

# Extensive Alternative Splicing within the Amino-propeptide Coding Domain of $\alpha 2(\text{XI})$ Procollagen mRNAs

EXPRESSION OF TRANSCRIPTS ENCODING TRUNCATED PRO- $\alpha$  CHAINS\*

(Received for publication, July 13, 1995, and in revised form, March 14, 1996)

Vincent C. H. Lui<sup>‡§</sup>, Ling Jim Ng<sup>‡</sup>, Eric W. Y. Sat, John Nicholls<sup>¶</sup>, and Kathryn S. E. Cheah<sup>||</sup>

From the Departments of Biochemistry and <sup>¶</sup>Pathology, The University of Hong Kong, 5 Sassoon Road, Hong Kong

**Heterogeneity in type XI procollagen structure is extensive because all three  $\alpha(\text{XI})$  collagen genes undergo complex alternative splicing within the amino-propeptide coding domain. Exon 7 of the human and exons 6–8 of the mouse  $\alpha 2(\text{XI})$  collagen genes, encoding part of the amino-propeptide variable region, have recently been shown to be alternatively spliced. We show that exon 6-containing mRNAs for human  $\alpha 2(\text{XI})$  procollagen are expressed at 28 weeks in fetal tendon and cartilage but not at 38–44 days or 11 weeks. In the mouse, exon 6 is expressed in chondrocytes from 13.5 days onward. We recently identified conserved sequences within intron 6 of the human and mouse  $\alpha 2(\text{XI})$  collagen genes, containing additional consensus splice acceptor and donor sites that potentially increase the size of exon 7, dividing it into three parts, designated 7A, 7B, and 7C. We show by reverse transcription polymerase chain reaction and *in situ* hybridization that these potential splice sites are used to yield additional  $\alpha 2(\text{XI})$  procollagen mRNA splice variants that are expressed in fetal tissues. In human, expression of exon 7B-containing transcripts may be developmental stage-specific. Interestingly, inclusion of exon 7A or exon 7B in human and mouse  $\alpha 2(\text{XI})$  procollagen mRNAs, respectively, would result in the insertion of an in-frame termination codon, suggesting that some of the additional splice variants encode a truncated pro- $\alpha 2(\text{XI})$  chain.**

The collagens are an important class of extracellular matrix components because the structural integrity and functional properties of different tissues are influenced by the characteristic combinations and amounts of these molecules (1). These triple helical glycoprotein molecules belong to a large family consisting of 19 different types with 30 genes encoding their constituent  $\alpha$  chains (2, 3). Diversity within a single type of collagen can arise not only through varying the combinations of their  $\alpha$  chains within heterotrimers, but also from the synthesis of different isoforms arising from alternative splicing of

mRNAs. There is increasing evidence for the syntheses of different isoforms within each collagen type arising from alternatively spliced transcripts such as for types II, VI, IX, XI, XIII, XIV, and XVIII collagens (4–7). In addition, alternative splicing of  $\alpha 2(\text{I})$  and  $\alpha 1(\text{III})$  collagen genes produces mRNAs that could encode a non-collagenous protein and/or truncated collagen, suggesting that some collagen genes may have alternative functions (8, 9).

Type XI collagen, a fibril-forming collagen, is a heterotrimer composed of  $\alpha 1(\text{XI})$ ,  $\alpha 2(\text{XI})$ , and  $\alpha 3(\text{XI})$  subunits (10) and is synthesized as a precursor procollagen molecule with amino- and carboxyl-terminal globular extensions (propeptides). The amino-propeptides in both  $\alpha 1(\text{XI})$  and  $\alpha 2(\text{XI})$  procollagen consist of several subdomains, a proline- and arginine-rich peptide region (PARP),<sup>1</sup> a variable region (VR), and a constant region (CR). For both of these procollagens the PARP and CR subdomains are similar in size, but the VR subdomain is highly variable both in length and amino acid sequence, raising questions about possible different functional roles of the VR (11, 12).

We have recently found that heterogeneity in type XI collagen structure may arise from differing combinations of the three  $\alpha(\text{XI})$  chains within heterotrimers and homotrimers (13). The heterogeneity in type XI procollagen structure may be more extensive because all three genes for type XI collagen have been shown to undergo alternative splicing within the amino-propeptide coding domain. The VR in the amino-propeptide of  $\alpha 2(\text{XI})$  procollagen is encoded by exons 5–9 in the gene. Within this coding region, exon 7 of the human and exons 6–8 of the mouse  $\alpha 2(\text{XI})$  collagen genes have been shown to undergo complex alternative splicing (14, 15). Alternative processing has also been demonstrated in the VR of  $\alpha 1(\text{XI})$  procollagen in human, chick, and rat (14, 16). The pro- $\alpha 3(\text{XI})$  chain of type XI procollagen is an overglycosylated variant of pro- $\alpha 1(\text{II})$  and therefore is the gene product of the  $\alpha 1(\text{II})$  collagen gene (17). Exon 2 of the  $\alpha 1(\text{II})$  collagen gene is also alternatively spliced, yielding two forms of  $\alpha 1(\text{II})$  procollagen mRNAs which either include or exclude an exon encoding a cysteine-rich domain in the amino-propeptide (18). Assembly of the translation products of these different pro- $\alpha(\text{XI})$  alternatively spliced transcripts within type XI procollagen heterotrimers would result in extensive heterogeneity in the structure of the amino-propeptide of type XI procollagen.

The genomic organization, intron-exon boundaries, and sequence conservation for the amino-propeptide coding domain of the human and mouse  $\alpha 2(\text{XI})$  collagen genes have been determined recently, partly by comparing cDNA and genomic sequences (14, 15, 19) (Fig. 1). In assessing the degree of inter-species sequence conservation for the  $\alpha 2(\text{XI})$  collagen gene, we

\* This work was supported by the Arthritis and Rheumatism Council (United Kingdom), the Croucher Foundation, and a Hong Kong University Strategic Research grant. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank<sup>TM</sup>/EBI Data Bank with accession number(s) U41067 and U41069.

<sup>‡</sup> The first two authors contributed equally to this work.

<sup>§</sup> Croucher Foundation Scholar and a University of Hong Kong Postdoctoral Fellow. Present address: Dept. of Cell Biology, Erasmus University, Rotterdam, The Netherlands.

<sup>||</sup> To whom correspondence should be addressed. Tel.: 852-281-99170 or 281-99240; Fax: 852-285-51254.

<sup>1</sup> The abbreviations used are: PARP, proline- and arginine-rich peptide; VR, variable region; CR, constant region; PCR, polymerase chain reaction; bp, base pair(s); RT, reverse transcription.

