Isolation of cDNAs encoding the complete sequence of bovine type X collagen

Evidence for the condensed nature of mammalian type X collagen genes

J. Terrig THOMAS,* Alvin P. L. KWAN, Michael E. GRANT and Raymond P. BOOT-HANDFORD

Department of Biochemistry and Molecular Biology, School of Biological Sciences, Stopford Building, University of Manchester, Oxford Road, Manchester M13 9PT, U.K.

The complete primary structure of the bovine $\alpha 1(X)$ collagen chain was determined by nucleotide sequencing of cDNA clones. The overlapping cDNA clones encode 3144 bp with a 5'-terminal untranslated region of 148 bp, a 2025 bp reading frame and a 3'-terminal untranslated region of 971 bp. This represents the first complete sequence of a mammalian type X collagen cDNA and has allowed a number of informative comparisons to be made with the previously published chick $\alpha 1(X)$ sequence. The primary translation products of both bovine and chick type X collagen are 674 amino acid residues in length and there is a 73.3 % identity at the amino acid level (67.8 % at the base level). Sequence analyses reveal that the greatest degree of identity between the two species occurs within the triple-helical domain and the C-terminal non-collagenous domain, whereas the identity within the N-terminal non-collagenous domain is markedly lower. The inter-chain disulphide-bonding observed previously within the triple helix of bovine type X collagen is explained by the presence of two cysteine residues within an imperfection of the triple-helical domain encoded by -Gly-Xaa-Cys-Xaa-Yaa-Cys-Xaa-Yaa-Gly-. Southern-blot analyses of bovine genomic DNA demonstrate that the bovine type X collagen gene is likely to have a condensed structure, similar to that of the chick, with at least 1.3 kb of the coding sequence being contained within one exon.

INTRODUCTION

The growth of mammalian long bones occurs by the process of endochondral ossification, which involves precise temporal and spatial changes in the deposition of new extracellular matrix macromolecules as well as the degradation and removal of a pre-existing matrix. The transition from cartilage to bone takes place at epiphyseal growth plates, where chondrocytes differentiate through a series of well-defined stages, namely proliferation, maturation and hypertrophy (for review see Caplan, 1988). Concomitant changes associated with chondrocyte hypertrophy are the synthesis of type X collagen (Kielty *et al.*, 1984; Schmid & Linsenmayer, 1985; Kwan *et al.*, 1986a), an increase in the number of matrix vesicles (for review see Wuthier, 1988), an increase in alkaline phosphatase activity (Väänänen, 1980) and the generation of a calcified matrix, which is subsequently replaced by bone.

To date, most studies on collagen type X have been conducted with chick tissues and antibodies (polyclonal and monoclonal), and cDNAs to chick collagen type X are now available (Schmid & Linsenmayer, 1985; Kwan *et al.*, 1986a; Ninomiya *et al.*, 1986). Success in the chick system has been, in large part, attributable to the fact that collagen type X is a major biosynthetic product of chick chondrocytes undergoing endochondral ossification. However, marked differences are apparent in the development of the long bones in birds compared with mammals. The latter exhibit very narrow epiphyseal growth plates in comparison with the avian system and consequently there have been technical difficulties in isolating significant quantities of mammalian type X collagen for biochemical analyses. In addition, studies at the mRNA and gene levels in mammalian species have not been possible because the chick cDNAs encoding collagen type X do not cross-hybridize with their mammalian mRNA counterparts.

Analyses of the small amounts of intact type X collagen extractable from bovine epiphyseal growth plate has shown that significant differences exist in the properties of bovine collagen type X compared with chick collagen type X (Ayad *et al.*, 1987). The bovine species contains inter-chain disulphide bonds, and pepsinization yields $\alpha 1(X)$ chains of slightly higher molecular size than is found in the chick system (Grant *et al.*, 1985; Ayad *et al.*, 1987; Kirsch & von der Mark, 1990; Thomas *et al.*, 1990a).

The experiments described below were initiated to isolate cDNAs for mammalian collagen type X and thereby open up new areas of research into endochondral bone formation. The recent discovery that type X collagen synthesis is a major feature of the biosynthetic activity of human chondrocytes from osteoarthritic joints, and therefore that inappropriate expression of this gene product may play a role in the pathogenesis of this condition (J. Hoyland, J. T. Thomas, M. E. Grant, R. P. Boot-Handford & C. J. Freemont, unpublished work), provides further stimulus for determining the structure and function of the mammalian gene and its translated product.

