present communication describes the isolation of cDNA clones that encode a large portion of the chicken cartilage matrix protein. These clones were isolated from an expressing cDNA library by using antiserum that had been prepared against a cartilage proteoglycan fraction (25) and that reacts with a M_r 54,000 protein from cartilage. The clones we describe here are related to this latter protein as demonstrated by the reactivity of an antiserum prepared against an insert-encoded protein.

The conclusion as to the identity of the clones is based on the following observations. The antibodies used to select the cDNA clones react with a protein from chicken cartilage that has similar characteristics to the bovine CMP. For instance, both the chicken and the bovine protein have nearly identical molecular sizes as determined on reducing NaDodSO₄/polyacrylamide gels. Both also form disulfide-crosslinked supramolecular weight aggregates. However, from the pres-

ent data, it is not clear whether the chicken CMP aggregate is an oligomer of M_r 54,000 subunits or if it is interacting with some other matrix component(s). Paulsson and Heinegård (18) reported that the bovine CMP formed a disulfide-bonded trimer with a M_r of 148,000. Although the chicken CMP forms similar disulfide-bonded multimers, their size differed from those of the bovine species. In unreduced chicken cartilage extracts, the antibodies bound predominantly to species having molecular weights greater than 200,000. Finally, the reactivity of the chicken protein with an antiserum directed against bovine CMP indicates that the two molecules are related. A further similarity of the chicken and bovine proteins is that, as with the bovine CMP (13), the chicken M_r 54,000 protein and cartilage proteoglycan can be shown to interact by using DEAE-cellulose chromatography (data not shown). Recently, a M_r 54,000 protein from rat chondrosarcoma has been reported (40). This molecule interacts with collagen as well as proteoglycan. Interestingly, we have evidence that CMP also binds to collagen. However, the rat chondrosarcoma protein differs from CMP in that it does not form disulfide-bonded multimers (40).

The hybridization of the cDNAs to a transcript in sternal, but not in calvarial, RNA and the appearance of the transcript concomitant with chondrogenesis indicates that CMP expression is cartilage specific and developmentally regulated. The conclusion was reinforced by the immunological data showing that the protein product was detectable only after *in vitro* chondrogenesis had taken place. The association of mRNA expression with cartilage and its absence in the other tissues analyzed supports the identity of the cDNA clones as CMP clones because Paulsson and Heinegård (18) have found CMP antigen to be restricted to cartilage. Our results extend this analysis to the mRNA level.

The amino acid sequence deduced from our clones represents the major portion of the mature chicken CMP. There are two striking features in the structure of CMP. The first is that the protein contains a homologous repeat sequence. These repeats are quite large, with each comprising over a third of the protein. The homology between the repeats is 39%. This high degree of similarity indicates that these regions have arisen as a result of duplication of common gene sequences. Perhaps this arrangement is the basis for some bifunctional property of the molecule. The second feature of the CMP structure is that the amino acid sequence contains an EGF-like sequence. While nothing is known about their function, EGF-like sequences have been found in other proteins (37, 41, 42). Interestingly, none of these proteins are extracellular matrix proteins. It is tempting to speculate that the presence of the EGF-like sequence in CMP might make this matrix molecule a growth promoter.

The chicken CMP sequence reported here will allow synthesis of peptides for generation of antibodies to specific regions of the protein. Such antibodies as well as the cDNA clones for CMP provide new probes for the study of chondrogenesis. Additionally, these probes will be useful for investigations of the function of CMP as a component of the extracellular matrix.

We thank Dr. John Baker (University of Alabama) for providing us with antiserum against bovine CMP. This work was supported by Grants HD 09174 and HD 22016 from the National Institute of Child Health and Human Development. The paper is Scientific Contribution no. 1143 of the Storrs Agricultural Experiment Station, University of Connecticut, Storrs, CT 06268.

1. Zwilling, E. (1968) *Dev. Biol. Suppl.* 2, 184–207.
2. Goetinck, P. F. (1983) in *Cartilage*, ed. Hall, B. K. (Academic, New York), Vol. 3, pp. 165–189.
3. Grobstein, C. (1975) in *Extracellular Matrix Influences on Gene*