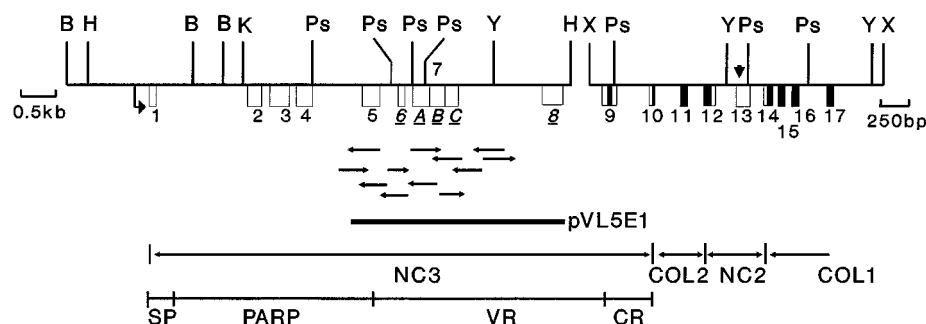


FIG. 1. Exon-intron organization of gene region coding the amino-propeptide of human  $\alpha 2(XI)$  collagen. Exons are shown in boxes and are numbered from the 5' of the *COL11A2* gene. Filled boxes represent exons or part of exons that encode Gly-X-Y sequences. Alternatively spliced exons characterized in this and other studies (14, 15, 19) are in *italics* and *underlined*. Symbols and abbreviations: bent arrow, position of transcription start of the *COL11A2* gene; filled downward arrow, the amino-protease cleavage site; horizontal filled arrows, location of sequences analyzed and direction of sequencing; kb, kilobases; SP, signal peptide; NC3, non-collagenous domain 3; COL2, collagenous domain 2; NC2, non-collagenous domain 2; COL1, central triple helical domain. Restriction enzyme sites: B, *Bam*HI; H, *Hind*III; K, *Kpn*I; Ps, *Pst*I; X, *Xho*I; Y, *Xba*I.

had noted a region of conserved sequences containing motifs corresponding to consensus splice acceptor and donor sites within intron 6 of the human and mouse genes (19). In the present study, we have aimed to assess whether these splice signals can function to generate alternative transcripts using RT-PCR and *in situ* hybridization. We show that these potential splice sites are used to yield additional  $\alpha 2(XI)$  procollagen mRNA splice variants that are expressed in human and mouse fetal tissues. Interestingly, inclusion of one of these additional exons in  $\alpha 2(XI)$  procollagen mRNAs would result in the insertion of an in-frame termination codon within the VR, suggesting that some of the additional splice variants encode a truncated pro- $\alpha 2(XI)$  chain.

#### EXPERIMENTAL PROCEDURES

##### Human Fetal Material

Human fetal tissues were obtained from elective therapeutic terminations of pregnancy. Approval was obtained from the Ethical Committee of the Medical Faculty, Hong Kong University, for work with fresh embryonic materials. The gestational age of specimens was determined by assessing the centers of ossification (20) and the histologic characteristics of the stained sections. The crown-rump length was also measured for intact fresh specimens (21).

##### DNA Sequencing and Sequence Analyses

Nucleotide sequences were determined on both strands by the dideoxy chain termination method (22) on double-stranded DNA using Sequenase 2.0 (U. S. Biochemical Corp.) following the manufacturer's protocol. Longer contiguous stretches of DNA sequence were obtained using the exonuclease/mungbean nuclease method of generating overlapping clones (23). Sequence comparison analyses were performed using the UWGCG computer program (24).

##### Oligonucleotide Primers

The following pairs of oligonucleotide primers were used for cDNA synthesis and PCR. To facilitate cloning of the PCR products, a *Hind*III adaptor (underlined) was added to the 5' ends of primers. The locations of these primers are also shown in Figs. 1 and 2.

**Amplification between Human Exons 5 and 7C**—An exon 7C primer (OVL142H), 5'-GGCAAGAGGCTCGACTCCAGGA-3', was used for cDNA synthesis. OVL142H and an exon 5 primer (OVL130U), 5'-GAGCCCAGAGATCTCCACAGCA-3', were used for first round PCR. Second round PCR was carried out using an exon 5 primer (OVL130R), 5'-AAAAGCTTATCAGGAACCCAGAGCCAG-3', and another exon 7C primer (OVL131), 5'-AAAAGCTTCTCTTCACTGGGGTGGGGT-3'.

**Human Exons 5 to 7A**—An exon 7A primer (OVL202), 5'-GCAGTTGGAGAGGGCTCCGG-3', was used for cDNA synthesis. OVL202 and OVL130U were used for first round PCR. An exon 7A primer (OVL203), 5'-AAAAGCTTGCAGGTGGGGTAGGCTTTCA-3', and OVL130R were used for second round PCR.

**Human Exons 5 to 7B**—An exon 7B primer (OVL204), 5'-GGA-AATCTCAGATCTTCAGCCCC-3', was used for cDNA synthesis. OVL204 and OVL130U were used for first round PCR. Second round PCR primers were OVL130R and OVL205, 5'-AAAAGCTTGGGGA-

GAGTGAACCTCCAAG-3'.

**OVL172 (5'-ACCTTGGAGGTTCACTCTCC-3')**—This was used as an exon 7B primer in conjunction with OVL131 in PCR to generate a human exon 7B probe for *in situ* hybridization.

**Mouse Exons 5 to 7C**—An exon 7C primer (OVL142), 5'-AAAAGCTTGGCAAGGGACTGGACTCCAG-3', was used for cDNA synthesis. OVL142 and an exon 5 primer (OVL140), 5'-AAAAGCTTGTTCAGGAGCTCCTCGTTGT-3', was used for first round PCR. Primers used for the second round PCR amplification were OVL141 (5'-AAAAGCTTTCAGGAGCCCCAGAGCAGGT-3', exon 5 primer) and OVL131. OVL131 was used as a common exon 7C primer for the mouse and human genes because this sequence was 100% conserved.

##### Cloning of RT-PCR Products

First strand cDNA was transcribed from 20  $\mu$ g of total human RNA or 10  $\mu$ g of total mouse RNA using avian myeloblastosis virus reverse transcriptase (35 units, Seikagaku America Inc. or 200 units, Superscript II, Life Technologies, Inc.) with a corresponding antisense primer as described previously (25). One-tenth of the reaction product was used for second strand DNA synthesis and amplification by PCR with the corresponding 5' sense primer. PCRs were carried out with 1-min denaturation of DNA at 94 °C, 1-min annealing reaction at 55 °C, 1-min extension reaction at 72 °C for 30 cycles. As controls, PCRs were performed with same set of RNAs but without the addition of reverse transcriptase to test for genomic contamination of RNA samples. Second round PCR amplifications were performed with one-fifteenth of the first round amplified products. The PCR amplifications were carried out under the same conditions as the first round PCR amplification reactions. RT-PCR products were either directly cloned into the *Eco*RV site of pUBS (gift of G. Murphy, AFRC, Norwich) or digested with *Hind*III (New England Biolabs) and subcloned into the *Hind*III site of pBlue-scribe (Stratagene).

