One cosmid and two overlapping phage clones covering the entire mouse \( \alpha 2(IX) \) collagen gene including 12 kilobase pairs (kb) of 5' - and 8 kb of 3' -flanking sequences were isolated from two genomic libraries. The overall gene structure was determined by restriction mapping and nucleotide sequencing. The gene spans 16 kb from the start of transcription to the polyadenylation site and contains 32 exons. It codes for a mRNA of 3 kb that translates into a polypeptide of 688 amino acids. The intron-exon junctions and mRNA structure were confirmed by amplification of cDNA made for mouse cartilage RNA. The coding sequence of the mouse \( \alpha 2(IX) \) collagen gene shows marked similarities to those for other type IX collagen chains. Although the overall exon-intron organization of the mouse gene is very similar to the chick \( \alpha 2(IX) \) gene, some unexpected differences were observed at the splice junctions. Split codons characteristic for the central triple helical domain of the chick were not found in the mouse gene that thus exhibited a long stretch of exons with sizes that are multiples of 9 base pairs in this domain. The promoter of the mouse \( \alpha 2(IX) \) collagen gene contains some G + C-rich elements including three Sp1 consensus recognition sites and a far upstream CCAT box but no TATAA box. Both primer extension and RNase protection assays revealed several transcription start sites within 418 base pairs of the promoter. The present study reports the first complete nucleotide sequence of any type IX collagen gene and forms the basis for comparative structural studies on this collagen type and for experiments involving transgenic mice.

In higher eukaryotes the collagen gene family consists of at least 30 genes making up a minimum of 17 different collagen types with important structural functions in the extracellular matrices (1, 2). Four of these collagen types (II, IX, X, and XI) have traditionally been considered specific for cartilage. Type II, IX, and XI collagens form a multicomponent fibrillar network (3) with two important functions. The fibrils provide structural strength to the tissue and entrap the proteoglycan molecules that provide cartilage with resilience. The unique properties of cartilage are thus closely associated with the interactions of its two major structural components. Due to its unique structural features and localization on the surface and at the intersections (5) of cartilage collagen fibrils, type IX collagen is likely to play an important role in this process. The interactions of type IX and type II collagens are stabilized through covalent cross-links (6–8). Furthermore, type IX collagen is a proteoglycan, i.e. many collagen molecules contain a covalently attached chondroitin sulfate side chain (9). Rotary shadowing has demonstrated this chain to be located at a distinct kink that is present in all type IX collagen molecules (10). In cartilage the amino-terminal end of type IX molecule that sticks away from fibril surface has a basic globular domain capable of interacting with the glycosaminoglycan side chains of proteoglycans (11). All these features suggest an important role for type IX collagen in mediating the interactions between the collagen network and the proteoglycans.

Type IX collagen is a heterotrimer of \( \alpha 1(IX), \alpha 2(IX), \) and \( \alpha 3(IX) \) chains and belongs to the subfamily of FACIT (fibril-associated) collagens with interrupted triple helices. Its molecular biologic characterization has focused on the chick system. cDNA clones are available for the \( \alpha 1(IX) \) (10–12), \( \alpha 2(IX) \) (12, 13), and \( \alpha 3(IX) \) (14, 15) collagen mRNAs. The exon structure is also known for the chick \( \alpha 2(IX) \) gene (12, 16, 17) and partially for the chick \( \alpha 1(IX) \) gene (12, 16, 18). The cDNAs and genes for rat and human \( \alpha 1(IX) \) collagen have also been characterized to variable degrees (19, 20), whereas those for mammalian \( \alpha 2(IX) \) and \( \alpha 3(IX) \) collagens remain uncharacterized. The gene structure of the chick type IX collagen shares several features common with genes coding for fibrillar collagens. The characteristic feature of these genes is the structure of exons coding for triple helical domains; these have sizes that are multiples of 9 bp (coding for one Gly-Xaa-Yaa repeat), begin with a codon for Gly, and consequently end with a complete codon for the amino acid in Yaa position (the so-called “9-bp rule”). In fibrillar collagens a majority of the triple helical exons are 54 bp in size; others have sizes of 108, 45, 99, and 162 bp, all conforming to the 9-bp rule. However, the genes for nonfibrillar collagens, including those for chick \( \alpha 1(IX) \) and \( \alpha 2(IX) \) chains, also exhibit some divergence from these rules. Discontinuities in the Gly-Xaa-Yaa repeat structure result in exon sizes that are not multiples of 9 bp, and split codons for Gly sometimes occur at exon junctions.

