Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 6040–6044, September 1987
Biochemistry

Type XII collagen: Distinct extracellular matrix component discovered by cDNA cloning

(recombinant DNA/nucleotide sequencing/collagen genes)

MARION K. GORDON, DONALD R. GERECKE, AND BJÖRN R. OLSEN

Department of Anatomy and Cellular Biology, Harvard Medical School, Boston, MA 02115

Communicated by Elizabeth D. Hay, May 7, 1987 (received for review March 18, 1987)

ABSTRACT **We have screened a cDNA library constructed from tendon fibroblast mRNA for the presence of collagenous coding sequences. Nucleotide sequence analysis of one isolated clone, pMG377, reveals that the clone encodes a polypeptide that is homologous to, yet distinctly different from, type IX short-chain collagen polypeptides. The structure of the conceptual translation product of the cDNA is also different from that of all other collagen types. Therefore, we have given the type IX-like collagen chain encoded by pMG377 the designation α1(XII). Ribonuclease protection assays with single-stranded cRNA probes demonstrate that α1(XII) mRNA is present in several tissues such as calvaria, tendon, and sternal cartilage of 17-day-old chicken embryo and in cornea from 6-day-old embryos. Using pMG377 as the hybridization probe, we isolated a fragment of the corresponding gene from a chicken genomic library. Partial nucleotide sequence analysis of the genomic clone DG12 shows that the exon/intron structure of the α1(XII) collagen gene appears to be homologous to that of the α1(IX) and α2(IX) collagen genes. Our data demonstrate that types IX and XII collagen are two homologous members of a family of unique collagenous proteins that show tissue-specific patterns of expression. Based on their structure and the properties of their genes, we conclude that this family of collagens is distinctly different from that of fibrillar collagens.**

The extracellular matrix components classified as collagens on the basis of their triple-helical domain constitute a heterogeneous group of proteins with a variety of molecular structures and interaction properties. Within this group, the fibrillar collagens I, II, III, and V form characteristic "quarter-staggered" fibrils in connective tissues (for a review, see ref. 1). However, types IV, VI, VII, VIII, IX, and X collagens do not form these kinds of fibrils and have molecular structures distinct from those of fibrillar collagens. We have isolated genes encoding the polypeptides of types IX and X collagens, and we have found that the genes for these two "short-chain" collagens also have an exon structure that is clearly distinct from that of fibrillar collagen genes (2, 3).

Although the function of type IX collagen in extracellular matrices is not known, some intriguing structural characteristics of this collagen are known. First, the molecules do not contain one long triple-helical domain like fibrillar collagens but instead are composed of three triple-helical domains interspersed with noncollagenous domains (4). Second, type IX collagen contains a sulfated glycosaminoglycan side-chain covalently attached to the polypeptide subunit designated the α2(IX) chain. This gives type IX collagen a proteoglycan character (5–8).

cDNA probes corresponding to the α1(IX) and α2(IX) chains (9, 10) have been used to demonstrate that type IX collagen mRNAs are expressed in a tissue-specific manner—i.e., they are expressed by chicken embryo chondrocytes but not by chicken embryo fibroblasts (9, 11). However, tendon fibroblasts do contain mRNAs capable of synthesizing bacterial collagenase-sensitive polypeptides that are similar in size to type IX polypeptides (11). To explore the nature of these polypeptides in detail, we have screened a cDNA library constructed from tendon fibroblast mRNA for the presence of collagenous coding sequences. Here we report on the isolation and characterization of one such clone, pMG377. Nucleotide sequence analysis reveals that the clone encodes a collagenous polypeptide that is homologous to, yet distinctly different from, type IX collagen polypeptides synthesized by chondrocytes. The structure of the conceptual translation product of pMG377 is different from that of collagens I–X and type XI (1α, 2α) chains. Therefore, we have given the type IX-like collagen chain encoded by pMG377 the designation α1(XII). Ribonuclease protection assays using cRNA derived from pMG377 as probe show that α1(XII) mRNA is present in several tissues, such as calvaria, tendon and sternal cartilage of 17-day-old chicken embryos, and cornea from 6-day-old embryos.

To investigate the homology between the genes encoding types XII and IX collagen, we used pMG377 as a hybridization probe to isolate a fragment of the corresponding gene from a chicken genomic library. Partial nucleotide sequence analysis of the genomic clone DG12 demonstrates that it encodes part of the α1(XII) collagen chain and that the exon/intron structure of the gene appears to be homologous to that of the α1(IX) and α2(IX) collagen genes.

Our results demonstrate that types IX and XII collagen are members of a family of unique collagenous proteins that show tissue-specific expression and are distinctly different from fibrillar collagens.