##### Southern Hybridization Analyses

Southern hybridizations using DNA probes labeled with [ $\alpha$ - $^{32}$ P]dCTP (Amersham Corp., 3,000 Ci/mmol) were at 65 °C for 18 h in a mixture containing 3  $\times$  SSC, 0.1% SDS, 12% polyethylene glycol 8000 (Sigma), 250  $\mu$ g/ml heparin (Sigma grade II) (1  $\times$  SSC: 0.15 M NaCl, 0.15 M sodium citrate, pH 7.5), with posthybridization washes in 1  $\times$  SSC, 0.1% SDS at 65 °C for human RT-PCR samples and in 0.1  $\times$  SSC, 0.1% SDS at 65 °C for mouse samples.

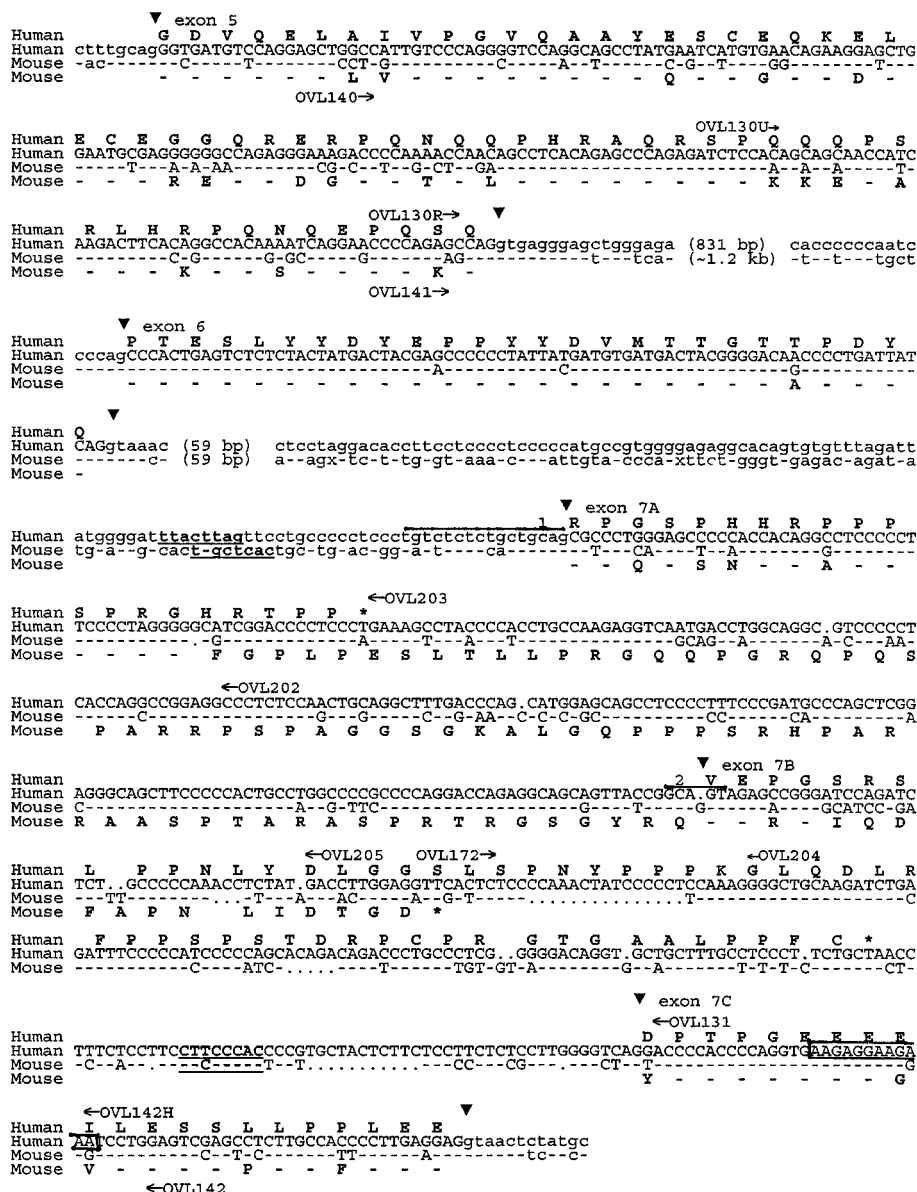
##### RNA Isolation and *in Situ* Hybridization

Total human and mouse RNAs were prepared by the lithium chloride-urea differential precipitation method (26). RNAs were isolated from week 11 human fetal cartilage and from week 28 human fetal cartilage and tendon. Mouse RNAs were isolated from 1-day-old neonatal cartilage and from day 15.5 whole mouse fetuses. Two human fetuses of 38–42 days and 42–44 days gestation and day 16.5 CBA/n mouse fetuses were collected and processed for *in situ* hybridization as described (27). Single-stranded  $^{35}$ S-labeled sense and antisense riboprobes were generated from subclones containing human and mouse pro- $\alpha 2(XI)$  collagen gene exons as described previously (25).

##### Riboprobes

Riboprobes for the human  $\alpha 2(XI)$  collagen gene (*COL11A2*) were as follows.

FIG. 2. Sequence conservation and additional splice signals between exons 5 and 7 in the amino-propeptide coding domains of the human and mouse  $\alpha 2(XI)$  collagen genes. Sequence identity is indicated by the dashes, and gaps introduced by the GAP program to optimize alignment are represented by stops (...). Exon sequences are in uppercase letters; intron sequences are in lowercase letters. The translated amino acid sequence is shown above and below the DNA sequence and in bold uppercase letters. Additional potential splice consensus sequences are overlined and numbered for exon 7A (site 1) and exon 7B (site 2). Donor sequences similar to the consensus splice site signals: *site 1* is located 450 (human) and 476 (mouse) bp upstream of the published splice acceptor site for exon 7; *site 2* is a potential donor site in human and 5' 227 and 198 bp 5' of the exon 7 splice acceptor site in human and mouse, respectively. The putative lariet sequence is underlined and in bold. The stop codon of translation is indicated by a star. The locations of oligonucleotide primers that were used in the RT-PCR analyses are indicated by horizontal arrows above and below the DNA sequence. x, sequence ambiguity not resolved even after sequencing both strands of DNA. The potential splicing regulating purine-rich sequences (see "Discussion") are boxed.



1. pVL50, a 0.6-kilobase *Pst*I fragment containing *COL11A2* exon 6, was subcloned into pBluescript KS<sup>+</sup> (Stratagene).
2. pVL51, a 170-bp *Pst*I fragment containing *COL11A2* exon 7A, was subcloned into pBluescript KS<sup>+</sup>.
3. pVL105, a probe for *COL11A2* exon 7B, was a 215-bp PCR fragment generated using primer pair OVL172 and OVL131, which anneal to exons 7B and the 5' end of exon 7C of *COL11A2*, respectively, and cloned into TA vector (Invitrogen).

Riboprobes for the mouse  $\alpha 2(XI)$  collagen gene (*Col11a-2*) were as follows.

