The human collagen X gene

Complete primary translated sequence and chromosomal localization

J. Terrig THOMAS,*†, Carl J. CRESSWELL,* Bharat RASH,* Hans NICOLAI,† Tania JONES,† Ellen SOLOMON,† 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., †Imperial Cancer Research Fund Laboratories, Lincoln's Inn Fields, London WC2A 3PX, U.K.

We report on the complete primary translated sequence of human $\alpha 1(X)$ collagen, deduced from a genomic clone, and the chromosomal localization of the human collagen X gene. The primary translated product of human collagen X is encoded by two exons of 169 bp and approx. 2940 bp. The 169 bp exon encodes 15 bp of 5'-end untranslated sequence, 18 amino acid residues (54 bp) of signal peptide and $33\frac{1}{2}$ amino acid residues (100 bp) of the *N*-terminal non-collagenous domain. The 2940 bp exon encodes $4\frac{2}{3}$ amino acid residues (14 bp) of the *N*-terminal non-collagenous domain, the complete triple-helical domain of 463 amino acid residues (1389 bp), the complete *C*-terminal non-collagenous domain of 161 amino acid residues (483 bp) and 1054 bp of 3'-end untranslated sequence up to and including a potential cleavage/polyadenylation signal. The size of the intron separating the two exons, as estimated by partial sequencing and Southern-blot analyses, is approx. 3200 bp. By a combination of somatic cell hybrid screening and hybridization *in situ* the human collagen X gene (*COL10A1*) has been assigned to the distal end of the long arm of chromosome 6 at the locus 6q21–6q22.3.

INTRODUCTION

Current understanding of the structure of collagen X has been derived largely from studies on chick collagen X extracted from embryonic tissues or cultured hypertrophic chondrocytes. The molecule is distinguished by being a homotrimer [$\alpha 1(X)_3$] in which the triple-helical domain is approximately half the length of that found in the classical fibrillar cartilage collagen II. The collagen X molecule also retains a large non-collagenous globular extension at its *C*-terminus, which undoubtedly plays a key role in the intracellular assembly of the triple-helical collagen X molecules. In addition, this region may also be important for the extracellular macromolecular aggregation of collagen X to form lattice-like networks similar to those of type IV and VIII collagens (Kwan *et al.*, 1991). Indeed, it is of note that collagen types VIII and X show a high degree of similarity at the base and amino acid level (Yamaguchi *et al.*, 1989; Thomas *et al.*, 1991) and also have similar gene structures (Yamaguchi *et al.*, 1991).

A great deal of interest has been focused on the possible role of type X collagen in the process of skeletogenesis in view of its specific and transient expression at sites of endochondral ossification (Schmid & Linsenmayer, 1985; Kwan *et al.*, 1986; Marriott *et al.*, 1991). The marked decrease of collagen X synthesis in vitamin D-deficiency rickets (Kwan *et al.*, 1989) and its increased synthesis in the presence of the mineralizing agent calcium β -glycerophosphate (Thomas *et al.*, 1990) suggest that collagen X expression and deposition in the matrix are essential for normal bone development.

In order to study the potential role of collagen X in human pathological diseases such as osteoarthritis and the chondrodysplasias it is essential to obtain information on the human collagen X gene.

We have recently obtained the complete nucleotide sequence of the bovine $\alpha 1(X)$ collagen cDNA (Thomas *et al.*, 1991), from which we were able to highlight the most conserved regions of the

molecule in comparison with the chick (i.e. the triple-helical domain and 90% of the *C*-terminal non-collagenous domain). We also demonstrated that the mammalian collagen X gene appears highly condensed in a fashion similar to that previously reported for the chick (LuValle *et al.*, 1988). The bovine $\alpha 1(X)$ cDNA restriction fragment encoding most of the highly conserved region of the *C*-terminal non-collagenous domain has now been used to isolate the human collagen X gene. Here we present the complete primary structure of human collagen X together with its chromosomal localization.