MATERIALS

Bacterial collagenase type 1A, ascorbate, isopropyl 1-thio- β -D-galactopyranoside, anti-(rabbit IgG) antibody conjugated with alkaline phosphatase, Nitro Blue Tetrazolium and 5-bromo-4-chloroindol-3-yl phosphate were from Sigma Chemical Co. (Poole, Dorset, U.K.). Trypsin, Dulbecco's modified Eagle's medium, foetal-calf serum, glutamine, penicillin, streptomycin, proteinase K, ultra-pure agarose and low-melting-point agarose were from Gibco-BRL (Paisley, Strathclyde, Scotland, U.K.). Oligo(dT)-cellulose was from Collaborative Research (Bedford, Beds., U.K.). Bovine growth-plate oligo(dT)-primed cDNA library was from Clontech Laboratories (Palo Alto, CA, U.S.A.). Nitrocellulose membrane was from Schleicher und Schüll

* To whom correspondence should be addressed.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X53556.

(Dassel, Germany). All restriction enzymes, M13mp18 and M13mp19 were from Boehringer Mannheim (Mannheim, Germany). L-[5-³H]Proline was from Amersham International (Amersham, Bucks., U.K.). [α -³²P]dCTP and [α -³²P]dATP were from ICN (Irvine, CA, U.S.A.). Vacugene blotting apparatus and oligolabelling kit were from Pharmacia-LKB (Milton Keynes, Bucks., U.K.). Sequenase sequencing kit was from United States Biochemical Corp. (Cleveland, OH, U.S.A.). GemSeq RT sequencing system was from Promega Biotech (Madison, WI, U.S.A.). Chick α 1(II) cDNA (pYN 2142) was kindly donated by Dr. B. R. Olsen (Harvard Medical School, Boston, MA, U.S.A.).

METHODS

Cell culture

Growth-plate cartilage was dissected from the proximal end of the tibia and the distal end of the femur of bovine fetuses ranging from 170 to 235 days. The cartilage was cleaned of adhering tissue, finely chopped and digested for 3.5 h at 37 °C with bacterial collagenase type 1A (100 mg/ml) and trypsin (0.3 mg/ml) in Dulbecco's modified Eagle's medium. The released chondrocytes were plated out at 1×10^6 cells/ml within three-dimensional type I collagen gels (1.42 mg/ml) as described by Gibson *et al.* (1983). The cultures were maintained with Dulbecco's modified Eagle's medium supplemented with 10% (v/v) foetal-calf serum, penicillin (0.1 mg/ml), streptomycin (300 μ g/ml), glutamine (2 mM) and ascorbate (50 μ g/ml). At day 6 in culture, cells were radiolabelled for 24 h with fresh medium containing L-[5-³H]proline (20 μ Ci/ml) for analyses of newly synthesized collagens as described previously (Thomas *et al.*, 1990b).

RNA isolation

Medium was removed from the cultures on day 6 and the collagen gels plus cells were homogenized in medium containing 4 M-guanidinium isothiocyanate, 5 mM-sodium citrate, 0.1 M-2-mercaptoethanol and 0.5% Sarkosyl. CsCl (1 g/2.5 ml) was added to the cell-culture homogenate, which was layered into Beckman Quick-Seal ultracentrifuge tubes (16 mm $\times$ 76 mm) containing 5.7 M-CsCl in 0.1 M-EDTA, pH 7 (2.5 ml). RNA was isolated by CsCl-density-gradient centrifugation as described by Chirgwin *et al.* (1979), with a Beckman Ti50 rotor (40000 rev./min for 16 h at 18 °C). The RNA pellet obtained after centrifugation was digested with proteinase K as described previously (Thomas *et al.*, 1990b). The digest was extracted once with phenol/chloroform (1:1, w/v) mixture and the aqueous phase was transferred to a fresh tube. RNA was precipitated in 0.1 vol. of 3 M-sodium acetate buffer, pH 5.2, and 2.2 vol. of ethanol at -20 °C for at least 2 h. RNA was recovered by centrifugation, washed with 70% (v/v) ethanol and dried under vacuum before being dissolved in diethyl pyrocarbonate-treated deionized water.

The total cellular RNA was used to isolate polyadenylated RNA by oligo(dT)-cellulose chromatography (Aviv & Leder, 1972). Polyadenylated RNA was shown to be intact by Northern blotting (as described below) and probing with the chick α 1(II) cDNA (pYN 2142), which cross-reacts with its bovine mRNA counterpart.

Library construction

The polyadenylated RNA was used to synthesize an oligo(dT)-primed cDNA library cloned into the *Eco*RI site of the λ gt11 expression vector (Clontech Laboratories). The library contained 1.4×10^6 independent clones with an insert size range of 0.8–4 kb.