1. pVL100, a 130-bp RT-PCR fragment containing *Col11a-2* exon 6, was subcloned into pBluescribe.
2. pVL102, a 290-bp RT-PCR fragment containing *Col11a-2* exon 7A, was subcloned into pBluescribe.
3. pVL103, a 490-bp RT-PCR fragment containing exons 7A and 7B subcloned in pBluescribe, was used to generate a probe for *Col11a-2* exon 7B. pVL103 contained a *Bgl*II site within exon 7B, 17 nucleotides from the junction between exons 7A and 7B. To generate a riboprobe specific for exon 7B, pVL103 was linearized at the *Bgl*II site for the production of antisense transcripts.

*In situ* hybridization slides were stained in Harris hematoxylin and eosin, and K5 photographic emulsion (Ilford) was used for autoradiography. Photographs of *in situ* hybridizations were taken using Kodak Ektachrome ASA 64 film on a Zeiss Axioskop microscope under dark-field illumination.

## RESULTS AND DISCUSSION

**Additional Splice Sites Flanking Exon 7 within the Variable Region Coding Domain**—A comparison of the genomic sequences of the human and mouse  $\alpha 2(XI)$  collagen genes<sup>2</sup> (19) showed that there was little sequence conservation for the introns between exons 5 and 7. An exception was a highly conserved region in the intron sequence between exons 6 and 7, immediately upstream of the splice junction of exon 7 (Fig. 2 and Table I). Interestingly, this conserved region contained additional potential splice acceptor and donor sequences similar to the consensus splice site signals 5'-T/CT/CTTT/CT/CT/CT/CT/CNCAG|G-3' (acceptor) and 5'-AG|GTA/GAGT-3' (donor) (28). One site (site 1, 5'-TCTCTGCTGCAG|CG-3' in human, 5'-TCTCTCACTGCAG|CG-3' in mouse) is located 450 and 476 bp upstream of the published splice acceptor site for exon 7. The other site (site 2) is a potential donor site (5'-TTACCGGCA|GTAGAG-3' in human and 5'-TCGGCAG|GTAGAG-3' in mouse) 227 and 198 bp 5' of the exon 7 splice acceptor site in human and mouse, respectively (Fig. 2). Two

<sup>2</sup> V. C. H. Lui, L. J. Ng, E. W. Y. Sat, J. Nicholls, and K. S. E. Cheah, unpublished data.



TABLE I  
Comparison of exon sizes and sequence identity for the  
amino-propeptide coding domain of the human and mouse  
 $\alpha 2(XI)$  collagen gene

Exon	Exon size bp	Identity	
		Amino acids	Nucleotide sequences
5 (I <sup>a</sup> )	192	78.1	78.6
6	78 (78 <sup>c</sup> )	96.1	96.1
7A	250 (252 <sup>c</sup> )	60	82.7
7B	227 (180 <sup>c</sup> )	50	80.2
7C(II <sup>a</sup> , 7 <sup>b</sup> )	63 (63 <sup>c</sup> )	76.2	85.7

<sup>a</sup> Human *COL11A2* exon number in Ref. 14.

<sup>b</sup> Mouse *Col11a-2* exon number in Ref. 15.

<sup>c</sup> Exon size in *Col11a-2* determined in this study.

putative branch point sequences, essential for the formation of the lariat intermediate in mRNA splicing, are present 32–42 nucleotides upstream of each of the exon 7 potential splice acceptor sites (29), suggesting that these signals could be functional.

**Consequences of Utilizing Additional Splice Signals within Intron 6 of the  $\alpha 2(XI)$  Collagen Gene**—The presence of additional signals within intron 6 of the  $\alpha 2(XI)$  collagen gene potentially increases the size of exon 7 and if used, would divide it into three parts, which we have designated 7A, 7B, and 7C as shown in Figs. 1 and 2. Exon 7C corresponds to the alternatively spliced exon 7 reported previously in human and mouse (14, 15). Utilization of 5' acceptor site 1 in conjunction with the closest donor (site 2) would yield a separate exon 7A of 250 and 252 bp in human and mouse, respectively. Acceptor site 1 could also be used in conjunction with the donor site at the 3' end of exon 7C to yield an enlarged exon 7 consisting of parts A, B, and C, 540 bp in size for human  $\alpha 2(XI)$  collagen mRNA (513 bp for the mouse homolog). Since the sequence at the 5' end of part 7B was more typical of a donor site than an acceptor, splicing of this exon as a separate exon was probably rare. The third splice variation would result from the use of previously reported splice signals for exon 7 flanking part C, yielding a separate exon 7C of 63 bp (14, 15).

Translation of the combined sequences within parts A, B, and C of exon 7 for the human gene revealed an open reading frame, in-frame with that for  $\alpha 2(XI)$  procollagen, which ended in exon 7A because of the presence of the termination codon TGA. Exon 7B also contained a termination codon TAA. In the mouse gene the open reading frame ended within exon 7B with the termination codon TGA. The difference in length of the open reading frame in human and mouse lies in the replacement of the human TGA codon in exon 7A with CCA in the mouse. There were no canonical polyadenylation signals within the intron 7 sequence; however, cleavage and polyadenylation of transcripts containing exon 7A/7B within intron 7 cannot be excluded since in other genes RNA processing can occur even in the absence of a clear consensus sequence (30).

The absence of an alternative long open reading frame in the combined sequence of exons 1–7C suggests that human pro- $\alpha 2(XI)$  mRNAs containing exon 7A or 7B and mouse transcripts containing exon 7B would encode a truncated  $\alpha 2(XI)$  procollagen containing approximately two-thirds of the intact amino-propeptide that terminates in the variable region. Since assembly of trimeric procollagen molecules requires association of the carboxyl-propeptides, these truncated pro- $\alpha 2(XI)$  chains are not expected to associate within trimers. Instead, a non-collagenous polypeptide of 312 and 395 amino acids would be synthesized for human and mouse, respectively. However, in the mouse alternatively spliced transcripts containing exon 7A and

not 7B would encode full-length pro- $\alpha 2(XI)$  collagen chains. The sequence identity between human and mouse for these alternative exons was high at the amino acid and nucleotide levels (Table I), suggesting a functional role for the truncated pro- $\alpha 2(XI)$  chain.

**Expression of Additional Splice Variants of Exon 7 in Human and Mouse**—Reverse transcription PCR and *in situ* hybridization were used to determine if the potential additional splice variants of exon 7 could be expressed *in vivo* in human and mouse fetal tissues.