MATERIALS

Isopropyl 1-thio- β -D-galactopyranoside was from Sigma Chemical Co. (Poole, Dorset). Ultra-pure agarose was from GIBCO BRL (Paisley, Strathclyde, Scotland, U.K.). Nitrocellulose membrane was from Schleicher und Schüll (Dassel, Germany). All restriction enzymes, bacteriophages M13mp18 and M13mp19 and plasmids pUC118 and pUC119 were from Boehringer Mannheim (Mannheim, Germany). The radiolabelled compounds [α - 32 P]dCTP and [α - 32 P]dATP were from ICN (Irvine, CA, U.S.A.). The Vacugene blotting apparatus and oligo-labelling kit were from Pharmacia-LKB (Milton Keynes, Bucks., U.K.). The Sequenase sequencing kit was from United States Biochemical Corp. (Cleveland, OH, U.S.A.) and the GemSeq RT sequencing kit was from Promega Biotech (Madison, WI, U.S.A.).

METHODS

Screening of genomic libraries and isolation of genomic $\alpha 1(X)$ collagen clones

A 492 bp *Hind*III–*Sst*I restriction fragment from the clone BX2.131 (Thomas *et al.*, 1991), encoding 360 bp of *C*-terminal

† 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 X60302.

non-collagenous domain and 132 bp of 3'-end untranslated sequence was used as a probe to screen a human genomic library cloned into the λ EMBL3 SP6/T7 vector (Clontech Laboratories, Palo Alto, CA, U.S.A.). The library was screened by standard plaque hybridization with randomly primed [α - 32 P]dCTP-labelled cDNA, prepared by using an oligolabelling kit, to specific radioactivities of at least 5×10^9 c.p.m./ μ g of DNA. Hybridization was performed in medium containing $4 \times$ SSC, 50% (v/v) formamide, $1 \times$ Denhardt's solution, 1 mg of salmon sperm DNA/ml and 0.1% SDS at 37 °C. ($1 \times$ SSC is 0.15 M-NaCl/15 mM-sodium citrate buffer, pH 7; Denhardt's solution is 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.002% BSA.) Filters were washed to a final stringency of $0.5 \times$ SSC/0.1% SDS at 55 °C. A number of positive plaques were picked, re-plated and re-screened until pure. Genomic DNA inserts were released from large-scale preparations of the cloned λ vectors by digestion with the restriction enzymes *Sst*I or *Xho*I and purified by gel electrophoresis.

Hybridization and nucleotide sequence analyses

Southern blots were prepared from positive genomic clones, or human placental DNA, digested with a number of restriction enzymes. The DNA digests were electrophoresed on 1% agarose gels and processed for hybridization as described previously (Thomas *et al.*, 1991). Following detailed restriction-endonuclease mapping of the positive genomic clones, restriction fragments containing exons (based on their hybridization to bovine collagen X cDNA clones) were subcloned into M13mp18 or M13mp19 (Messing & Vieira, 1982) or pUC118 or pUC119 (Vieira & Messing, 1987) for preparation of single-stranded DNA for sequence analysis.

Both strands of the isolated genomic restriction fragments encoding exons in the human collagen X gene were sequenced with the use of [α - 32 P]dATP by the dideoxy chain-termination method (Sanger *et al.*, 1977) with the Sequenase or GemSeq RT sequencing systems. The primers used were either the M13 universal primer or sequence-specific 17-mer oligonucleotides (synthesized with an Applied Biosystems DNA synthesizer). Sequences were ordered, the predicted amino acid sequence was deduced and sequence similarity searches were conducted by using either the SEQAID program (D. D. Rhoads & D. J. Roufa, Kansas State University, 1986) or the University of Wisconsin Genetics Computer Group sequence analysis package.

Chromosomal localization

Southern-blot hybridization of collagen X to human $\times$ rodent hybrids. DNA (7 μ g) from human and rodent parental lines and human $\times$ rodent hybrids, CTP34B4, CTP412A2, Horp9.5, DT-1.2.4, Dur4.3, F4SC13C112, Mog34A4, FG10, Mog1B/9, Sif-4A31, SIR74ii, Twin/D12, 3W4C15, MCP6, EDAG2.8.9, EDAG3R, 56.47 and clone 2D, whose chromosome contents have been previously established (Gross & Harris, 1975; Jones *et al.*, 1976; Povey *et al.*, 1980; Kietly *et al.*, 1982; Edwards *et al.*, 1985; Phillips *et al.*, 1985; Solomon *et al.*, 1985; Wong *et al.*, 1987; Willson *et al.*, 1987) was digested to completion with *Hind*III. The DNA samples were size-fractionated by electrophoresis on a 0.8% agarose gel and alkaline-blotted on to Hybond N+ membranes (Reed & Mann, 1985) after depurination by a 10 min treatment in 0.25M-HCl. A 750 bp DNA fragment (containing 40 bp M13 polylinker, 349 bp 3'-end untranslated sequences and 361 bp coding for most of the C-terminal non-collagenous region of human collagen X) was labelled to a specific radioactivity of 1×10^8 c.p.m./ μ g of DNA with [3 H]deoxynucleotides and oligonucleotide primers (Feinberg & Vogelstein, 1983). Hybridization of the filter-bound DNAs to the probe was carried out in medium containing $5 \times$ SSPE,