Library screening and isolation of clones

The library was screened with a polyclonal antibody raised to chick collagen type X collagen (Kwan *et al.*, 1986a). The library was plated on to *Escherichia coli* Y1090 and allowed to grow for approximately 3 h at 42 °C. Nitrocellulose filters, which had been soaked in 10 mM-isopropyl 1-thio- β -D-galactopyranoside and allowed to dry, were then placed on to the cultures and the cultures were incubated at 37 °C for 3 h. The filters were removed, replaced with similarly treated second filters and incubated for a further 3 h at 37 °C. Both sets of filters were blocked for 15 min in TBST (10 mM-Tris/HCl buffer, pH 8.0, containing 150 mM-NaCl and 0.05% Tween 20) containing 20% (v/v) donor-calf serum, rinsed in TBST and incubated for 30 min with the polyclonal antibody to chick type X collagen (1/500 dilution in TBST). Filters were washed with TBST (3 $\times$ 5 min) and incubated for 30 min with anti-(rabbit IgG) antibody conjugated with alkaline phosphatase (1/7500 dilution in TBST). Filters were again washed (3 $\times$ 5 min with TBST) and developed in alkaline phosphatase reaction buffer (100 mM-Tris/HCl buffer, pH 9.5, containing 100 mM-NaCl and 5 mM-MgCl₂) containing 0.03% Nitro Blue Tetrazolium and 0.016% 5-bromo-4-chloroindol-3-yl phosphate. Immunopositive plaques were picked, re-plated and re-screened until pure. Inserts were released from large-scale preparations of the cloned λ vectors by digestion with the restriction enzyme *Eco*RI and purified by agarose-gel electrophoresis. cDNA inserts were subcloned into M13mp18 or M13mp19 (Messing & Vieira, 1982) for preparation of single-stranded DNA templates for sequence analysis.

Overlapping clones were isolated by screening the library by plaque hybridization with randomly primed [α -³²P]dCTP-labelled cDNA restriction fragments prepared by using an oligolabelling kit.

The cDNA probe BII4.2, specific for α 1(II) collagen, was also isolated from the same bovine growth-plate cDNA library. The above library was screened by using the chick α 1(II) collagen cDNA (pYN2142) as ³²P-labelled probe. A number of clones were isolated, one of which (BII4.2) contained a cDNA insert encoding 400 bp of the non-collagenous C-terminal domain of bovine type II collagen.

Isolation of a cDNA encoding the complete N-terminal domain of bovine type X collagen by the polymerase chain reaction

The polymerase-chain-reaction protocol used was based on that described by Goblet *et al.* (1989). The template for the PCR reaction was single-stranded M13mp18 containing the 5'-terminal *Eco*RI-digested cDNA fragment from BX10.10. The primers used were the M13 universal primer and a 17-mer oligonucleotide prepared to the most 5'-terminal region of the cDNA BX10.10 encoding part of the N-terminal non-collagenous domain (see Fig. 2). The polymerase-chain-reaction product was random primer labelled and used to screen the λ gt11 library by plaque hybridization as described above.

DNA sequencing and analysis

Both strands of the isolated cDNA clones were sequenced in the M13 vector utilizing [α -³²P]dATP by the dideoxy chain-termination method (Sanger *et al.*, 1977) with the Sequenase or the GemSeq RT sequencing system. The combination of the two methods eliminated ambiguous regions in the sequence. The primers used were either the M13 universal primer or sequence-specific 17-mer oligonucleotides (made with an Applied Biosystems DNA synthesizer). Sequences were ordered, the predicted amino acid sequence was deduced and sequence similarity to the chick α 1(X) cDNA determined by using the SEQAID program (D. D. Rhoads and D. J. Roufa, Kansas State University, 1986).

Hybridization analysis

Northern blots of polyadenylated RNA were prepared by electrophoresis of heat-denatured RNA as described previously (Boot-Handford *et al.*, 1987) followed by blotting on to nitrocellulose, with a Vacugene blotting apparatus. The blots were air-dried, baked under vacuum at 80 °C for 2 h and pre-hybridized for 6 h at 37 °C in medium containing 50% (v/v) formamide, 4×SSC, 1 mg of sheared salmon sperm DNA/ml, 1×Denhardt's solution (Maniatis *et al.*, 1982) and 0.1% SDS. (1×SSC is 0.15 M-NaCl/15 mM-sodium citrate buffer, pH 7; 1×Denhardt's solution is 0.02% Ficoll 400/0.02% polyvinylpyrrolidone/0.002% BSA.) Southern blots were prepared from bovine placental DNA digested with a number of restriction enzymes. The DNA digests were electrophoresed on a 1% agarose gel (20 V overnight). The gel was soaked in 0.5 M-NaOH containing 1.5 M-NaCl for 1 h, neutralized in 0.5 M-Tris/HCl buffer, pH 8.0, containing 1.5 M-NaCl (3×15 min) and equilibrated in 4×SSC before blotting, baking and pre-hybridizing as above. The desired cDNA fragments were released from double-stranded M13mp18 or M13mp19 vectors by restriction-endonuclease digestion, purified by agarose-gel electrophoresis and labelled with [α -³²P]dCTP by random primer labelling to specific radioactivities of at least 4×10⁸ c.p.m./μg of DNA.