MATERIALS AND METHODS

Isolation of RNA, Cell-Free Translations, and cDNA Synthesis. Total RNA from 6-day-old chicken embryo cornea and from 15- or 17-day-old tendons, sterna, and calvarial bones was isolated by using the guanidine thiocyanate method (12) as adapted by Maniatis *et al.* (13). Poly(A)$^+$ RNA, obtained by oligo(dT)-cellulose chromatography (14), was fractionated by centrifugation through low-salt sucrose gradients (15). Fractions were dialyzed versus water to remove sucrose, quantitated by absorbance at 260 nm, and translated in a rabbit reticulocyte cell-free system (Amersham) with 20 μCi (1 Ci = 37 GBq) of [^{35}S]methionine as the radiolabel. Half of each translation reaction was made 25 mM in *N*-ethylmaleimide and 50 mM in $CaCl_2$ for incubation with 3 units of bacterial collagenase. Following the 45-min 37°C collagenase digestions, both collagenase-treated and nontreated translation samples were diluted into $NaDodSO_4$/PAGE sample buffer containing 5% 2-mercaptoethanol for electrophoresis through $NaDodSO_4$/10% polyacrylamide gels (16).

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.

Fractionated RNA that translated into 50- to 100-kDa collagenase-sensitive polypeptides was used as template for cDNA synthesis. The double-stranded cDNA was prepared by using a kit from Amersham, based on the method of Gubler and Hoffman (17). The cDNA was dC-tailed, hybridized to dG-tailed pBR322 (Bethesda Research Laboratories), and used to transform *Escherichia coli* strains MC1061 or HB101.

Screening of cDNAs. Minipreparative isolations (18) of DNA were performed on approximately 1000 transformants. The isolated plasmid DNAs were digested with restriction enzymes and electrophoresed on agarose gels. The 77 cDNAs with the largest inserts were chosen for subsequent analysis by *Sau*96I digestion. After incubation with 1 unit of *Sau*96I, 20-μl samples were treated with 0.5 μg of RNase A for 10 min at 37°C and then radiolabeled with 2 μCi of [α-^{32}P]dGTP by the Klenow fragment of DNA polymerase I for 10 min at room temperature (9). After electrophoresis at 300 V through 10% polyacrylamide gels, the gels were exposed to x-ray film to determine which cDNAs exhibited the *Sau*96I "ladder" pattern typical of collagens. Southern blots (19) of these *Sau*96I-positive cDNAs were hybridized to α1(I) and α2(I) nick-translated probes by the sarkosyl method (20). The pMG377 cDNA insert, which did not hybridize to type I collagen probes, was mapped with restriction enzymes and subjected to nucleotide sequence analysis (21).

RNase Protection Assays. The two *Ava* I fragments of the pMG377 insert were cloned into pSP64 and pSP65 vectors (shown in Fig. 4). Complementary RNAs were transcribed (22) from linearized pSP6 plasmids and from a pSP62 plasmid containing DNA corresponding to the C-propeptide region of chicken α2(I) collagen and from a pSP62 plasmid containing DNA from the α2(IX) cDNA, pYN1731 (10). The transcribed cRNAs, which were labeled to a high specific activity with ^{32}P-labeled nucleotides (23), were hybridized to 50 μg of total RNA from various chicken embryonic tissues. After hybridization in 80% formamide/0.4 M NaCl/40 mM Pipes buffer, pH 6.4/1 mM EDTA at 45°C (24) for 8–10 hr, the RNA·RNA hybrids were digested with RNases A and T1 (24) at 30°C for 30 min. After further processing of the samples as described (25), the samples were electrophoresed through 4% denaturing polyacrylamide gels, which were dried before exposure to x-ray film.

Screening of the Chicken Genomic DNA Library. The insert of pMG377, isolated from a *Pst* I digest of the plasmid DNA by preparative agarose gel electrophoresis, was labeled by nick-translation and used as the probe to screen a chicken genomic library. This library was constructed by using chicken genomic DNA fragments generated by partial *Eco*RI digestion and cloned in the bacteriophage λ Charon 4A vector (unpublished work). Filters were screened with the hybridization probe in the presence of sarkosyl. Phage purification and recombinant DNA isolations were performed as described elsewhere (13).

Construction of Plasmid Genomic Subclones. The DNA of the recombinant genomic clone DG12 was digested with *Eco*RI. The resulting fragments were inserted into the *Eco*RI site of pBR322, and isolated subclones were characterized by restriction endonuclease analysis and Southern blotting. For nucleotide sequence analysis, restriction endonuclease fragments of the genomic subclone pDG30 were inserted in both orientations into M13mp18 and M13mp19 by standard techniques and were sequenced (26).