Our interest has focused on the mammalian \( \alpha 2(IX) \) colla-
gens. We have recently cloned a short cDNA for the mouse a2(IX) mRNA (21) and a full-length cDNA for the human a2(IX) mRNA (22). In the present study the murine probe was employed for isolation of the corresponding gene that was subsequently characterized by nucleotide sequencing. Detailed information of the entire genomic structure of the mouse a2(IX) collagen gene is needed for further experiments that include generation of transgenic mice harboring various mutations of the gene.

**EXPERIMENTAL PROCEDURES**

**Genomic Libraries**—A mouse genomic library (liver DNA from BALB/c mice, Clonetech) in cosmids pWE15 (5 × 10⁶ clones) was screened with a ³²P-labeled 444-bp cDNA fragment of mouse a2(IX) collagen cDNA clone pMCol9a2-1 (21). The hybridization was performed in 5 × SSC (1 × SSC is 0.15 M NaCl, 0.015 M trisodium citrate, pH 7.0), 10 × Denhardt’s solution, 0.5% sodium dodecyl sulfate, and 200 mg/ml denatured herring sperm DNA at 66 °C overnight. A high stringency wash was performed for the filters in 0.2 × SSC, 0.5% sodium dodecyl sulfate at 66 °C for 3 h. After this wash one colony exhibiting a strong positive signal was selected for DNA purification and further characterization by restriction mapping and Southern hybridization. A commercial genomic library prepared with DNA from NIH/3T3 cells in λ FIX² vector (Stratagene) was screened similarly with the same probe. Two additional clones were identified and isolated for further characterization.

**Subcloning and DNA Sequencing**—For sequencing approximately 36 kb of the gene from the cosmid clone was subcloned first as HindIII fragments in Bluescript™ KS+ vector (Stratagene) and then further as 85 smaller subclones. In addition to oligonucleotide primers corresponding to the T3 and T7 recognition sites of the vector, 26 synthetic oligonucleotides were used as sequencing primers. Sequencing was performed on the double-stranded DNA using the Sanger dideoxy method (Sequenase® reagent kit). The sequences were stored and analyzed using the University of Washington GCG software.

**RNA Extraction, cDNA Synthesis, and Amplification by Polymerase Chain Reaction**—Total RNAs were extracted from rib and epiphyseal cartilages of newborn mice using the guanidinium isothiocyanate method (23). Total RNA (1 μg) was used as the template for cDNA synthesis by Moloney murine leukemia virus reverse transcriptase under conditions suggested by the supplier (Life Technologies, Inc.). Both oligo(dT) and random hexamers were used as primers. Aliquots of cDNA were used for amplification by the polymerase chain reaction (Gene Amp®, Perkin-Elmer) using specific oligonucleotide primers. The reactions were cycled by denaturing at 94 °C for 1 min, annealing at 50 °C for 2 min, and extension at 72 °C for 2 min. After 30 amplification cycles aliquots of the reactions were fractionated by electrophoresis on 1.5% agarose gels; the specific fragments were purified and cloned by blunt-end ligation into the EcoRV site of the Bluescript vector.

**Primer Extension and RNase Protection**—The transcription initiation site was determined by primer extension as previously described (24). A specific oligonucleotide MP-16 (5′-CAGAGTCTTCTGGA AGTCCC-3′) labeled at the 5′-end and 20 μg of total RNA isolated from mouse cartilage were mixed. The annealed oligonucleotide served as the primer for the reverse transcriptase reaction. The primer-extended products were analyzed on a 6% polyacrylamide sequencing gel in parallel with dideoxy sequencing reactions primed with the same oligonucleotide. For RNase protection analyses transcripts were synthesized for several genomic subclones spanning exon 1 from the T3 and T7 RNA polymerase sites in the Bluescript vector (24). The protected fragments were fractionated on 4–6% denaturing polyacrylamide gels.