$5 \times$ Denhardt's solution, 0.5% SDS and 1 mg of yeast RNA/ml. ($1 \times$ SSPE is 0.18 M-NaCl/10 mM-sodium phosphate buffer, pH 7.4, containing 1 mM-EDTA.) Following hybridization, filters were washed at 65 °C in medium containing 0.1% SDS and $0.1 \times$ SSC.

Regional assignment of the human collagen X gene by hybridization *in situ*. Human lymphocytes obtained from normal female 46(XX) blood were cultured with phytohaemagglutinin for 72 h at 37 °C (Watt & Stephens, 1986). Bromodeoxyuridine was then added to give a final concentration of 200 μ g/ml to ensure good-quality post-hybridization chromosomal banding. The cells were washed with fresh medium 16–17 h later and incubated in thymidine-rich (10 μ M) medium for an additional 6–7 h. Harvesting of the cells and all further procedures were performed under subdued lighting. Metaphase spreads were prepared by using a modification of previously described methods (Watt & Stephens, 1986). Cells were allowed to swell in 75 mM-KCl at 37 °C for 10 min before being centrifuged at 1500 *g* for 5 min, resuspended and fixed in methanol/acetic acid (3:1, v/v). Cells were spread on to twin frosted slides and air-dried.

Hybridization *in situ* was essentially carried out as described by Harper & Saunders (1981) and Zabel *et al.* (1983), with the 750 bp DNA fragment described above as a probe. Hybridization to metaphase spreads was carried out for 18–22 h at 37 °C in medium containing 50% (v/v) formamide, 0.24 M-NaCl, 0.03 M-sodium citrate, 0.04 M-sodium phosphate, 10% (w/v) dextran sulphate and 1000 times the probe DNA concentration of salmon sperm DNA at pH 7.0. The experiment was performed with probe concentrations of 0.02 and 0.004 μ g/ml. Post-hybridization washes were at 39 °C in medium containing 50% (v/v) formamide, 0.3 M-NaCl and 0.03 M-sodium citrate, pH 7.0, followed by $2 \times$ SSC, pH 7.0. The slides were then dehydrated and immersed in Ilford K5 emulsion. After 6–10 days slides were developed and G-banded by using a variation of the method of Wolff & Perry (1974). Slides were stained for 30 min in Hoechst 33258 (0.5 μ g/ml), exposed to long-wave u.v. light for 15 min and stained with Wright's stain (diluted 1:3 in 50% Sørensen's buffer).

RESULTS AND DISCUSSION

Isolation and characterization of human collagen X genomic clones

A number of positive clones were isolated on screening the human genomic library with the 492 bp *Hind*III–*Sst*I restriction fragment (see the Methods section) derived from the bovine α 1(X) cDNA. Detailed restriction-endonuclease mapping revealed that one clone (HX.3) contained an insert size of approx. 15 kb, potentially encoding the complete human collagen X gene (Fig. 1). Restriction fragments containing exon sequences were identified on Southern blots by hybridization with bovine collagen X cDNAs. These fragments were subcloned into M13 and pUC vectors for sequence analysis (see the Methods section). Extensive nucleotide sequence analysis of the human genomic subclones and comparison with the nucleotide sequences of the complete bovine α 1(X) cDNA (Thomas *et al.*, 1991) and the chick α 1(X) gene (LuValle *et al.*, 1988) revealed the following. (1) The complete primary translated product of human collagen X is encoded by two exons of 169 bp and approx. 2940 bp. These two exons have equivalent counterparts in exons 2 and 3 in the chick α 1(X) collagen gene. (2) The 169 bp exon encodes 15 bp of 5'-end untranslated sequence, 18 amino acid residues (54 bp) of signal peptide and $33\frac{1}{2}$ amino acid residues (100 bp) of the N-terminal non-collagenous domain (Fig. 2). The exon ends in a G, which forms the first base of the next codon. (3) The 2940 bp