RESULTS AND DISCUSSION

Isolation and characterization of bovine type X collagen cDNA clones

An oligo(dT)-primed cDNA library was constructed in λgt11 (Clontech Laboratories) with the use of polyadenylated RNA from foetal bovine growth-plate chondrocytes cultured within three-dimensional collagen gels and known to be synthesizing large amounts of type X collagen (Thomas *et al.*, 1990a). The library was screened with a polyclonal antibody raised to chick type X collagen (Kwan *et al.*, 1986a), and a clone containing a 1.6 kb cDNA insert (BX2.131) was isolated. When polyadenylated RNA extracted from foetal bovine growth-plate cultures was subjected to Northern-blot analysis, BX2.131 hybridized with an RNA species of approx. 3 kb (Fig. 1). Furthermore, BX2.131 did not hybridize with polyadenylated RNA extracted from either foetal bovine epiphyseal cartilage taken from a non-mineralizing region (Fig. 1), or from kidney, placenta or tendon (results not shown). The size of the mRNA recognized by BX2.131 and its pattern of tissue expression is consistent with that of type X collagen.

Nucleotide sequencing of the 1.6 kb insert from BX2.131 revealed that this cDNA contained 480 bp coding for 160 amino acid residues of helical domain, 465 bp encoding the complete C-terminal non-collagenous domain (155 amino acid residues) and 638 bp of 3'-terminal untranslated region of the bovine α1(X) mRNA (Fig. 2a).

Additional cDNA clones were isolated from the λgt11 library by re-screening with the 1 kb *EcoRI*/*HindIII* restriction fragment from BX2.131 (Fig. 2b). A number of clones were isolated (Fig. 2a), one of which (BX8.4) contained a cDNA insert of 2836 bp encoding the 3'-end of the N-terminal non-collagenous domain (10 bp), the complete triple-helical domain, the C-terminal non-collagenous domain and 972 bp of 3'-terminal untranslated region up to and including the poly(A) addition signal.

cDNAs encoding the complete N-terminal domain of bovine type X collagen were obtained by screening the λgt11 library with a polymerase-chain-reaction fragment containing 90 bp of the 5'-end of the cDNA BX10.10 (Fig. 2a) and 71 bp of M13mp18 polylinker. A number of clones were isolated, one of which (BX26.2) contained an insert that extended 147 bp 5' of the proposed ATG translation-initiation start codon (Fig. 2a). A

Fig. 1. Size and distribution of mRNAs recognized by cDNAs isolated from the bovine growth-plate cDNA library

(a) Northern blot of 5 μg of polyadenylated RNA extracted from bovine growth-plate chondrocytes cultured within type I collagen gels hybridized with the type X collagen cDNA BX2.131. (b) Northern-blot analysis of 15 μg of foetal bovine total RNA obtained from (lane 1) growth-plate cartilage and (lane 2) non-mineralizing epiphyseal cartilage hybridized with cDNA probes for type X collagen (BX2.131) and type II collagen (BII4.2). Size markers refer to the position of 28 S and 18 S rRNA.

Fig. 2. cDNAs encoding bovine type X collagen

(a) Diagram showing the positions of a number of overlapping cDNAs encoding bovine type X collagen. PCR indicates the region of clone BX10.10 amplified by polymerase chain reaction to produce the N-terminal non-collagenous-domain-specific probe with which clone BX26.2 was identified (see the Methods section). (b) A partial restriction map of the full-length cDNA sequence. (c) The deduced domain structure obtained from the complete sequence. ■, N-terminal non-collagenous domain; □, triple helix; ▨, C-terminal non-collagenous domain. Straight lines indicate the position and lengths of the 5'-terminal and 3'-terminal untranslated regions.

partial restriction map of the complete cDNA encoding bovine collagen X is shown in Fig. 2(b) and the complete domain structure of the bovine α1(X) cDNA is shown in Fig. 2(c).

The cDNA sequence across the internal *EcoRI* site (Fig. 2b) was determined by sequence analysis of the 2 kb restriction

[illegible]

fragment obtained upon digestion of λ clone BX8.4 with *Kpn*I, with the use of sequence-specific oligonucleotide primers.