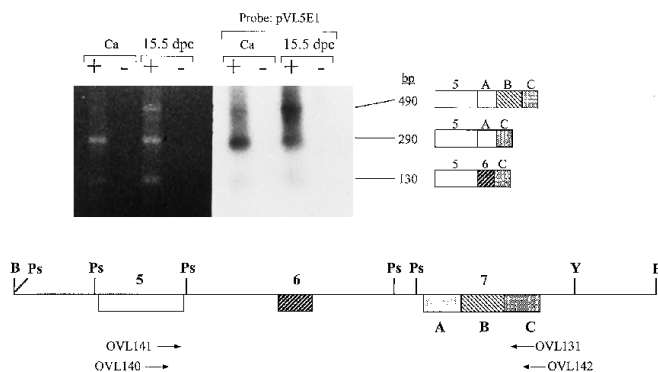
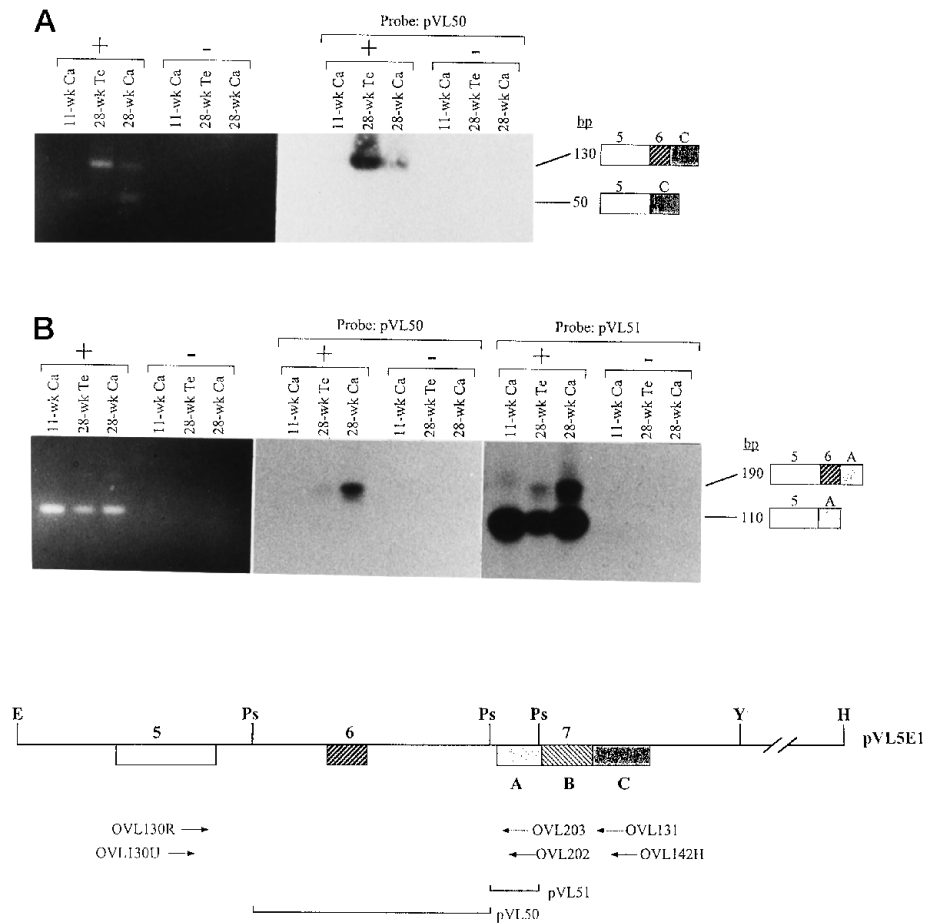
**RT-PCR Analyses**—RT-PCR assays using human fetal RNA and oligonucleotide primers to exons 5 and 7C (see "Experimental Procedures") generated two fragments of 130 and 50 bp from week 28 cartilage and tendon (Fig. 3A). Only the 50-bp fragment was obtained from week 11 cartilage RNA. The sizes of these fragments were consistent with an exon composition of exon 5–6–7C (130 bp) and exon 5–7C (50 bp). The presence of exon 6 within the 130-bp fragment was confirmed by hybridization with an exon 6-specific probe (pVL50; Fig. 3A). Further confirmation of the exon composition of these fragments was obtained by cloning and sequencing the RT-PCR fragments (data not shown). Comparison of the RT-PCR sequence with the human genomic sequence showed that the 130-bp fragment contained exons 5–6–7C and the 50-bp fragment, exons 5–7C; however, no RT-PCR fragments containing exons 7A and 7B were obtained. Although alternative transcripts containing exon 6 have been demonstrated in the mouse (15), expression in human has not been reported (14).

To determine if the failure to detect exons 7A and 7B in the RNA samples truly reflected their absence, primers specific to exons 7A and 7B were also used for RT-PCR (see "Experimental Procedures"). No RT-PCR product was obtained using primers to exons 5 and 7B (data not shown); however, two fragments of 190 and 110 bp were generated from week 28 cartilage and tendon RNA using the exon 7A primer (Fig. 3B). For week 11 cartilage RNA only the 110-bp fragment was obtained (Fig. 3B). The 190-bp fragment could not be cloned because of the trace amount of this fragment obtained; however, hybridization of the 190-bp fragment with both exon 6- and exon 7A-specific probes (Fig. 3B) is consistent with an exon 5–6–7A composition. The 110-bp fragment contained exon 5–7A since it hybridized only with the exon 7A probe and not exon 6. This predicted composition of the 110-bp fragment was confirmed by cloning and sequencing the RT-PCR product (data not shown). Since the presence of the termination codon in exon 7A (Fig. 2), predicted from the genomic sequence was confirmed, this fragment represents a transcript encoding a truncated  $\alpha 2(XI)$  procollagen. Although the relative amounts of the different exon 7-containing transcripts were not measured, in human, mRNAs containing exon 7A were probably of low abundance at the developmental stage studied, since they could only be revealed using exon 7A-specific primers.

Total RNAs from mouse fetuses 15.5 days postcoitus and from 1-day-old neonatal mouse cartilage were also analyzed by RT-PCR using oligonucleotide primers to exons 5 and 7C (see "Experimental Procedures"). For both RNA samples, amplified products of 490, 290, and 130 bp were obtained (Fig. 4). These fragments hybridized with probe pVL5E1, containing exons 5–7 of the human *COL11A2* gene (Fig. 3), suggesting complex alternative splicing within this region of the gene. The exon composition within these fragments was determined by cloning and sequencing the RT-PCR fragments (data not shown) and by comparing the DNA sequences with the mouse genomic sequence.

Sequence analyses of the cloned RT-PCR fragments revealed that the compositions of the different fragments were as fol-

**FIG. 3. Expression of alternatively spliced exons between exons 5 and 7 in the human *COL11A2* gene.** *Top and center*, RT-PCR analyses were of RNA isolated from human week 11 and 28 fetal cartilage (Ca) and from week 28 fetal tendon (Te). RT-PCR products were obtained from reactions priming between exons 5 and 7C (panel A) and exons 5 and 7A (panel B). Photographs of RT-PCR fragments separated on ethidium bromide-stained 2.5% (w/v) agarose gels are shown on the *extreme left panels*, and the autoradiograms of Southern hybridizations are to the *right*. Probes used for hybridization are indicated above each panel. The *plus signs* indicate reactions with added reverse transcriptase; the *minus signs* indicate control reactions without added reverse transcriptase. The absence of PCR products and hybridization signal for control reactions show fragments that were not the result of amplification of genomic DNA. The sizes and exon composition of the RT-PCR fragments are shown to the *right* of the autoradiograms. *Bottom*, schematic diagram showing the exon-intron organization and locations of hybridization probes (horizontal bars) and oligonucleotide primers (arrows; see "Experimental Procedures"). Exons 5–7, open and numbered patterned boxes. Abbreviations: E, *EcoRI*; Ps, *PstI*; Y, *XbaI*; H, *HindIII*.