RESULTS AND DISCUSSION

Cell-free translations of mRNA isolated from 15- and 17-day-old chicken embryonic tendons showed that it directed the synthesis of type I preprocollagen chains as well as short bacterial collagenase-sensitive polypeptides that migrated between the 92.5-kDa globular molecular mass marker and the α2(I) preprocollagen chain (data not shown). To obtain further enrichment of the mRNAs specific for the short collagenous polypeptides and to separate such mRNAs from those of type I preprocollagen, tendon mRNAs were size-fractionated by centrifugation through a sucrose gradient. RNA recovered from the fractions of this gradient were translated in Amersham's rabbit reticulocyte cell-free system and analyzed by NaDodSO$_4$/PAGE. The results (Fig. 1) clearly showed that mRNAs specific for the short (≈100 kDa) collagenous polypeptides could be separated from the mRNAs of type I preprocollagen. The results also suggested that the short collagenous polypeptides did not represent premature terminations of translations of larger collagenous (i.e., type I procollagen) mRNAs.

Sucrose-gradient fractionated mRNA was used to construct a cDNA library with pBR322 as vector. The library was screened by a combination of procedures. (*i*) Approximately 1000 cDNAs were isolated on a small scale ("minipreps"), digested with the restriction endonuclease *Hin*dIII, and analyzed by electrophoresis in 1.5% agarose gels to determine the approximate size of the recombinant plasmid DNAs. (*ii*) Seventy-seven plasmids had inserts of 500 base pairs (bp) or larger and were chosen for digestion with the restriction endonuclease *Sau*96I. Due to the Gly-Pro-Xaa repeats, cDNAs corresponding to collagens contain frequent GGN-CCN codons. *Sau*96I cleaves at GGNCC and, therefore, digests collagen cDNAs to fragments that are multiples of 9 bp. Of the 77 clones, 10 produced the *Sau*96I ladder pattern typical of collagen cDNAs, as seen by electrophoresis on acrylamide gels (Fig. 2). One of these clones, pMG377, is described here.

pMG377 was digested with *Eco*RI and *Hin*dIII for electrophoresis through 1.5% agarose gels, blotted onto nitrocellulose, and probed with nick-translated α1(I) and α2(I) cDNA probes. The results showed that it did not hydridize to type

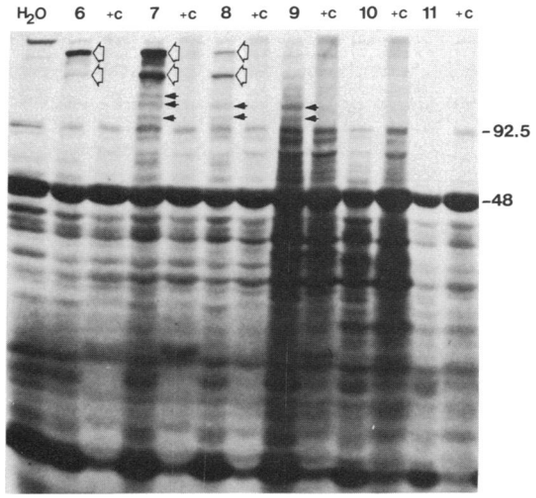
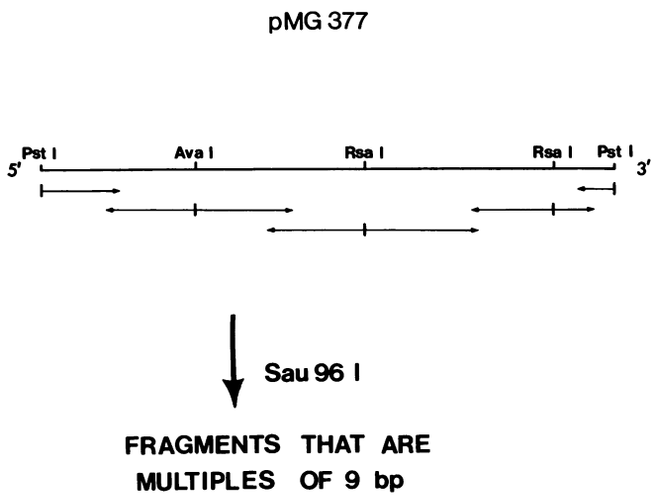


FIG. 1. Polyacrylamide gel electrophoresis (10% gel) of [^{35}S]-methionine-labeled cell-free translation products. Poly(A)$^+$ RNA was isolated from 15-day-old chicken embryo tendons and size-fractionated on sucrose gradients. RNA from each fraction (0.017 μg) was translated in a final volume of 20 μl. Half of each translation reaction was treated with bacterial collagenase (lanes +c) before electrophoresis. Note that mRNAs coding for type I preprocollagen chains are mostly found in fractions 6, 7, and 8 (open arrows), while mRNAs encoding shorter collagenase-sensitive polypeptides (compare to +c lanes) are enriched in fractions 7, 8, and 9 (arrowheads). Molecular weights are indicated on the right. Note that the intensities of translation products in different lanes may not accurately reflect the levels of the corresponding mRNAs in the total poly(A)$^+$ population.