The translation of the bovine collagen X mRNA probably starts at the methionine codon defined by nucleotide residues 1–3 (Fig. 3). The methionine residue at this position is followed by a typical hydrophobic signal peptide sequence (Watson, 1984), 18 amino acid residues in length (Fig. 3), and shows a high degree of sequence similarity to that previously published for chick α 1(X) (LuValle *et al.*, 1988). Two other putative ATG start codons encoded by base numbers –145 to –143 and –119 to –117 are followed by in-frame stop codons encoded by base numbers –118 to –116 and –107 to –105 respectively. The *N*-terminal non-collagenous domain contains 38 amino acid residues, four more than in the corresponding region of the chick. The triple-helical domain of bovine type X collagen contains 463 amino acid residues compared with 460 in chick α 1(X), and the *C*-terminal domain contains 155 amino acid residues compared with 162 in chick α 1(X) (Yamaguchi *et al.*, 1989).

The primary translation products of both bovine and chick type X collagen are 674 amino acid residues long, and there is an overall 73.3 % identity at the amino acid level (67.8 % at the base level). Bovine type X collagen has a calculated M_r of 65557 for the non-hydroxylated non-glycosylated form compared with M_r 66444 for chick type X collagen (Yamaguchi *et al.*, 1989).

Sequence comparisons between bovine and chick α 1(X) collagen

Signal peptide and *N*-terminal non-collagenous domain. The putative signal peptide of the bovine collagen X cDNA consists of 18 amino acid residues and is 67 % identical at the amino acid level (61 % at the base level) with that previously published for chick type X collagen (LuValle *et al.*, 1988). The similarity between the signal peptide sequences of bovine and chick type X collagen suggests that the signal-peptidase cleavage site in the bovine collagen X is between the glycine and valine residues encoded by the bases 52–57 (Fig. 3).

Sequence comparisons between the *N*-terminal non-collagenous domains of bovine and chick type X collagen show that, compared with the triple-helical domain and *C*-terminal non-collagenous domain, the *N*-terminal non-collagenous domain contains the least amount of sequence similarity to chick α 1(X). There are two distinct regions of sequence similarity interspersed with dissimilar sequences (Fig. 3). The first region of sequence similarity, of 17 amino acid residues, is 53 % identical at the amino acid level (68.6 % at the base level), whereas the second region, of 15 amino acid residues, is 40 % identical at the amino acid level (55.6 % at the base level). As the sequences in the *N*-terminal non-collagenous domain are less conserved than in the remainder of the molecule, it would suggest that this domain is less important in the structure/function of type X collagen.

Fig. 3. Complete nucleotide sequence of the bovine type X collagen cDNA, the amino acid sequence of the conceptual translation product and comparison with the coding region of chick α 1(X) nucleotide sequence

The *N*-terminal and *C*-terminal non-collagenous domains are indicated by boxed areas. Numbers refer to the number of nucleotides counting from the proposed initiation ATG codon of the bovine α 1(X) cDNA. The positions of the eight imperfections in the -Gly-Xaa-Yaa- repeat are indicated by double horizontal lines (==). Gaps created by sequences being present in one species only are indicated by dashes (---). Locations of cysteine and conserved tyrosine residues in the bovine α 1(X) cDNA are indicated by circles (●) and squares (■) respectively. The putative *N*-linked oligosaccharide attachment site is indicated by triangles (▼▼). The location of the poly(A) addition signal (AATAAA) is indicated in bold type.

Triple-helical domain. Compared with chick α 1(X), the bovine triple-helical domain contains one additional Gly-Xaa-Yaa triplet. These extra amino acid residues account, in part, for the observation that when the pepsin-resistant triple-helical domains of bovine and chick type X collagen are subjected to SDS/PAGE the bovine species migrates at a slightly lower rate (Ayad *et al.*, 1987).

Within the triple-helical domain there is 75.8 % identity with the previously published sequence for chick α 1(X) at the amino acid level (68.7 % at the base level). Furthermore, 65 % of the codon changes take place at the third base position of codons and thus have little influence on the sequence of the translated product. It has recently been demonstrated that there is considerable similarity between the sequences encoding chick α 1(X) and rabbit α 1(VIII) collagen (Yamaguchi *et al.*, 1989). Computer analysis of the sequences encoding the triple-helical domains of bovine α 1(X) and rabbit α 1(VIII) reveal 57.4 % identity at the amino acid level compared with 56 % identity between chick α 1(X) and rabbit α 1(VIII) (Fig. 4).