Fig. 1. Diagram showing the structure of the human $\alpha 1(X)$ collagen gene in the region containing the two exons encoding the complete primary translated sequence

A partial restriction map of the human genomic clone HX.3 is illustrated above and the locations of the exon-intron splice junctions within the mRNA are shown below the diagram of the gene structure. Key to restriction enzymes: S, *Sst*I; H, *Hind*III; E, *Eco*RI. The exons are represented by the black boxed areas and the 750 bp restriction fragment used in the chromosomal localization studies is indicated by a bar. The domain structure of human collagen X deduced from nucleotide sequencing is also shown below the diagram of the gene structure: $\square$, N-terminal non-collagenous domain; $\square$, triple helix; $\square$, C-terminal non-collagenous domain.

exon encodes 4 $\frac{2}{3}$ amino acid residues (14 bp) of the N-terminal non-collagenous domain, the complete triple-helical domain of 463 amino acid residues (1389 bp), the complete C-terminal non-collagenous domain of 161 amino acids residues (483 bp) and 1054 bp of 3'-end untranslated sequence up to and including a potential cleavage/polyadenylation signal (Fig. 2). The exon structure of the human collagen X gene is shown in Fig. 1.

The exon/intron boundaries were defined by a combination of Southern-blot analyses and sequence similarity to the chick $\alpha 1(X)$ gene. Both exons are defined at their 5'-ends by the presence of acceptor splice junctions, which are preceded by consensus lariat branch-point sequences 35–40 bp 5'. Similarly, the 3'-end of the smaller 169 bp exon is defined by the presence of a donor splice junction. The size of the intron separating the two exons, as estimated by partial sequencing and Southern-blot analyses, is approximately 3200 bp, which is at least 1000 bp longer than the equivalent intron in the chick gene (LuValle *et al.*, 1988).

Sequence comparisons of collagen X across species

We have previously reported an extensive sequence comparison between bovine and chick collagen X and rabbit $\alpha 1(VIII)$ (Thomas *et al.*, 1991), and consequently we here only present comparative data between human and chick $\alpha 1(X)$ where significant differences are apparent.

The putative primary translation product of human collagen X is 680 amino acid residues, whereas both bovine and chick $\alpha 1(X)$ are 674 amino acid residues in length. The overall sequence identities between human, bovine and chick $\alpha 1(X)$ are shown in Table 1. As expected, human and bovine collagen X share a higher degree of identity than do the human and chick proteins. Human $\alpha 1(X)$ has a calculated M_r of 66128 for the non-hydroxylated non-glycosylated form compared with M_r 65557 for bovine $\alpha 1(X)$ and M_r 66444 for chick $\alpha 1(X)$.

Signal peptide and N-terminal non-collagenous domain. The putative signal peptide sequence located within the putative exon 2 of the human collagen X gene is 18 amino acid residues in length, the same as in bovine and chick $\alpha 1(X)$, and is 77.8% identical at the amino acid level (88.9% at the base level) with

bovine $\alpha 1(X)$. In our previous study we noted that there was a distinct reduction of sequence conservation between the N-terminal non-collagenous domains of bovine and chick collagen X when compared with the triple-helical and C-terminal non-collagenous domains. However, the N-terminal non-collagenous domain of human and bovine $\alpha 1(X)$ are both 38 amino acid residues in length, four more than in the chick sequence, and share an identity of 81.6% at the amino acid level (86.8% at the base level). This finding indicates that, although the N-terminal non-collagenous domain is not highly conserved across a broad spectrum of species, it is highly conserved within the two mammalian species studied to date and, potentially, may have an important functional role.

Triple-helical domain. The triple-helical domains of both human and bovine $\alpha 1(X)$ are 463 amino acid residues in length, compared with 460 in the chick, and there is 86.6% sequence identity at the amino acid level (87% at the base level) between the human and bovine proteins. The locations and form of all the imperfections are conserved with the exception of the cysteine-containing imperfection (-Gly-Xaa-Cys-Xaa-Yaa-Cys-Xaa-Yaa-Gly-) present in the bovine $\alpha 1(X)$ that is replaced by a -Gly-Xaa-Gly- interruption in the human sequence (bases 574–582) (Fig. 2). In this respect the human collagen X sequence is more like that of chick. The absence of the cysteine residues indicates that human collagen X does not contain any interchain disulphide bonds. This prediction is supported by evidence from studies at the protein level, which demonstrate that the migration of human collagen X on SDS/PAGE is not affected by reducing agents (Marriott *et al.*, 1991). Indeed, it appears from other studies that bovine is the only species studied so far that possesses disulphide-bonded collagen X. Studies on the rabbit (Remington *et al.*, 1983) and rat (P. Rooney, M. E. Grant & J. McClure, unpublished work) indicate that, in these species, collagen X does not contain disulphide bonds.