**RESULTS AND DISCUSSION**

Screening of the mouse genomic (BALB/c) cosmid library with a mouse a2(IX) collagen cDNA fragment from plasmid pMCol9a2-1 was performed under high stringency and yielded one positive clone cMMP9a2. Two additional clones (ARR9a2-1 and λRR9a2-2) were found during screening of a mouse genomic (NIH/3T3) phage library. Restriction mapping of these three clones showed considerable overlap indicating that they all coded for the same gene (Fig. 1) that was consequently identified as the a2(IX) collagen gene. Clone cMMP9a2 contains the entire coding sequence of the gene and approximately 12 kb of 5′- and 8 kb of 3′-flanking sequences.

The gene for the a2 chain of mouse type IX collagen spans 16 kb from the transcription start site to the polyadenylation site and contains 32 exons. A total of 21 kb of the nucleotide sequence was determined (Fig. 2). Exons were identified by flanking consensus splice signals and by comparison with the corresponding chick gene (12, 13, 16, 17) and murine cDNA sequences. The intron-exon boundaries and all exon sequences were confirmed by sequencing of cDNA clones spanning all exon boundaries of the mRNA. The amplification strategy of cDNAs by polymerase chain reaction is shown in Fig. 3. The sequences at the intron-exon boundaries conform well with the general splice consensus sequences. The consensus sequence for the 3′- and 5′-ends of the introns was as follows.

The subscript numbers denote the frequency of the most common nucleotides in percent (total 31). Only 3 of the 32 exons begin with a split codon in the mouse (exons 30–32). In the chick gene also exons 20, 22, and 24 begin with a split codon (12). The chick gene is therefore more compact than the mouse gene, which has an average intron size (419 bp) quite similar to the mouse pro-α1(II) collagen gene (451 bp) (25). Like the type II collagen gene, the a2(IX) gene is distinctly more compact toward the 3′-end; the 5′-half of the coding sequence spans 11.4 kb, and the 3′-half spans only 3.5 kb of genomic DNA. The sizes of exons between mouse and chick are remarkably similar except in the 3′-end of the central COL2 domain where an interesting size difference was observed as will be discussed below. The longest intron of the mouse gene (intron 17) contains 3353 bp, and the shortest one (intron 23) contains 77 bp. The sizes
the sequences for the collagenous domains contain gaps at discontinuities of the Gly-Xaa-Yaa repeat pattern to allow for their alignment. All exon sequences have been confirmed by the sequencing of the corresponding cDNA.

FIG. 2. The nucleotide sequence of the collagenous domains.

Mouse α2(Ix) Collagen Gene

The nucleotide sequence of the mouse α2(Ix) collagen gene. The sequence begins with the translation start site and shows all exons (capital letters) with 10 bp of flanking intron sequences and the 3'-untranslated region. The exon numbers and sizes are shown on the left. Underlining highlights the single polyadenylation signal. The sequences coding for the noncollagenous domains are boxed. The sequences for the collagenous domains contain gaps at discontinuities of the Gly-Xaa-Yaa repeat pattern to allow for their alignment. All exon sequences have been confirmed by the sequencing of the corresponding cDNA.
of introns do not share any apparent correlation between mouse and chick (12). Although almost 100 kb in size, the known areas of the chick α1(IX) collagen gene also share marked similarities with the α2(IX) genes in exon organization (12).