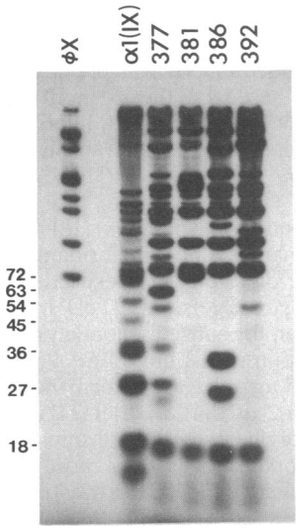


FIG. 2. (*Right*) Polyacrylamide gel electrophoresis of cDNAs digested with *Sau*96I and labeled with ^{32}P as described. As control for the collagenase ladder pattern, the pYN1738 cDNA, specific for α1(IX) collagen, was used. The leftmost lane contains φX174 *Hae* III restriction fragments, and the positions of fragments 18–72 bp long are indicated by the numbers along the left side of the autoradiograph. (*Left*) Restriction endonuclease-cleavage map of the insert of the recombinant clone pMG377. Below the map, arrows indicate the direction of sequence analysis after end-labeling of DNA fragments at different restriction endonuclease sites.

I mRNAs (not shown). Therefore, pMG377 was subjected to nucleotide sequence analysis. A restriction endonuclease cleavage map of pMG377 and the strategy for sequencing are shown in Fig. 2; the result of the sequence analysis is shown in Fig. 3. The conceptual translation product encoded by the insert of pMG377 proved to be unlike any other known collagen and displayed several unique characteristics: (*i*) a short noncollagenous domain, called NC1, of 28 amino acid residues at the carboxyl terminus; (*ii*) two imperfections in the triple-helical domain called COL1, which is found on the amino side of the NC1 domain; and (*iii*) the presence of a cysteinyl residue at both the junction between COL1 and NC1 and also at the fifth residue into NC1.

```
          630
pMG377    AGA GGA GAG CCT GGA CCT GGA GGG AGA CCA GGA TTC
α1(XII)   Arg-Gly-Glu-Pro-Gly-Pro-Gly-Gly-Arg-Pro-Gly-Phe
α1(IX)    Leu  *  Pro Arg  *  Leu Pro  *  Leu Lys  *  Pro
α2(IX)    Met  *  Pro Arg  *  Val Pro  *  Leu Leu  *  Ala

          642
pMG377    CCA GGA CCT CCT GGG TTC CAA GGA CCA CCT GGT GAA
α1(XII)   Pro-Gly-Pro-Pro-Gly-Phe-Gln-Gly-Pro-Pro-Gly-Glu
α1(IX)     *   *  Glu Ile  *  Arg Lys  *   *  Lys  *   *
α2(IX)    Ala  *  Gln Ile  *  Asn Ile  *   *  Lys  *  Lys

          654
pMG377    AGA GGA ATG CCT GGA GAG AAA GGT GAA AGA GGG ACT
α1(XII)   Arg-Gly-Met-Pro-Gly-Glu-Lys-Gly-Glu-Arg-Gly-Thr-
α1(IX)    Ala  *  Glu Arg  *   *  Arg  *  Phe Pro  *  Arg
α2(IX)     *   *  Glu Lys  *   *  Arg  *  Asp Thr  *  Arg

          666
pMG377    GGA TCT CAA GGA CCA CGA GGT TTG CCA GGA CCA CCT
α1(XII)   Gly-Ser-Gln-Gly-Pro-Arg-Gly-Leu-Pro-Gly-Pro-Pro-
α1(IX)     *  Val Lys  *  Leu Pro  *  Pro Arg  *  Leu  *
α2(IX)     *  His Pro  *  Met Pro  *  Pro Pro  *  Ile  *

          678
pMG377    GGT CCA CAG GGA GAA TCC AGA ACA GGT CCA CCA GGC
α1(XII)   Gly-Pro-Gln-Gly-Glu-Ser-Arg-Thr-Gly-Pro-Pro-Gly-
α1(IX)     *  Glu Pro  *  Lys Pro Ser Tyr  *  Arg Glu  *
α2(IX)     *  Leu Pro  *  Ile Pro Gly His Ala Leu Asp  *

          690
pMG377    TCC ACA GGA TCA CGA GGA CCA CCC GGG CCT CCA GGT
α1(XII)   Ser-Thr-Gly-Ser-Arg-Gly-Pro-Pro-Gly-Pro-Pro-Gly-
α1(IX)    Arg Asp  *  Val  *   *   *   *  Val Ala  *
α2(IX)    Lys Asp  *  Glu  *   *   *   *  Val  *   *
α3(IX)        Asp  *  Ala Gln  *  Glu  *   *  Leu  *   *

          702
pMG377    CGT CCT GGA AAT GCG GGT ATT CGA GGT CCT CCA GGG
α1(XII)   Arg-Pro-Gly-Asn-Ala-Gly-Ile-Arg-Gly-Pro-Pro-Gly-
α1(IX)    Gln  *   *  Ile Pro  *  Pro Pro  *   *   *   *
α2(IX)    Asp Ala  *  Arg Pro  *  Ser Pro  *   *  Ala  *
α3(IX)    Asp  *   *  Thr Pro  *  Ala Val  *  Ala Gln  *

          714
pMG377    CCA CCA GGA TAT TGT GAT TCA TCC CAG TGT GCT AGC
α1(XII)   Pro-Pro-Gly-Tyr-Cys-Asp-Ser-Ser-Gln-Cys-Ala-Ser-
α1(IX)     *   *   *   *   *  Glu Pro  *  Ser  *  Arg Met
α2(IX)    Leu  *   *  Phe  *  Glu Pro Ala Ala  *  Leu Gly
α3(IX)    Thr  *   *  Ile  *  Asp  ?   ?  Ala  *   ?  Gly

          726
pMG377    ATT CCT TAC AAT GGC CAA GGA TTC CCA GAA CCA TAC
α1(XII)   Ile-Pro-Tyr-Asn-Gly-Gln-Gly-Phe-Pro-Glu-Pro-Tyr-
α1(IX)    Gln Ala Gly Gln Arg Ala Ala Gly Lys Asn Met Lys
α2(IX)    Ala Leu Pro Thr Pro Arg His Gly

          738
pMG377    GTT CCT GAA AGC GGA CCC TAT GCA GGC TGA AGG TGA
α1(XII)   Val-Pro-Glu-Ser-Gly-Pro-Tyr-Ala-Gly
α1(IX)    Gly  *
```