**FIG. 4. Expression of alternatively spliced exons between exons 5 and 7 in the mouse *Col11a-2* gene.** *Top*, RT-PCR analyses were of RNA isolated from 1-day-old neonatal mouse cartilage (Ca) and from a whole mouse embryo 15.5-days postcoitus (dpc). *Left*, photographs of RT-PCR fragments separated on ethidium bromide-stained 2.5% (w/v) agarose gel. *Right*, autoradiograms of Southern hybridizations using human genomic fragment pVL5E1 (see Fig. 3) as probe. *Plus signs* indicate reactions with added reverse transcriptase; *minus signs* indicate control reactions without added reverse transcriptase. *Right*, sizes and exon composition of RT-PCR fragments. *Bottom*, diagram of exons 5–7 (numbered open and patterned boxes) and the locations of primers used for RT-PCR (arrows; see "Experimental Procedures"). Abbreviations: Ps, *PstI*; Y, *XbaI*; B, *BamHI*.

lows: the 490-bp fragment, exons 5–7A–7B–7C; the 290-bp fragment, exons 5–7A–7C; and the 130-bp fragment, exons 5–6–7C (data not shown). These analyses showed that mouse *Col11a-2* exon 6 was 78 bp, in agreement with the previous report (15). The sizes of exons 7A and 7B, 252 and 198 bp, respectively, were as predicted from the genomic sequence (Fig. 2). Transcripts containing exons 5–7B–7C were not detected as may be

expected if the junction of 7A and 7B acted mainly as a donor site (Fig. 2). Because of the termination codon in exon 7B, the transcript represented by the 490-bp fragment would encode truncated  $\alpha 2(XI)$  procollagen.

**In Situ Hybridization Analyses**—The fetal tissue pattern of expression of transcripts containing alternatively spliced exons was studied by *in situ* hybridization using antisense riboprobes specific for the each of the individual exons 6, 7A, or 7B (Fig. 5). Hybridization signals were at background level for the sense controls for all three riboprobes (reviewed, data not shown). Despite the high sequence homology (80%) between human and mouse exon 7B, background hybridization signals were obtained for the day 42–44 human fetal tissues using the mouse exon 7B probe. In contrast, a strong hybridization signal was obtained in the vertebral bodies using the human exon 7B probe (data not shown), demonstrating the species and sequence specificity of the antisense riboprobes used.

Transcripts containing exon 6 were not detected in human fetal chondrocytes at 38–42 days and at 42–44 days (Fig. 5, A and B). It is interesting that exon 6-containing transcripts were also not expressed by week 11 chondrocytes but were found in week 28 cartilage and tendon (Fig. 4). In human, expression of exon 6 transcripts may be restricted to later developmental stages, unlike the mouse where exon 6-containing mRNAs were clearly found in mouse fetal chondrocytes at 16.5 days (Fig. 5C) and at 1 day postpartum (Fig. 4). The expression of mRNAs containing exon 6 by mouse fetal chondrocytes is in contrast to another study where most of the mouse *Col11a-2* mRNAs in 3-week-old neonate cartilage lacked exon 6, and expression was found mainly in non-chondrogenic tissues such as muscle, brain, heart, and calvaria (15). However, exon 6 mRNAs were reported in developing forelimb bud in embryos between 10 and



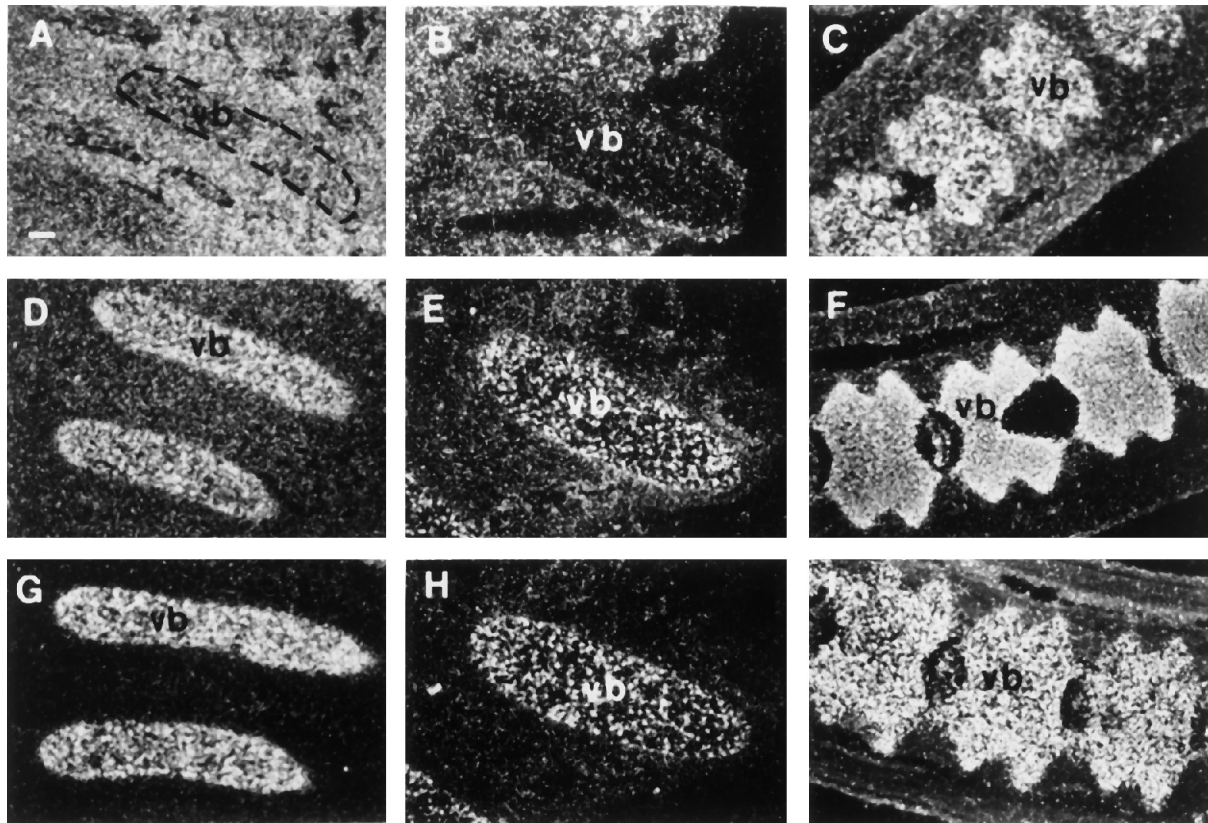


FIG. 5. Expression of alternatively spliced exons 6, 7A, and 7B in human and mouse embryos. Panels A–C, D–F, and G–I are hybridizations with antisense riboprobes for exons 6, 7A, and 7B, respectively. Each panel was photographed under the same magnification under dark-field illumination. There was a lack of exon 6 expression in vertebral bodies (vb, outlined by dashed line) in a week 38–42 (panel A) and week 42–44 human embryo but strong hybridization in a day 16.5 mouse fetus (C). There was strong expression of exons 7A and 7B in the vertebral bodies of a day 38–42 (panels D and G) and 42–44 (panels E and H) human embryo and in a day 16.5 mouse fetus (panels F and I). Bar = 50  $\mu$ m.