**Table I**

| Exon No. | Exon size | Intron size | Exon No. | Exon size | Intron size | Exon No. | Exon size | Intron size |
|----------|-----------|-------------|----------|-----------|-------------|----------|-----------|-------------|
| 1        | 156*      | 1179        | 12       | 54        | 288         | 23       | 54        | 77          |
| 2        | 75        | 1260        | 13       | 54        | 386         | 24       | 72        | 151         |
| 3        | 36        | 86          | 14       | 54        | 91          | 25       | 36        | 136         |
| 4        | 63        | 1217        | 15       | 54        | 105         | 26       | 46        | 93          |
| 5        | 54        | 112         | 16       | 54        | 3353        | 27       | 36        | 130         |
| 6        | 36        | 97          | 17       | 54        | 369         | 28       | 147       | 234         |
| 7        | 24        | 259         | 18       | 54        | 225         | 29       | 55        | 275         |
| 8        | 54        | 326         | 19       | 54        | 804         | 30       | 189       | 294         |
| 9        | 54        | 194         | 20       | 46        | 392         | 31       | 78        | 265         |
| 10       | 48        | 128         | 21       | 54        | 283         | 32       | 706*       |             |
| 11       | 57        | 84          | 22       | 54        | 205         |          |           |             |

* The exon and intron sizes are given in base pairs.

Fig. 3. The restriction map and amplification strategy for mouse α2(IX) collagen cDNA. Top, cDNA fragments amplified by polymerase chain reaction. For sequencing each fragment was cloned into Bluescript by blunt-end ligation. Middle, the domain structure of the mouse α2(IX) collagen and the restriction map of the overlapping cDNA clones. Bottom, the scale in base pairs.

Fig. 4. A schematic presentation of the exon-intron organization of the mouse α2(IX) collagen gene. Top, the exons shown as boxes are numbered from the 5'-end. The open boxes in exons 1 and 32 denote noncoding regions. Bottom, the bar shows the scale in base pairs.

Mouse α2(IX) Collagen Gene

Sequence comparison of the mouse α2(IX) exons (cDNAs) and deduced amino acids (Fig. 5) with the corresponding human and chick sequences are summarized in Table II. The overall sequence identity between the mouse and chick exons is 70% and between mouse and human exons 87%. The differences occur most frequently in third positions, and the overall amino acid similarities of the α2(IX) chains are 76 and 91%, respectively. The conservation of nucleotide and amino acid sequences varies, however, between the different domains of the chain (Table II). The α2(IX) collagen chain contains seven domains analogous to other known α chains of type IX collagen. The specific features of the mouse α2(IX) collagen gene and its polypeptide product will be discussed below, starting with the promoter and proceeding in the 3' direction through the three collagenous (COL) domains and four short noncollagenous (NC) domains of the α chain. In the present study we follow the customary numbering of the COL and NC domains from the carboxyl-terminal end. Exons and introns, however, are numbered from the 5'-end.

Promoter and 5'-Untranslated Sequence—We initially determined the transcription start site by primer extension analysis using newborn mouse cartilage RNA. Two start sites were found 383 and 502 bp upstream of the ATG codon (Fig. 6A). However, RNase protection analyses revealed several additional transcription start sites downstream. The predominant one is 84 bp upstream of the translation initiation codon (Fig. 6C). Three perfect copies of the hexanucleotide GCCGCG (Sp1 binding site) are located upstream of this site. Within the first 200 bp of this promoter the G + C content is 67%, similar to the mammalian α1(II) collagen promoters (25, 26). The promoter contains a single CCAAT element at -371 and a TATAA sequence at -689. We compared the structure of the promoter with that of the mouse α1(II) collagen since common regulatory motifs could be expected to be found in these genes that are coexpressed. In addition to the three Sp1 sites, two sequences of 26 and 28 bp with sequence identities of 77 and 71% were seen in similar positions relative to the Sp1 sites in the two promoters. We therefore propose that transcription of the mouse α2(IX) gene predominantly starts at the downstream site marked in Fig. 6B. When genomic subclones covering only sequences upstream of -140 were used as probes in Northern analysis, only faint hybridization to a 2.9-kb α2(IX) collagen mRNA was seen in limb cartilage RNA, whereas in samples containing elastic cartilage of the ear lobes a strong band of similar size was seen (data not shown). This supports our data on the major transcription start site but also suggests alternative use of start sites in different cartilages.