FIG. 3. Nucleotide sequence and conceptual translation product [designated α1(XII)] of pMG377. For comparison, corresponding regions of α1(IX), α2(IX) (9, 10), and α3(IX) collagen chains (courtesy of Michel van der Rest) are shown. Question marks indicate amino acid residues that could not be positively identified in the α3(IX) chain. Asterisks indicate amino acid residues identical to those of the pMG377 translation product. Imperfections in the triple helix are underlined. The boxed region contains the NC1 domains of each chain. The number above the first amino acid in each row corresponds to the residue's position in the α1(IX) chain (9).

Since these three features are characteristic of type IX collagen polypeptides, the sequence of pMG377 and its conceptual translation product [designated α1(XII) in Fig. 3] was compared to those of type IX chains (Fig. 3). This comparison showed that pMG377 was homologous to the type IX collagen chains at both the amino acid and nucleotide sequence levels. For example, the α1(IX) and α1(XII) polypeptides share 47 of 110 amino acid residues, or 42.7%. The homology between α2(IX) and α1(XII) is 38.5%, while α1(IX) and α2(IX) chains are 55.8% homologous in the compared regions. In addition to showing that the pMG377 conceptual translation product is not the α3(IX) chain, amino acid sequence analysis of a proteolytic fragment of the α3(IX) chain (Michel van der Rest, personal communication) showed that α3(IX) and the pMG377 translation product have the lowest homology when all four chains are compared in corresponding regions. Nucleotide homologies between pMG377 and α1(IX) and α2(IX) cDNAs are somewhat higher than the amino acid homologies. pMG377 is identical with the α1(IX) cDNA, pYN1738, in 145 positions of 312 nucleotides, corresponding to 46.5% homology. The homology with the α2(IX) cDNA, pYN1731, is 44.2%. In the same region, the homology between the type IX cDNAs is 55.4%. Nucleotide sequences are not yet available for the α3(IX) chain.

Because of the similarity to type IX collagen, the tissue distribution of the pMG377 mRNA was studied. Blot-hybridization analysis of embryonic tendon, calvarial, sternal, and corneal mRNAs with the nick-translated pMG377 insert did not show hybridization to any distinct mRNA species, even after long exposures to x-ray film. This suggests that the mRNA corresponding to pMG377 is not very abundant. Because of their higher sensitivity, RNase protection assays were used to determine the tissue distribution of the pMG377 mRNA. *Ava* I fragments of the pMG377 insert DNA were cloned into the SP6 vectors to synthesize a radiolabeled cRNA probe of high specific activity (Fig. 4 *Upper*). The cRNA probes were hybridized to 17-day-old chicken embryo calvarial, sternal, and tendon mRNA as well as to 6-day-old corneal mRNA. The pMG377 cRNA probes are protected by a completely complementary mRNA species present in calvaria, cornea, tendon, and sterna as determined by the size of the protected cRNAs (Fig. 4 *Lower*). Since this rules out cross-hybridization to type IX mRNAs, we have desig-