- Expression*, eds. Slavkin, H. C. & Grenlich, R. C. (Academic, New York), pp. 9–16.
4. Bernfield, M., Banerjee, S. D., Koda, J. E. & Rapraeger, A. C. (1984) in *The Role of the Extracellular Matrix in Development*, ed. Trelstad, R. (Liss, New York), pp. 545–572.
5. LeDourain, N. M., Cochard, P., Vincent, M., Duband, J., Tucker, G., Teillet, M. & Thiery, J. (1984) in *The Role of Extracellular Matrix in Development*, ed. Trelstad, R. L. (Liss, New York), pp. 373–398.
6. Hascall, V. C. (1981) in *Biology of Carbohydrates*, ed. Ginsberg, C. (Wiley, New York), Vol. 1, pp. 1–49.
7. Miller, E. J. & Matukas, V. J. (1969) *Proc. Natl. Acad. Sci. USA* 64, 1264–1268.
8. Noro, A., Kimata, K., Oike, Y., Shinomura, T., Maeda, N., Yano, S., Takahashi, N. & Suzuki, S. (1983) *J. Biol. Chem.* 258, 9323–9331.
9. Shinomura, T., Kimata, K., Oike, Y., Noro, A., Hirose, N., Tanabe, K. & Suzuki, S. (1983) *J. Biol. Chem.* 258, 9314–9322.
10. Rosenberg, L. C., Choi, H. U., Tang, L., Johnson, T. L., Pas, S., Webber, C., Reiner, A. & Poole, R. (1985) *J. Biol. Chem.* 260, 6304–6313.
11. McKeown-Longo, P. J., Sparks, K. J. & Goetinck, P. F. (1982) *Collagen Relat. Res.* 2, 232–244.
12. Hewitt, A. T., Kleinman, H. K., Pennypacker, J. P. & Martin, G. R. (1980) *Proc. Natl. Acad. Sci. USA* 77, 385–388.
13. Paulsson, M. & Heinegård, D. (1981) *Biochem. J.* 197, 367–375.
14. Neame, P. J., Périn, J.-P., Bonnett, F., Christner, J. E., Jollès, P. & Baker, J. R. (1985) *J. Biol. Chem.* 260, 12402–12404.
15. Deák, F., Kiss, I., Sparks, K. J., Argraves, W. S., Hampikian, G. & Goetinck, P. F. (1986) *Proc. Natl. Acad. Sci. USA* 83, 3766–3770.
16. Doege, K., Hasell, J. R., Caterson, B. & Yamada, Y. (1986) *Proc. Natl. Acad. Sci. USA* 83, 3761–3765.
17. Paulsson, M. & Heinegård, D. (1984) *Collagen Relat. Res.* 4, 219–229.
18. Paulsson, M. & Heinegård, D. (1979) *Biochem. J.* 183, 539–545.
19. Lord, N. D., Baker, J. R. & Christner, J. E. (1984) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* 43, 1695a (abstr.).
20. Adams, S. L., Sobel, M. B., Howard, B. H., Olden, K., Yamada, K. M., de Crombrughe, B. & Pastan, I. (1977) *Proc. Natl. Acad. Sci. USA* 74, 3399–3403.
21. Payvar, F. & Schimke, R. T. (1979) *J. Biol. Chem.* 254, 7636–7642.
22. Helfman, D. M., Feramisco, J. R., Fiddes, J. C., Thomas, G. P. & Hughes, S. H. (1983) *Proc. Natl. Acad. Sci. USA* 80, 32–35.
23. Hanahan, D. (1983) *J. Mol. Biol.* 166, 557–580.
24. Hanahan, D. & Meselson, M. (1983) *Methods Enzymol.* 100, 333–342.
25. Sparks, K. J., Lever, P. L. & Goetinck, P. F. (1980) *Arch. Biochem. Biophys.* 199, 579–590.
26. Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* 76, 4350–4354.
27. Clewell, D. B. & Helinski, D. R. (1972) *J. Bacteriol.* 110, 1135–1141.
28. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
29. Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977) *Biochemistry* 16, 4743–4751.
30. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5201–5205.
31. Oegema, T. R., Jr., Hascall, V. C. & Dziewiatkowski, D. D. (1975) *J. Biol. Chem.* 250, 6151–6159.
32. Hamburger, V. & Hamilton, H. L. (1951) *J. Morphol.* 88, 49–92.
33. Goetinck, P. F., Pennypacker, J. P. & Royal, P. D. (1974) *Exp. Cell Res.* 87, 241–248.
34. Upholt, W. B., Vertel, B. M. & Dorfman, A. (1979) *Proc. Natl. Acad. Sci. USA* 76, 4847–4851.
35. Deák, F., Argraves, W. S., Kiss, I., Sparks, K. J. & Goetinck, P. F. (1985) *Biochem. J.* 229, 189–196.
36. Scott, J., Urdea, M., Quiroga, M., Sanchez-Pescador, R., Fong, N., Selby, M., Rutter, W. J. & Bell, G. I. (1983) *Science* 221, 236–240.
37. Doolittle, R. F., Feng, D. F. & Johnson, M. S. (1984) *Nature (London)* 307, 558–560.
38. Steiner, D. F., Quinn, P. S., Chan, S. J., March, J. & Tager, H. S. (1980) *Ann. N.Y. Acad. Sci.* 243, 1–16.
39. Hughes, R. C. (1973) *Progr. Biophys. Mol. Biol.* 26, 189–268.
40. Chandrasekhar, S., Laurie, G. W., Cannon, F. B., Martin, G. R. & Kleinman, H. K. (1986) *Proc. Natl. Acad. Sci. USA* 83, 5126–5130.
41. Sudhof, T. C., Goldstein, J. L., Brown, M. S. & Russell, D. W. (1985) *Science* 228, 893–895.
42. Bender, W. (1985) *Cell* 43, 559–560.