14 days. Differences in our findings may reflect stage-specific transient expression of alternative transcripts containing exon 6.

Exon 7A and 7B probes hybridized to chondrogenic tissues such as the chondrocytes within the cartilaginous masses within vertebral bodies in both human and mouse (Fig. 5, D–I). The sensitivity of *in situ* hybridization is usually insufficient for the detection of splicing precursors. The strong hybridization signals for these exons suggest that the inclusion of exons 7A and 7B in transcripts is the result of alternative splicing and does not represent intermediates of mRNA processing. Expression of transcripts containing exon 7B may be stage-specific for human, since exon 7B could only be detected in mRNAs at earlier stages of embryonic development (summarized in Table II). In contrast for the mouse, exon 7B was expressed throughout chondrogenesis from 13.5 days onward<sup>2</sup> (this study).

**Synthesis of a Truncated  $\alpha 2(XI)$  Amino-propeptide**—Although the relative levels of the different exon 7 variants were not determined, it is interesting that exon 7C contains a conserved purine-rich sequence 5'-GAAGAGGAAGAA-3' (Fig. 2), which is similar to that found to act as an exon splicing enhancer in some alternatively spliced genes (for review, see Refs. 31–33). Purine-rich motifs within exons have also been shown to act as splicing silencers in viral genes (34). Exons 7A and 7B also contained purine-rich motifs, but these were shorter. It would be important to determine if these motifs have a role in regulating usage of the alternative exon 7 splice sites.

The presence of premature termination codons in genes has been shown to accelerate the decay rate of mRNAs, a phenomenon called *nonsense-mediated mRNA decay* (35, 36). Skipping

TABLE II  
Developmental stage expression of mRNAs containing exons 6, 7A, 7B, and 7C in human and mouse  
ND, not determined; dpc, days postcoitum; dpp, days postpartum; +, expressed; –, not expressed.

Developmental stage	Exon expressed			
	6	7A	7B	7C
Human				
38–42 days <sup>a</sup>	–	+	+	ND
42–44 days <sup>a</sup>	–	+	+	ND
11 weeks <sup>b</sup>	–	+	–	+
28 weeks <sup>b</sup>	+	+	–	+
Mouse				
15.5 dpc <sup>a,b</sup>	+	+	+	+
16.5 dpc <sup>a</sup>	+	+	+	ND
1 dpp <sup>b</sup>	+	+	+	+

<sup>a</sup> Determined by *in situ* hybridization.

<sup>b</sup> Determined by RT-PCR.

of exons containing a premature termination codon has also been observed for fibrillin and ornithine  $\delta$ -aminotransferase genes (37). The presence of the premature termination codon in mRNAs that include exon 7A and 7B therefore raises questions about the stability and the efficiency of translation of these transcripts and whether levels of the truncated pro- $\alpha 2(XI)$  polypeptide may be low. An additional question would be whether the premature termination affects splice site selection frequency. The *in situ* hybridization data did not show an especially low level of exon 7A- or exon 7B-containing transcripts in human and mouse. The presence of the termination codon therefore probably had not affected mRNA stability. Although the physiologic significance of the synthesis of trun-

cated  $\alpha 2(XI)$  collagen is unknown, the synthesis of truncated  $\alpha 2(XI)$  collagen chain may be important since it is conserved among different species. Another question arising would be whether the PARP peptide previously thought to be a processed product of  $\alpha 2(XI)$  procollagen (38) is the product of mRNAs encoding a truncated pro- $\alpha$  chain. This truncated polypeptide, although larger, includes the PARP domain, and PARP peptide could be derived from it by further processing.

**Heterogeneity within the Amino-propeptide of  $\alpha 2(XI)$  Procollagen Created by Alternative Splicing**—Our data have shown by RT-PCR and *in situ* hybridization analyses that the splice signals present in intron 6 are utilized in both human and mouse, giving rise to additional splice variants of  $\alpha 2(XI)$  procollagen mRNA. Interestingly, the expression of exons 6 and 7B may be developmental stage-specific in human (Table II). In addition to the alternatively spliced exons 6, 7A, and 7B identified in the present study, exons 7C and 8 have also been shown to be alternatively spliced in the mouse (15). With the exception of the alternative transcripts that encode truncated  $\alpha 2(XI)$  procollagen, the added differential usage of these exons could generate great diversity in the structure of the amino-propeptide of  $\alpha 2(XI)$  procollagen and therefore in heterotrimeric type XI procollagen.

Several functions have been proposed for the amino-propeptide of fibril-forming collagens. Proposals based on studies on types I and III procollagen include a role of retained propeptide in preventing premature fibril formation intracellularly and in regulating fibril diameter (39, 40). Biosynthetic studies have shown that parts of the amino-propeptides of pro- $\alpha 1(XI)$  and pro- $\alpha 2(XI)$  are retained within fibrils (41, 42), which may be important for their proposed roles in regulating fibrillogenesis and modulating fibril diameter (43–45).

It is also interesting that the amino-propeptides of type I and III procollagen have been shown to be capable of down-regulating collagen synthesis by cultured fibroblasts and in cell-free translation assays (39, 46, 47). The possible synthesis of truncated amino-propeptide for  $\alpha 2(XI)$  procollagen raises further questions about potential roles for these polypeptides in regulating biosynthesis at the translational level.

The complex alternative splicing found for the  $\alpha 1(XI)$  and  $\alpha 2(XI)$  collagen genes with its resulting structural diversity within the amino-terminal region and the developmental stage differences in expression raise questions about possible differing functional roles for the amino-propeptide. The roles are expected to differ depending on whether the amino-propeptide is retained, cleaved, or synthesized as a separate polypeptide.