The other imperfections are all highly conserved. Human collagen X has two imperfections corresponding to mammalian collagenase cleavage sites (Gly-Ile-Yaa-Xaa-Yaa-Gly), one being identical with that found in bovine $\alpha 1(X)$ (bases 451–468 in Fig. 2, residues 151–156 in Fig. 3), the other (bases 1435–1452 in Fig. 2, residues 479–484 in Fig. 3) exhibiting a conservative substitution at the second Xaa position (valine for threonine). The other imperfection of the Gly-Xaa-Yaa-Xaa-Yaa-Gly type (bases 301–318 in Fig. 2, residues 101–106 in Fig. 3) also has one amino acid change in the human compared with the bovine $\alpha 1(X)$ chain and is a valine-for-threonine substitution at the second Yaa position. Two of the imperfections of the type Gly-Xaa-Gly are identical in human and bovine (bases 652–660 and 889–897 in

Table 1. Overall sequence identities at the amino acid level for the N-terminal non-collagenous (N), the helical (H) and the C-terminal non-collagenous domains (C) of human $\alpha 1(X)$ compared with bovine and chick $\alpha 1(X)$

The numbers in parentheses refer to the sequence identities at the base level.

	Sequence identity (%)		
	Human		
	N	H	C
Bovine	78.6 (86.9)	86.6 (87)	89.4 (86.1)
Chick	55.4 (60.6)	74.8 (71)	77.7 (72.3)

3132 TTTGCTGTTCTCAACCATTCTTTCAGGCTTTTCATTGACACAAATAAATAACATCAATAGATTTTATAGGGATGGGTGGCTTT 3214

HUMAN X	MLPQIPFLLVSLNLVHGVFYAERYQTPTGIKGLPNTKTQFFIPYTIKSKGIAVRGEQG	60
BOVINE X	TAL M T PS A G VLS	
	TPGPPGPAGPRGHPPGSPGPKPGYGSPLQGEPLPGPPGSAVGKPGVPGLPGKPGER	120
	I T P Q T L	
	GPYGPKGVDVGPAGLPGRGPPGPGIPGPAGISVPGKPGQGPAGPRGFPGEKGAPG	180
	LN I P E TS	
	VPGMNGQKGEKMGYAGPRPGERGLPGPQGPPTGSPGPGVVKRGENGVPQPGIKGDRGFP	240
	L HCT C P L L Q V	
	GEMGPIGPPGPGPPGERGPEGIGKPGAAGAPGQPGIPGTKGLPGAPGIAGPPGPPGFGK	300
	R AA S Q P I M Q T L A	
	PGLPGLKGERGPAGLPGGPGAKGEQGPAGLPKGLTGPPGNMGPPGPKGIPGSHGLPGP	360
	Q V S H EA P S NP	
	KGETGPAGPAGYPGAKGERGSPGSDGKPGYPGKPGLDGPKGNPGLPGPKGDPGVGGPPGL	420
	M V N S L E N IA S	
	PGPVGPAGAKGMPGHNGEAGPRGAPGIPGTRGPIGPPGIPGFPGSKGDPSPGPPGPGI	480
	V V V T	
	ATKGLNGPTGPPGPPGRHSGEPGLPGPPGPPGPPGQVMPEGFIKAGQASLSGTPLV	540
	V NA VAL D V ***** F	
	SANQGVGTGMPVSAFTVILSKAYPAIGTPIPFDKILYNRQOHYDPTGTGIFTQIPGIYYFS	600
	K K	
	YHVHVKGTHVWVGLYKNGTPVMYTYDEYTKGYLDQASGSAIDLTLTENDQVWLQLPNAESN	660
	I A I V G	
	GLYSSEYVHSSFSGLVAPM * 680	
	P	

Fig. 3. Comparison of the amino acid sequences of human and bovine $\alpha 1(X)$ collagen

The *N*-terminal and *C*-terminal non-collagenous domains are indicated by the boxed areas. The gap created by the extra six amino acid residues present in the human sequence compared with the bovine is indicated by asterisks (***). The position of the eight imperfections in the Gly-Xaa-Yaa repeat are indicated by double horizontal lines (==) and the positions of the unusual -Gly-Gly- sequences are indicated by single horizontal lines (—).