The triple helix of bovine type X collagen contains four imperfections of the type Gly-Xaa-Gly, three of the type Gly-Xaa-Yaa-Xaa-Yaa-Gly and one imperfection of the type Gly-Xaa-Cys-Xaa-Yaa-Cys-Xaa-Yaa-Gly (Fig. 3). All the imperfections within the bovine type X collagen triple-helical domain are at positions identical with those in the chick α 1(X) triple helix (Fig. 3). Of the three imperfections of the type Gly-Xaa-Yaa-Xaa-Yaa-Gly, two start with the sequence Gly-Ile, and are thus potential mammalian-collagenase-cleavage sites (Fig. 3). All the imperfections are the same amino acid number in length, with the exception of the cysteine-containing interruption encoded by the sequence (Gly-His-Cys-Thr-Pro-Cys-Arg-Pro-Gly), which occurs at the position of a Gly-Xaa-Gly sequence in chick α 1(X). The presence of the two cysteine residues within the triple-helical domain accounts for the disulphide-bonding observed in bovine type X collagen (Grant *et al.*, 1985; Ayad *et al.*, 1987; Kirsch & von der Mark, 1990; Thomas *et al.*, 1990a). The two cysteine residues present in the bovine α 1(X) triple helix appear to have originated by single base substitutions of glycine codons present in the chick α 1(X).

Earlier work carried out by Schmid *et al.* (1986) has shown that chick type X collagen is unusual among the other collagen types in possessing two mammalian-collagenase-cleavage sites. Analysis of the predicted amino acid sequence for bovine type X collagen suggests that a 32000- M_r fragment will be produced upon digestion with mammalian collagenase and it will be disulphide-bonded. The functional significance of a 32000- M_r disulphide-bonded triple-helical mammalian-collagenase-resistant domain within bovine type X collagen is not known. However, work on chick type X collagen (Schmid *et al.*, 1986; Gadher *et al.*, 1988) has demonstrated that the 32000- M_r mammalian-collagenase-cleavage fragment is thermally stable and may persist *in vivo*.

***C*-Terminal non-collagenous domain.** The *C*-terminal non-collagenous domain of bovine α 1(X) is 155 amino acid residues in length compared with 162 amino acid residues in chick α 1(X) (Yamaguchi *et al.*, 1989). The first 15 amino acid residues of bovine α 1(X) show little sequence similarity to chick α 1(X), whereas the remaining 140 residues show an identity of 83 % at the amino acid level (76 % at the base level) (Fig. 3). As in the triple-helical domain, the majority of the base changes take place in the third base position (71.7 %). Of particular note are the conserved locations of 13 tyrosine residues and a putative *N*-linked oligosaccharide-attachment site (Fig. 3). In addition to the two cysteine residues within the triple helix, a third cysteine residue is present in the *C*-terminal non-collagenous domain at

Fig. 4. Comparison of the amino acid sequences of bovine and chick $\alpha 1(X)$ and rabbit $\alpha 1(VIII)$ collagens

The N-terminal and C-terminal non-collagenous domains are indicated by boxed areas. Gaps created by sequences not being present in all three species are indicated by dotted lines (...).

the same location as that of the only cysteine residue present in chick $\alpha 1(X)$ (Yamaguchi *et al.*, 1989). It is not clear to what extent this cysteine residue contributes to disulphide-bonding of type X collagen. For the most part neither chick nor bovine type X collagen appears to contain disulphide bonds within the C-terminal non-collagenous domain. However, there is evidence that a proportion of chick type X collagen molecules do become disulphide-bonded in this region (Kwan *et al.*, 1986b).

The highly conserved nature of most of the C-terminal domain between the bovine and chick species (Fig. 3) suggests that this region of the molecule has an important role in the structure/assembly of type X collagen *in vivo*. Further evidence for the C-terminal domains having a significant function has come from the complete cDNA sequence encoding rabbit type VIII collagen (Yamaguchi *et al.*, 1989), demonstrating 58% identity at the amino acid level between the C-terminal three-quarters of the C-terminal non-collagenous domains of rabbit type VIII collagen and chick type X collagen. A similar comparison of bovine collagen X and rabbit collagen VIII in the same region reveals 55% identity at the amino acid level.

It should be noted that the 5'-terminal and 3'-terminal

untranslated regions of bovine and chick $\alpha 1(X)$ mRNA showed no significant sequence similarity.

Codon usage

Analysis of codon usage for proline and alanine in the bovine $\alpha 1(X)$ triple-helical domain demonstrates that, in the third base position, guanine occurred least frequently (Table 1), as has been noted for many other collagen types including chick collagen X (Ninomiya *et al.*, 1986; Baldwin *et al.*, 1989). Similarly, as has been noted for the fibrillar collagens (Baldwin *et al.*, 1989), cytosine was rarely used as the third base of codons for proline in the Yaa position within the repeating Gly-Xaa-Yaa sequences of rabbit type VIII collagen and both chick and bovine type X collagen (Table 1). It has been suggested that such a bias in codon usage may be due to the dinucleotide sequence C-G being prone to mutation (Bird, 1986). However, the codon usage for glycine is not conserved in the manner observed for proline and alanine. In the fibrillar collagens and chick $\alpha 1(X)$ guanine is the least common third base, whereas in rabbit type VIII collagen and bovine type X collagen there is a more random distribution in the third base position of codons for glycine (Table 1). Such

Table 1. Codon usage in the triple-helical domains of human type II, chick and bovine type X and rabbit type VIII collagens

Data for human type II collagen are taken from Baldwin *et al.* (1989), those for chick type X collagen are taken from Ninomiya *et al.* (1986) and those for rabbit type VIII collagen are taken from Yamaguchi *et al.* (1989).