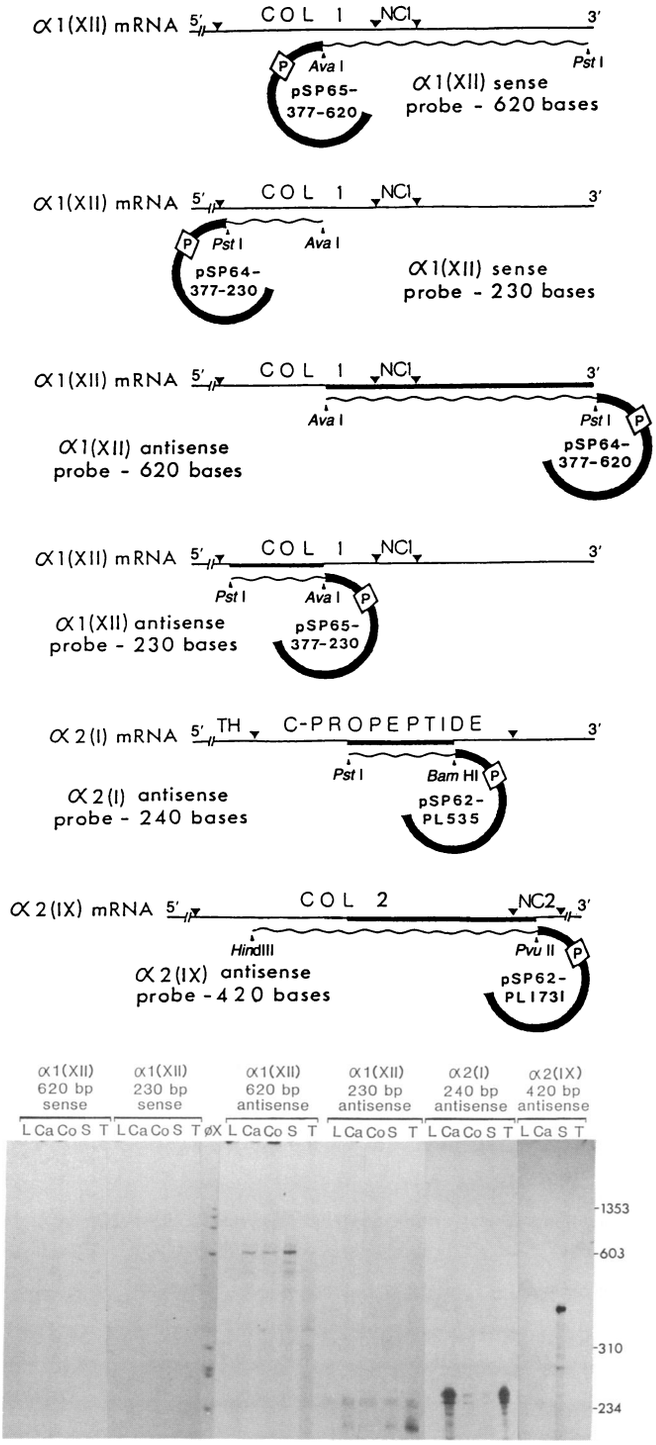


FIG. 4. (*Upper*) Schematic diagrams of sense and antisense cRNA probes used for RNase protection assays. The probes were generated by cloning portions of the plasmids pMG377, pYN535 [specific for α2(I) collagen], and pYN1731 [specific for α2(IX) collagen] into pSP6 vectors. Only the *Ava* I and *Bam*HI sites indicated in the diagrams are present in the cDNA inserts; the *Pst* I sites define the junction between the inserts and their original plasmid vectors. Domains encoding COL1 and NC1 are indicated above the α1(XII) mRNA. Domains encoding COL2 and NC2 are indicated above the α2(IX) mRNA. Triple-helical (TH) and C-propeptide coding domains are shown above the α2(1) mRNA. The SP6 plasmids are shown below the mRNAs. The promoter of the SP6 plasmid is indicated by a P inside a diamond. The wavy line in the linearized plasmids corresponds to transcribed cRNAs. The heavy line in the mRNAs indicates regions of complementarity with the cRNA probes. Note that while the transcribed insert of the α2(IX) plasmid is 680 nucleotides long, only 420 nucleotides are complementary to α2(IX) mRNA; the remaining 260 nucleotides are derived from simian virus 40 DNA and correspond to the "linker" used for cloning the α2(IX) cDNA pYN1731 (10). (*Lower*) Autoradiograph of probes protected by 17-day-old chicken embryonic calvarial, sternal, tendon, and 6-day corneal total RNA. A 10^6-cpm cRNA probe was used for each hybridization. Probes are indicated above the six sets of lanes. Fifty micrograms of the following chicken embryonic total RNA was used: 17-day liver (lanes L), 17-day calvaria (lanes Ca), 6-day cornea [not used with α2(IX) probe] (lanes Co), 17-day sterna (lanes S), and 17-day tendon (lanes T). The time of exposure to film for the α2(I) and α2(IX) cRNA lanes is 0.1 that of the exposure for the lanes of the other four cRNAs; however, even when the film was exposed 10 times longer, no bands were seen with the α2(IX) probe in the calvarial and tendon RNA lanes. Lane φX is as in Fig. 2. The lengths (in base pairs) of size markers are indicated on the right.

nated the protecting mRNA as α1(XII) mRNA in Fig. 4. By comparing the intensity of the bands seen with the pMG377 and α2(I) cRNA probes, we estimate that the α2(I) mRNA is 100–1000 times more abundant in tendon and calvaria than the pMG377 mRNA.