Fig. 2, residues 218–220 and 297–299 respectively in Fig. 3), whereas the one closest to the *N*-terminal non-collagenous domain (bases 250–258 in Fig. 2, residues 84–86 in Fig. 3) has a tyrosine residue in the human $\alpha 1(X)$ sequence and a threonine residue in its bovine counterpart.

Two unusual sequence motifs found within the triple-helical domain of human $\alpha 1(X)$ that are not present in either chick or bovine $\alpha 1(X)$ are Gly-Xaa-Yaa- triplets in which glycine is also represented in either the Xaa or the Yaa position (Gly-Gly-Pro-Gly and Gly-Val-Gly-Gly). The biological significance of such sequences (residues 317–320 and 413–416 respectively) is not known, but, together with the other imperfections, their presence may well add to the flexibility of the collagen X triple helix. Analysis of codon usage demonstrated the same pattern of third-base preferences for glycine, proline and alanine as reported for the bovine $\alpha 1(X)$ cDNA (Thomas *et al.*, 1991).

Fig. 2. Composite nucleotide sequence encoding the complete primary translated product of human $\alpha 1(X)$ collagen and comparison with the bovine $\alpha 1(X)$ collagen cDNA sequence

The sequence is initiated at the 5'-end of the 169 bp exon and the numbers refer to nucleotides counting from the proposed initiation ATG codon of human $\alpha 1(X)$ collagen. The location of the intron-exon splice junction is indicated by an arrow at the boundary between the two exons. The boxed areas define the *N*- and *C*-terminal non-collagenous domains. The positions of the eight imperfections in the Gly-Xaa-Yaa- repeat are indicated by double horizontal lines (==). The positions of the unusual -Gly-Gly- sequences are indicated by dotted lines (.....). Gaps created by sequences being present in one species only are indicated by dashes (—). The 34 bp repeated sequences in the 3'-end untranslated region are underlined and the locations of the two cleavage/polyadenylation signals (AATAAA) are in bold type.

C-Terminal non-collagenous domain. The *C*-terminal non-collagenous domain of human $\alpha 1(X)$ is 161 amino acid residues in length, six more than in bovine $\alpha 1(X)$ and one less than in chick $\alpha 1(X)$. This domain has an overall identity of 91.6% at the amino acid level (87.2% at the base level) in human compared with that of bovine $\alpha 1(X)$. However, the first 12 amino acid residues share only 66.7% identity at the amino acid level (83.3% at the base level), whereas 143 of the remaining 149 amino acid residues (residues 538–680 in Fig. 3) are 93.7% identical (86.9% at the base level). The six additional amino acid residues (residues 532–537 in Fig. 3) present in human compared with bovine collagen X represent an extension of the region of relatively low sequence conservation, and are in the same location as the seven extra amino acid residues present in chick $\alpha 1(X)$. Similar variabilities in the *N*-terminal region of *C*-terminal non-collagenous domains have also been noted in fibrillar collagen types I, II and III (Ninomiya *et al.*, 1984). The significance of such variability has been postulated to be due to either one of two possibilities. Firstly, the sequence connecting the *C*-terminal end of the triple-helical domain and the *C*-terminal non-collagenous domain could represent a flexible hinge region. Secondly, the different *N*-termini of the *C*-terminal non-collagenous domains could have collagen-type-specific functions and may be required for the specific associations that occur during assembly of the triple helix (Ninomiya *et al.*, 1984). The data that we have obtained showing marked sequence conservation throughout the *C*-terminus of the $\alpha 1(X)$ across species, with the exception of *N*-terminal end, indicates that this variable junction probably represents a flexible hinge.

As for the remainder of the human $\alpha 1(X)$ *C*-terminal non-collagenous domain, the locations of the 13 tyrosine residues, the putative *N*-linked oligosaccharide-attachment site and the single cysteine residue are all conserved as in bovine and chick $\alpha 1(X)$.