Amino acid	Human $\alpha 1(\text{II})$	Chick $\alpha 1(\text{X})$	Bovine $\alpha 1(\text{X})$	Rabbit $\alpha 1(\text{VIII})$	Third base
Glycine	0.44	0.29	0.17	0.18	U
	0.27	0.21	0.25	0.16	C
	0.23	0.41	0.33	0.30	A
	0.06	0.08	0.25	0.36	G
Codons examined	342	153	152	160	
Proline (total)	0.60	0.40	0.27	0.29	U
	0.24	0.16	0.14	0.19	C
	0.13	0.43	0.55	0.50	A
	0.03	0.01	0.04	0.02	G
Codons examined	227	116	120	130	
Proline (Yaa position)	0.73	0.44	0.25	0.33	U
	0.08	0.04	0.08	0.02	C
	0.15	0.52	0.62	0.62	A
	0.04	0.00	0.04	0.02	G
Codons examined	117	65	71	79	
Alanine	0.57	0.28	0.46	0.22	U
	0.24	0.19	0.18	0.44	C
	0.14	0.52	0.32	0.33	A
	0.03	0.00	0.03	0.00	G
Codons examined	97	21	22	9	

a lack of conservation in the codon usage for glycine in $\alpha 1(\text{X})$ across species suggests that the low abundance of guanine in the third base position is not necessary for the normal function of the type X collagen gene.

Evidence for the bovine type X collagen gene having a condensed structure

The chick collagen X gene has been characterized recently (LuValle *et al.*, 1988) and its structure appears unique in comparison with all other vertebrate collagen genes analysed to date (for review see Kuhn, 1987). Whereas the latter genes contain numerous exons with intron sequences representing at least 60% of the total DNA content of each gene, the chick collagen X gene contains only three exons (LuValle *et al.*, 1988), one of which (exon 3) encodes the entire triple helix, the C-terminal non-collagenous domain and continues into the 3'-terminal untranslated region of the mRNA (Ninomiya *et al.*, 1986).

Given that the structures of collagen genes are highly conserved in a type-specific manner across vertebrate species (Kuhn, 1987), we have conducted initial studies to determine whether the bovine collagen X gene is highly condensed in a fashion similar to that found in the chick. Genomic DNA extracted from bovine placenta and λ gt11 containing the near-full-length cDNA, BX8.4 (Fig. 2a), were digested with a number of restriction enzymes, Southern-blotted on to nitrocellulose and analysed by cDNA hybridization with the 1.6 kb cDNA, BX2.131 (Fig. 2a). The banding patterns obtained suggested that the bovine type X collagen gene was indeed likely to have a condensed structure, similar to that of the chick $\alpha 1(\text{X})$ collagen gene.

Restriction of the genomic DNA with any one of the five

Fig. 5. Evidence for the condensed nature of the bovine type X collagen gene

(a) Southern blot of bovine genomic DNA unrestricted (U) or restricted with *Bam*HI (B), *Eco*RI (E), *Pst*I (P), *Hind*III (H) and *Sst*I (S). (b) Southern blot of bovine genomic cDNA (first two lanes) and near-full-length cDNA (BX8.4) (third lane) restricted with *Sst*I (S) or *Eco*RI and *Sst*I (E/S). Both filters were hybridized with the 1.6 kb cDNA insert from BX2.131 (shown in Fig. 2).

enzymes tested to date that do not cut within the 1.6 kb cDNA always yields only one band when Southern blots are probed with this cDNA (Fig. 5a). Restriction of the bovine genomic DNA with enzymes known to cleave the cDNA probe (BX2.131) at a single locus (*Sst*I and *Hind*III) produces two bands upon Southern-blot analysis (Fig. 5a). Furthermore, restriction with *Hind*III produces two genomic bands (Fig. 5a), the lower- M_r band having a mobility that appears to be identical with that of the 990 bp fragment predicted from the partial restriction map of the full-length bovine $\alpha 1(\text{X})$ cDNA (Fig. 2b). In addition, restriction of the bovine genomic DNA with both *Eco*RI and *Sst*I produces a band having a mobility identical with that of the 1.3 kb fragment derived by *Eco*RI and *Sst*I digestion of the near-full-length cDNA BX8.4 (Fig. 5b). The combination of the above results demonstrates that the bovine type X collagen gene has a condensed structure and that the exon recognized by the 1.6 kb cDNA (BX2.131) must be at least 1.3 kb in length.