The 5'-untranslated sequence of the mouse α2(IX) collagen mRNA thus contains 84-502 nucleotides. This sequence exhibits very little similarity with the corresponding chick sequence that spans at least 250 nucleotides (12, 18). The longest murine 5'-untranslated sequence contains 77 AUG codons before the AUG that begins the open reading frame for the α2(IX) chain. The latter is, however, the first AUG codon in a sequence context that is in good agreement with the consensus for translation initiation GCC(A/G)CCAUUG (27). In the shortest transcript the first AUG begins the open reading frame.

The NC4 Domain—The α2(IX) polypeptide chain begins with a presumptive signal peptide of 21 amino acids, with the cleavage site between Ala and Glu residues conserved between all other type IX collagen chains (12, 14, 15, 20). Within the signal peptide the mouse and chick sequences diverge considerably (Table I, Fig. 5). The size of the NC4 domain in the α2(IX) chain is only 3 amino acids, similar to the α3 chain.
Due to the two alternative promoters the sizes of the NC4 domains in the a1(IX) chain are 243–245 amino acids and 2 amino acids, respectively, excluding the signal peptide of 23 amino acids (12, 20).

The first intron of the mouse a2(IX) collagen gene spans 1179 bp. Work on the pro-a1(II) gene in several species has located highly conserved sequences (25, 26) and a tissue-specific enhancer in the first intron (28). Comparison of the first intron of the mouse a2(IX) gene with the human and mouse pro-a1(II) genes revealed several short segments of sequence conservation including a region containing the putative cartilage-specific enhancer of the type II collagen gene (29).

The COL3 Domain—The length of this domain is 137 amino acids in all the three chains of type IX collagen in all species studied. The domain is coded for by exons 2–10 as in the chick a2(IX) gene (12). Exons 2 and 10 are joining exons coding for noncollagenous domains and for eight and two Gly-Xaa-Yaa repeats, respectively. The sizes of triple helical exons, 63, 54, 36, and 24 bp, conform to the 9-bp rule except for the 24-bp exon, where a discontinuity in the Gly-Xaa-Yaa repeat structure is reflected in the disappearance of one codon. This changes the size of the 27-bp exon 7 to 24 bp. All exons, 63, 54, 36, and 24 bp, conform to the 9-bp rule except for the 24-bp exon, where a discontinuity in the Gly-Xaa-Yaa repeat structure is reflected in the disappearance of one codon. The arrow marks the possible site for proteolytic cleavage.
chick α1(IX) collagen gene within this domain are identical in size except for exon 10, which is only 33 bp due to the shorter noncollagenous domain.

The NC3 Domain—This domain consists of 17 amino acids and is coded for by exons 10 and 11. Both exons are fusion exons, i.e., they also code for two and four perfect Gly-Xaa-Yaa repeats of triple helical sequences of COL4 and COL3 domains, respectively. The triple helical domains of exons 10 and 11 thus make another 64-bp exon that contains in the middle sequences coding for the NC3 domain and one intron. In the chick the NC3 domain contains the sequence Gly-Ser-Ala-Asn, where the Ser residue has been shown to be the attachment site for the glycosaminoglycan side chain (17). This sequence and the entire NC3 domain are highly conserved in the mouse (Fig. 2, Table I) and human (22). In the chick, rat, and human α1(IX) chain the NC3 domain consists of only 12 amino acids and in the α3(IX) chain of 15 amino acids (14, 15). The larger size of the NC3 domain in the α2 chain is necessary to accommodate the GAG attachment site and is probably the cause for the kink observed in all known type IX collagen molecules (4, 30). In all species the three α chains of type IX collagen share a similar amino acid sequence of Cys-Pro-Xaa-Xaa-Cys-Pro-Xaa at the end of this domain that makes it possible for the Cys residues to form interchain disulfide bridges (8, 14, 31).