The presence of type IX collagen mRNA in sternal cartilage has previously been demonstrated by RNA blot analysis (9). With this technique, calvarial RNA did not appear to contain type IX collagen mRNA (9). To make a meaningful comparison between the tissue distributions of type XII and type IX mRNAs, we determined levels of type IX mRNA with RNase protection assays using cRNA probes also. As shown in Fig. 4, an α2(IX) cRNA probe was protected by sternal cartilage RNA but not by RNA from liver, calvaria, or tendons. Therefore, we conclude that type IX and XII collagen mRNAs are indeed expressed differently in different tissues. The detection of α1(XII) mRNA in embryonic tendons, calvaria, corneae, and sterna shows that type XII collagen mRNA is expressed in these tissues. However, since all of these tissues contain more than one cell type, we cannot make conclusions about the cell-specific expression of the mRNA. Obviously, *in situ* hybridization with type XII collagen probes will be needed to address questions about which cells express the mRNA.

To further explore the similarity between the collagen encoded by pMG377 and type IX collagen, we used the cDNA to screen a genomic library. This led to the isolation of a genomic clone, DG12, containing a 13.7-kilobase (kb) genomic fragment. Partial nucleotide sequence analysis of a 3.1-kb subclone called pDG30 showed that DG12 represents a portion of the gene that codes for the α1(XII) mRNA (Fig. 5). The sequence analysis also showed that the position of the intron between exons 1 and 2 (as counted from the 3′ end) of the α1(IX) and α2(IX) genes is conserved in DG12 (Fig. 5). However, within the region that codes for the NC1 domain, DG12 contains an intron that is not found in the type IX collagen genes, suggesting that exon 1 in the type IX genes is represented by at least two exons in the homologous DG12 collagen gene.

What is the identity of the collagen encoded by the cDNA pMG377? Based on the homology between the pMG377 translation product and type IX collagen chains, we suggest that pMG377 codes for one of the polypeptide subunits of a molecule that shares some of the characteristics of type IX collagen. Whether this molecule would contain three distinct polypeptide subunits, of which the pMG377-encoded polypeptide is one, is as yet unknown. Also, until the protein is identified and isolated, it cannot be known whether it contains, like type IX collagen, a covalently bound glycosaminoglycan chain. However, based on the homology between the COL1 region of α1(IX) and α2(IX) chains and the pMG377-encoded polypeptide, we believe that each polypeptide of the molecule would contain three triple-helical

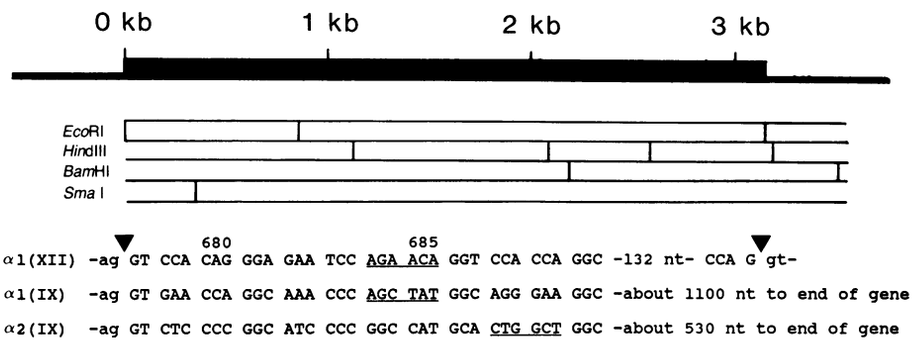


FIG. 5. Diagram showing the restriction endonuclease cleavage map of the subclone pDG30 derived from the α1(XII) genomic clone DG12. A portion of the nucleotide sequence of pDG30 is shown below the diagram. This sequence defines a 171-nucleotide (nt) exon that is homologous to the 5′ region of exon 1 in the α1(IX) and α2(IX) collagen genes (2). Imperfections in the Gly-Xaa-Yaa triplet structure are underlined. Exon/intron junctions are indicated by arrowheads. The numbers above the sequences are the amino acid codon numbers previously published for the α1(IX) cDNA pYN1738 (9).

domains connected by noncollagenous sequences as in type IX collagen. This structure, in addition to the unique sequence of the pMG377 translation product, provides a compelling argument for the conclusion that pMG377 codes for a subunit of a type IX-like but distinct collagen type. We suggest that this collagen be designated type XII and that the pMG377 translation product represents the carboxyl-terminal part of α1(XII).