#### REFERENCES

- Kielty, C. M., Hopkinson, I., and Grant, M. E. (1993) in *Connective Tissue and Its Heritable Disorders: Molecular, Genetic, and Medical Aspects* (Royce, P. M., and Steinmann, B., eds) pp. 103–148, John Wiley & Sons, New York
- Kivirikko, K. I. (1993) *Ann. Med.* **25**, 113–126
- Mayne, R., and Brewton, R. G. (1993) *Curr. Opin. Cell Biol.* **5**, 883–890
- Chu, M. L., and Prockop, D. J. (1993) *Collagen: Gene Structure in Connective Tissue and Its Heritable Disorders: Molecular, Genetic, and Medical Aspects*, pp. 149–166, John Wiley & Sons, New York
- Tillet, E., Mann, K., Nischt, R., Pan, T., Chu, M.-L., and Timpl, R. (1995) *Eur. J. Biochem.* **228**, 160–168
- Juvonen, M., Pihlajaniemi, T., and Autio-Harmainen, H. (1993) *Lab. Invest.* **69**, 541–551
- Rehn, M., and Pihlajaniemi, T. (1995) *J. Biol. Chem.* **270**, 4705–4711
- Bennett, V. D., and Adams, S. L. (1990) *J. Biol. Chem.* **265**, 2223–2230
- Nah, H.-D., Niu, Z., and Adams, S. L. (1994) *J. Biol. Chem.* **269**, 16443–16448
- Burgeson, R. E., and Hollister, D. W. (1979) *Biochem. Biophys. Res. Commun.* **87**, 1124–1131
- Zhidkova, N. I., Brewton, R. G., and Mayne, R. (1993) *FEBS Lett.* **326**, 25–28
- Yoshioka, H., and Ramirez, F. (1990) *J. Biol. Chem.* **265**, 6423–6426
- Lui, V. C. H., Kong, R. Y. C., Nicholls, J., Cheung, A. N. Y., and Cheah, K. S. E. (1995) *Biochem. J.* **311**, 511–516
- Zhidkova, N. I., Justice, S. K., and Mayne, R. (1995) *J. Biol. Chem.* **270**, 9486–9493
- Tsumaki, N., and Kimura, T. (1995) *J. Biol. Chem.* **270**, 2372–2378
- Oxford, J. T., Doege, K. J., and Morris, N. P. (1995) *J. Biol. Chem.* **270**, 9478–9485
- Eyre, D., and Wu, J. J. (1987) in *Structure and Function of Collagen Types* (Mayne, R., and Burgeson, R. E., eds) pp. 261–281, Academic Press, Orlando, FL
- Ryan, M. C., and Sandell, L. J. (1990) *J. Biol. Chem.* **265**, 10334–10339
- Lui, V. C. H., Ng, L. J., Sat, E. W. Y., and Cheah, K. S. E. (1996) *Genomics* **32**, 401–412
- Patten, B. M. (1976) *Human Embryology*, McGraw Hill, Maidenhead, OK
- O'Rahilly, R., and Muller, F. (1987) *Developmental Stages in Human Embryos*, Carnegie Institution of Washington, Washington, D. C.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1992) *Bio/Technology* **24**, 104–108
- Henikoff, S. (1987) *Methods Enzymol.* **155**, 156–165
- Devereux, J., Haeblerli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395
- Ng, L. J., Tam, P. P. L., and Cheah, K. S. E. (1993) *Dev. Biol.* **159**, 403–417
- Lovell-Badge, R. H. (1987) in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* (Robertson, E. J., ed) pp. 153–182, IRL Press, Oxford
- Wilkinson, D. G., and Green, J. (1990) in *Postimplantation Mammalian Embryos: A Practical Approach* (Copp, A. J., and Cockcroft, D. L., eds) pp. 155–171, IRL Press, Oxford
- Shapiro, M. B., and Senapathy, P. (1987) *Nucleic Acids Res.* **15**, 7155–7174
- Green, M. R. (1991) *Annu. Rev. Cell Biol.* **7**, 559–599
- Wahle, E., and Keller, W. (1992) *Annu. Rev. Biochem.* **61**, 419–440
- Watanabe, A., Tanaka, K., and Shimura, Y. (1993) *Genes Dev.* **7**, 407–418
- Amendt, B. A., Si, Z.-H., and Stoltzfus, C. M. (1995) *Mol. Cell. Biol.* **15**, 4606–4615
- Humphrey, M. B., Bryan, J., Cooper, T. A., and Berget, S. M. (1995) *Mol. Cell. Biol.* **15**, 3979–3988
- Staffa, A., and Cochrane, A. (1995) *Mol. Cell. Biol.* **15**, 4597–4605
- Lim, S.-K., Sigmund, C. D., Gross, K. W., and Maquat, L. E. (1992) *Mol. Cell. Biol.* **12**, 1149–1161
- Cui, Y., Hagan, K. W., Zhang, S., and Peltz, S. W. (1995) *Genes Dev.* **9**, 423–436
- Dietz, H. C., Valle, D., Francomano, C. A., Kendzior, R. J., Jr., Pyeritz, R. E., and Cutting, G. R. (1993) *Science* **259**, 680–683
- Neame, P. J., Young, C. N., and Treep, J. T. (1990) *J. Biol. Chem.* **265**, 20401–20408
- Kuhn, K., Weistner, M., Krieg, T., and Muller, P. K. (1982) *Connect. Tissue Res.* **10**, 43–50
- Fleischmajer, R., Perlsh, J. S., Burgeson, R. E., Shaikh-Bahai, F., and Timpl, R. (1990) *Ann. N. Y. Acad. Sci.* **580**, 161–175
- Morris, N. P., and Bachinger, H. P. (1987) *J. Biol. Chem.* **262**, 11345–11350
- Thom, J. R., and Morris, N. P. (1991) *J. Biol. Chem.* **266**, 7262–7269
- Birk, D. E., Fitch, J. M., Barbiariz, J. P., Doane, K. J., and Linsenmayer, T. F. (1990) *J. Cell Sci.* **95**, 649–657
- Birk, D. E., Silver, F. H., and Trelstad, R. L. (1991) in *Cell Biology of Extracellular Matrix* (Hay, E. D., ed) pp. 221–254, Plenum Press, New York
- Linsenmayer, T. F., Gibney, E., Iggo, F., Gordon, M. K., Fitch, J. M., Fessler, L. I., and Birk, D. E. (1993) *J. Cell Biol.* **121**, 1181–1189
- Paglia, L. M., Wilczek, J., de Leon, L. D., Martin, G. R., Horlein, D., and Muller, P. K. (1979) *Biochemistry* **18**, 5030–5034
- Wu, C. H., Walton, C. M., and Wu, G. Y. (1991) *J. Biol. Chem.* **266**, 2983–2987

**Nucleic Acids, Protein Synthesis, and  
Molecular Genetics:**

**Extensive Alternative Splicing within the  
Amino-propeptide Coding Domain of  $\alpha$   
2(XI) Procollagen mRNAs: EXPRESSION  
OF TRANSCRIPTS ENCODING  
TRUNCATED PRO- $\alpha$  CHAINS**

Vincent C. H. Lui, Ling Jim Ng, Eric W. Y.  
Sat, John Nicholls and Kathryn S. E. Cheah  
*J. Biol. Chem.* 1996, 271:16945-16951.

Access the most updated version of this article at <http://www.jbc.org/content/271/28/16945>

Find articles, minireviews, Reflections and Classics on similar topics on the [JBC Affinity Sites](#).

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 39 references, 24 of which can be accessed free at  
<http://www.jbc.org/content/271/28/16945.full.html#ref-list-1>