The COL2 Domain—This central collagenous domain consists of 339 amino acids in an uninterrupted (Gly-Xaa-Yaa)13 configuration. The length of the domain in known α1(IX) and α3(IX) chains is also 339 amino acids; however, in the chick the α3 chain contains one interruption in the triple helical structure (14, 15). In the α2(IX) gene the COL2 domain is

| Domain | Mouse/chick | Mouse/human |
|--------|-------------|-------------|
| NC4   | 40          | 51          |
| COL3  | 80          | 70          |
| COL2  | 75          | 70          |
| NC2   | 87          | 82          |
| COL1  | 75          | 70          |
| NC1   | 8           | 77          |

* The data for the chick sequences are from Refs. 12 and 13.
* aa, amino acid sequence.
* nt, nucleotide sequence.
* The data for the human sequences are from Ref. 22 and (for NC1) from our unpublished data.
* Similarity in percent.
* NA, not available.
* Due to an apparent sequencing error of the chick gene, no comparison of amino acid sequence was performed.

**TABLE II**
Comparison of amino acid and nucleotide sequence similarities between mouse, human, and chick α2(IX) collagen domains and their coding sequences

**A**

**B**

**C**

**FIG. 6. Determination of the transcription start site.** A, for primer extension of α2(IX) collagen mRNA, a 21-mer oligonucleotide complementary to nucleotides -197 to -217 was synthesized and used to prime cDNA synthesis by reverse transcriptase. The reaction products were resolved on a denaturing sequencing gel (lane X) with the corresponding sequencing reaction run on the left (lanes C, T, A, and G). The arrows and symbols highlight the two extension products. B, nucleotide sequence of the promoter and 5'-untranslated region of exon 1. +1 marks the major transcription start site. Underlining marks from the 5'-end a TATAA sequence, a CCAAT sequence, the two regions of homology between the α2(IX) α1(III) collagen promoters (dotted underlining), the three Sp1 recognition sites (double underlining), and the translation initiation codon. The seven upstream ATG codons and the shortest 5'-untranslated sequence are shown in capital letters. C, RNase protection analysis of the α2(IX) transcripts using a probe made from a 1047-bp EcoRI-Apal genomic subclone. The reaction products were resolved on a denaturing 4.5% acrylamide gel (lane 1) with MspI-digested pBR322 standards (lane 2). The symbols in panels A and C are also shown in panel B above the nucleotides that correspond to the 5'-ends of the transcripts.
collagen. Therefore two mutations must have occurred si-
the chick but not in the mouse is very unexpected. Mainte-
repeats and two copies of the murine B1 repetitive sequence
building blocks of collagen genes (32). Two other exons are
54 bp, one is 72 bp, and one is 36 bp long; all these conform
to the 9-bp rule. The last two triple helical exons are 33 and
147 bp and contain the intron in an unusual location between
the Xaa and Yaa codons. Within this domain is located the
3353-bp intron 17, the largest intron in the mouse a2(IX)
collagen gene. This intron contains both short dinucleotide
repeats and two copies of the murine B1 repetitive sequence
(13) in antiparallel orientation.

The presence of split codons for glycine in exons 19–24 in
the chick but not in the mouse is very unexpected. Mainte-
nance of the correct reading frame is a prerequisite for the
glycerine-Xaa-Yaa repeat structure and crucially important for all
collagens. Therefore these mutations must have occurred si-
multaneously, one to remove one G at the 5'-end of an exon
and another to add a G at the 3'-end of the preceding exon.
Even more surprising is the fact that this event appears to
have occurred at three consecutive exon-intron-exon units.

The NC2 Domain—This domain of 30 amino acids is coded
for by exons 29 and 30, both of which are fusion exons coding
for 1 and 16 Gly-Xaa-Yaa repeats, respectively. The size of the
domain is 30 amino acids also in the known al(1X) chains,
whereas the a3(IX) chain consists of 31 amino acids. Unlike
the a1(IX) and a3(IX) chains, the a2(IX) chain contains no
Cys residues in the NC2 domain; thus this chain cannot
participate in interchain disulfide bridging in this domain
that only occurs between the al(1X) and a3(IX) chains (14,
31).

The COL1 Domain—This domain of 115 amino acids is coded
for by exons 30–32. In this domain the divergence from the
54-bp exon structure and from the 9-bp rule is evident
both in the mouse and in the chick (12, 16). The exons contain
16, 9, and 14 Gly-Xaa-Yaa repeats, respectively. Both exons
31 and 32 contain one interruption, the former of the Gly-
Xaa/Yaa and the latter of the Gly-Xaa-Yaa/Xaa-Yaa type.
In addition, split codons for Gly exist at the 5'-end of exons
31 and 32, the first G nucleotide being located at the end of
the preceding exon. Similar split codons also exist in the chick
both in the a2(IX) and a1(IX) collagen genes (16).