Conclusive evidence that the human type X collagen gene is also condensed has come from initial sequence analysis of a human type X collagen genomic clone containing a continuous open reading frame encoding 156 bp of triple helix, 486 bp of the complete C-terminal domain and 339 bp of 3'-terminal untranslated region (J. T. Thomas, M. E. Grant & R. P. Boot-Handford, unpublished work).

The isolation of the complete cDNA for bovine type X collagen and comparison with the previously published sequence for chick type X collagen (Yamaguchi *et al.*, 1989) has highlighted potentially important regions of the molecule. Future studies, using cell-culture models, will be able to focus on the consequences of site-directed mutations in collagen X. In addition, the full characterization of the human gene for collagen X will be of vital importance in determining whether the type X collagen gene locus is linked with heritable disorders affecting cartilage such as some forms of osteoarthritis and the chondrodysplasias.

We thank Anne Marriott for kindly providing a number of cultures for RNA isolation and for her skilled dissection of bovine growth plates.

We also thank Carl Cresswell and Ailsa Parry for providing excellent technical assistance. The support of the Nuffield Foundation (Oliver Bird Trust) is gratefully acknowledged.

REFERENCES

- Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408–1412
- Ayad, S., Kwan, A. P. L. & Grant, M. E. (1987) *FEBS Lett.* **220**, 181–186
- Baldwin, C. T., Reginato, A. M., Smith, C., Jimenez, S. A. & Prockop, D. J. (1989) *Biochem. J.* **262**, 521–528
- Bird, A. P. (1986) *Nature (London)* **321**, 209–213
- Boot-Handford, R. P., Kurkinen, M. & Prockop, D. J. (1987) *J. Biol. Chem.* **262**, 12475–12478
- Caplan, A. I. (1988) *Ciba Found. Symp.* **136**, 3–21
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
- Gadher, S. J., Eyre, D. R., Duance, V. C., Wotton, S. F., Heck, L. W., Schmid, T. M. & Woolley, D. E. (1988) *Eur. J. Biochem.* **175**, 1–7
- Gibson, G. J., Kielty, C. M., Garner, C., Schor, S. L. & Grant, M. E. (1983) *Biochem. J.* **211**, 417–426
- Goblet, C., Prost, E. & Whalen, R. G. (1989) *Nucleic Acids Res.* **17**, 2144
- Grant, W. T., Sussman, M. D. & Balian, G. (1985) *J. Biol. Chem.* **260**, 3798–3803
- Kielty, C. M., Hulmes, D. J. S., Schor, S. L. & Grant, M. E. (1984) *FEBS Lett.* **169**, 179–184
- Kirsch, T. & von der Mark, K. (1990) *Biochem. J.* **265**, 453–459
- Kuhn, K. (1987) in *Biology of the Extracellular Matrix: Structure and Function of Collagen Types* (Mayne, R. & Burgeson, R. E., eds.), pp. 1–42, Academic Press, New York
- Kwan, A. P. L., Freemont, A. J. & Grant, M. E. (1986a) *Biosci. Rep.* **6**, 155–162
- Kwan, A. P. L., Sear, C. H. J. & Grant, M. E. (1986b) *FEBS Lett.* **206**, 267–272
- LuValle, P., Ninomiya, Y., Rosenblum, N. D. & Olsen, B. R. (1988) *J. Biol. Chem.* **263**, 18378–18385
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor
- Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276
- Ninomiya, Y., Gordon, M., van der Rest, M., Schmid, T., Linsenmayer, T. & Olsen, B. R. (1986) *J. Biol. Chem.* **261**, 5041–5050
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Schmid, T. M. & Linsenmayer, T. F. (1985) *J. Cell Biol.* **100**, 598–605
- Schmid, T. M., Mayne, R., Jeffrey, J. J. & Linsenmayer, T. F. (1986) *J. Biol. Chem.* **261**, 4184–4189
- Thomas, J. T., Boot-Handford, R. P., Marriott, A., Kwan, A. P. L., Ayad, S. & Grant, M. E. (1990a) *Ann. N.Y. Acad. Sci.* **580**, 477–479
- Thomas, J. T., Boot-Handford, R. P. & Grant, M. E. (1990b) *J. Cell Sci.* **95**, 639–648
- Väänänen, H. K. (1980) *Histochemistry* **65**, 143–148
- Watson, M. E. E. (1984) *Nucleic Acids Res.* **12**, 5145–5164
- Wuthier, R. E. (1988) *ISI Atlas Sci. Biochem.* **1**, 231–241
- Yamaguchi, N., Benya, P. D., van der Rest, M. & Ninomiya, Y. (1989) *J. Biol. Chem.* **264**, 16022–16029

Received 30 April 1990/18 June 1990; accepted 26 June 1990