Although pMG377 has been isolated from a cDNA library constructed with size-selected mRNA that translates into type IX-sized collagenase-sensitive polypeptides (Fig. 1), we have no direct evidence that those polypeptides are due to translation of α1(XII) mRNA. However, if α1(XII) collagen chains are similar to α1(IX) chains in size and our model of type XII collagen is correct, pepsin extracts of tissues containing this collagen should contain triple-helical fragments similar in size to the type IX fragments obtained when cartilage is extracted with pepsin. One of these fragments should contain the COL1 domain of the molecule. This pepsin-resistant fragment should be disulfide-bonded and give rise to three 10-kDa polypeptides after reduction. Prompted by this prediction, B. Dublet and M. van der Rest (private communication) recently examined pepsin extracts of embryonic chicken tendons for potential type XII collagen fragments. They indeed found a collagenous fragment with the predicted characteristics. A similar fragment has been isolated from bovine periodontal ligament by Yamauchi *et al.* (27). Although the reported cysteine content is much higher in the bovine preparation than in the chicken embryo peptide, it is possible, given the species difference, that the two fragments are homologous and that they represent a portion of type XII collagen. In fact, amino acid sequencing of a tryptic peptide, 13-amino acid residues long, derived from the 10-kDa pepsin-resistant fragment isolated by Dublet and van der Rest, shows complete identity with the amino acid sequence predicted from the nucleotide sequence of pMG377 (unpublished observations). Therefore, chicken embryonic tendons not only contain α1(XII) mRNA as demonstrated by cloning of cDNA and by the results of the RNase protection assay as reported here but also contain the corresponding translation product as part of a collagenous molecule with triple-helical domains.

We thank Michel van der Rest for permission to include the partial α3(IX) sequence and Kathy Svoboda for the 6-day-old chicken cornea RNA. We also thank Eunduck Kay and Christine Towle for help during the initial phase of this work. This work was supported in part by research grants AR36819 and AR36820 from the National Institutes of Health.

1. Miller, E. J. (1985) *Ann. N. Y. Acad. Sci.* **460,** 1–13.
2. Lozano, G., Ninomiya, Y., Thompson, H. & Olsen, B. R. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4050–4054.
3. Ninomiya, Y., Gordon, M., van der Rest, M., Schmid, T., Linsenmayer, T. & Olsen, B. R. (1986) *J. Biol. Chem.* **261,** 5041–5050.
4. van der Rest, M., Mayne, R., Ninomiya, Y., Seidah, N. G., Chretien, M. & Olsen, B. R. (1985) *J. Biol. Chem.* **260,** 220–225.
5. Noro, A., Kimata, K., Oike, Y., Shinomura, T., Maeda, N., Yano, S., Takahashi, N. & Suzuki, S. (1983) *J. Biol. Chem.* **258,** 9323–9331.
6. Vaughan, L., Winterhalter, K. H. & Bruckner, P. (1985) *J. Biol. Chem.* **260,** 4758–4763.
7. Bruckner, P., Vaughan, L. & Winterhalter, K. H. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2608–2612.
8. Konomi, H., Seyer, J. M., Ninomiya, Y. & Olsen, B. R. (1986) *J. Biol. Chem.* **261,** 6742–6746.
9. Ninomiya, Y. & Olsen, B. R. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3014–3018.
10. Ninomiya, Y., van der Rest, M., Mayne, R., Lozano, G. & Olsen, B. R. (1985) *Biochemistry* **24,** 4223–4229.
11. Gordon, M. K. (1986) Dissertation (Rutgers-The State University, New Brunswick, NJ).
12. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
13. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
14. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
15. Cantatore, P. & Attardi, G. (1980) *Nucleic Acids Res.* **8,** 2605–2625.
16. King, J. & Laemmli, U. K. (1971) *J. Mol. Biol.* **62,** 465–477.
17. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25,** 263–269.
18. Holmes, D. S. & Quigley, M. (1971) *Anal. Biochem.* **114,** 193–197.
19. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
20. Overbeek, P. A., Merlino, G. T., Peters, N. K., Cohn, V. H., Moore, G. P. & Kleinsmith, L. J. (1981) *Biochim. Biophys. Acta* **656,** 195–205.
21. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 497–559.
22. Butler, E. T. & Chamberlin, M. J. (1982) *J. Biol. Chem.* **257,** 5772–5778.
23. Melton, D. A., Krieg, P. A., Rebagliati, R. R., Maniatis, T., Zinn, K. & Green, M. R. (1984) *Nucleic Acids Res.* **12,** 7035–7056.
24. Baron, M. H. & Maniatis, T. (1986) *Cell* **46,** 591–602.
25. Zinn, K., DiMaio, D. & Maniatis, T. (1983) *Cell* **34,** 865–879.
26. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
27. Yamauchi, M., Kuboki, Y., Sasaki, S. & Mechanic, G. L. (1986) *Biochemistry* **25,** 1997–2002.