NC1 Domain—In the mouse this domain consists of 25
amino acids. Previously the chick NC1 domain has been
reported to consist of 15 amino acids (16). Comparison of the
chick, mouse, and human nucleotide sequences in the NC1
and 3'-untranslated domains revealed conservation of the
translated amino acid sequence beyond the reported chick
stop codon if an additional nucleotide was inserted 22 bases
before the stop codon. This addition, however, does not result
in a stop codon in the location where it is in the mouse
transcript but warrants further characterization of this do-
mun. The length of the corresponding domain in the chick
a2(IX) chain is 17 amino acids (14, 15), in the chick and rat
a1(IX) chains 21 and 20 amino acids, respectively, and in the
human a1(IX) chain 30 amino acids (19). Again, if an addi-
tional nucleotide is added to the human a1(IX) sequence,
the NC1 domain also becomes 20 amino acids in size. The
cysteine residue at the end of the COL1 domain and another one at
position 5 of the NC1 domain are conserved between all three
a chains of type IX collagen in all species known, making this
the second location where a set of interchain disulfide bridging
occurs between all three (31). The size of exon 31 is also
conserved between mouse and chick a1(IX) and a2(IX)
genesis.

3'-Untranslated Domain—The single polyadenylation site
(AATAAA) detected in the mouse gene begins 498 bp beyond
the translation stop codon (Fig. 2). In Northern analysis a
single a2(IX) collagen mRNA band of approximately 3 kb is
consistently observed (21). The size of the major a2(IX)
collagen transcripts (assuming a poly(A) tail of 200 bases)
can be calculated to be approximately 2850–2900 bases cor-
responding well with our size estimate (2.9 kb) in Northern
analysis (21). Since the gene contains no other AATAAA or
related sequence within the 1 kb of 3'-flanking sequences,
only one polyadenylation site is apparently used for these
transcripts. In the chick gene two polyadenylation sites have
been observed 166 bp and approximately 330 bp downstream
of the translation stop codon (16). Interestingly, essentially
no conservation is detectable within the 3'-untranslated
sequence between chick and mouse gene, a situation quite
different from that in the type II collagen genes where more
than 70% sequence identity exists between the two species
(25). The 3'-untranslated sequences of mouse and human
a2(IX) genes also exhibit considerably less sequence similarity
than the corresponding pro-a1(II) collagen genes.

Whereas the functions of cartilage during development and
in adult organism are fairly well understood, the roles of the
individual constituent molecules are much more difficult to
assess. Type IX collagen is mainly found in the hyaline
cartilage codistributed with type II collagen. In addition to its
participation in fibrillogenesis, type IX collagen appears to
mediate extracellular damage to collagenous type II
collagen. Expression of type IX collagen has been observed at
least in the notochord and the eye (34, 35). Presently it is not
known what the role of type IX collagen is at these sites
during embryonal development.

No mutations in the type IX collagen genes have been
identified in diseases. However, the genes for type IX collagen
are clearly candidate genes in the various chondrodysplasias
and degenerative diseases of joints and spine. The list of
human osteochondrodysplasias currently contains over 175
clinical diagnoses, of which only approximately 40 have been
connected to a specific protein, gene, or locus (36). Recently
a transgenic mouse line expressing a mutated a1(IX) cDNA
construct under the type collagen promoter has been reported
to exhibit osteoarthritic lesions on articular surfaces (37). It
seems obvious that transgenic mice harboring other mutations
in type IX collagen genes will help in defining the role of this
collagen type in cartilage and other tissues. The information
provided in this report forms the basis for such experiments
involving the a2(IX) chain.

Acknowledgments—We are grateful to Merja Lakkisto and Tuula
Oivanen for expert technical assistance.

* M. Peralá and E. Vuorio, unpublished observations.