Fig. 4. Chromosomal localization of the human collagen X gene

(a) Southern-blot hybridization of genomic DNA from human, mouse, hamster and a panel of human $\times$ rodent somatic hybrids containing fragments of chromosome 6. Clone 2D is a somatic hybrid that only contains the human X chromosome and is present as a control for MCP6, which contains both the human X chromosome and a fragment of chromosome 6. (b) Idiogram of the human G-banded chromosome 6 illustrating the locations of the somatic hybrids used in (a) and the distribution of autoradiographic grains in 50 cells after hybridization *in situ* with 0.004 μ g of the 3 H-labelled collagen-X-specific probe DNA/ml.

3'-End untranslated region. In our earlier study on the bovine $\alpha 1(X)$ cDNA we noted that there was very little sequence identity between the 3'-end untranslated regions of bovine and chick $\alpha 1(X)$. In contrast, comparison of the same region in the human and bovine $\alpha 1(X)$ sequence revealed an overall identity of 80%. However, unlike the coding domains, sequence identity was only achieved after the introduction of gaps in the sequences to allow for numerous insertions and deletions (Fig. 2). Of significant interest was the conservation in the human collagen X gene of the potential cleavage/polyadenylation signal located at the 3'-end of the bovine $\alpha 1(X)$ cDNA (Fig. 2). In the human collagen X gene a 34 bp sequence containing the cleavage/polyadenylation signal is repeated 60 bp further downstream, suggesting that there may be two sites for cleavage/polyadenylation in the human collagen X mRNA (Fig. 2).

Chromosomal localization

A panel of human $\times$ rodent somatic hybrids, whose chromosome content had been previously established (see the Methods section) were analysed with the 750 bp DNA fragment containing 361 bp of the C-terminal non-collagenous domain and 349 bp of 3'-end untranslated region (Fig. 1) by Southern-blot hybridization. The co-segregation frequency of the collagen X gene locus with the different somatic-cell hybrids was determined and the percentage discordancy calculated (results not shown). The lowest percentage discordancy (0%) was for chromosome 6. To

localize the position of the collagen X gene on chromosome 6 the DNA probe was subsequently hybridized to a panel of somatic-cell hybrids whose human chromosome content consisted solely of fragments of chromosome 6 (Fig. 4a). The hybrid clone 2D, which only contains chromosome X, was included as a control because chromosome X is also present in the hybrid MCP 6. Analysis of the Southern-blot hybridization revealed that the collagen X gene is localized to the distal end of the long arm of chromosome 6 in the region 6q21–6q22.3 (Fig. 4a).

Regional assignment of the human collagen X gene by hybridization *in situ*. In the experiment with 0.02 μ g of probe DNA/ml, 50 metaphase spreads were examined and 213 autoradiographic grains were scored on chromosomes. Of these, 22 grains (10.33%) were localized on chromosome band 6q21–22.3. In the experiment using 0.004 μ g/ml probe DNA, 50 metaphase spreads were examined and 220 autoradiographic grains were scored on chromosomes. Of these, 24 (10.9%) were localized on chromosome band 6q21–6q22.3 (Fig. 4b). Grain counts were not significantly above average on any other chromosome or any other region of chromosome 6.

The specific location of the human collagen X gene at 6q21–6q22.3 does not coincide with the location of any previously identified disease phenotypes (O'Brien, 1990), although a number of genes have been mapped to this region (O'Brien, 1990; Ziegler *et al.*, 1990). It is of interest that two other cartilage collagen genes have been mapped to other regions of chromosome 6. $\alpha 1(IX)$ is located on the long arm at the q12–q14 locus (Kimura *et al.*, 1989) and $\alpha 2(XI)$ has been assigned to the short arm (Law *et al.*, 1990).

We are currently screening for the 5'-end of the human collagen X gene including the putative exon 1 and promoter elements. From sequence analyses to date, this exon and the control elements are in excess of 580 bp upstream of the 169 bp exon described here.

Although a number of biochemical studies have alluded to a possible role for collagen X in endochondral bone formation (Kwan *et al.*, 1989; Thomas *et al.*, 1990), its function is still unknown. The availability of the human $\alpha 1(X)$ sequence now permits studies on collagen X expression in a number of human diseases. Preliminary results obtained from human osteoarthritic cartilage have demonstrated by hybridization *in situ* the expression of collagen X mRNA at sites where there appears to be a re-initiation of the endochondral bone formation process, including osteophytes and subchondral bone sclerosis (Hoyland *et al.*, 1991). One would predict that in heritable diseases affecting endochondral bone formation, such as the chondrodysplasias, collagen X would be an obvious candidate gene.

The support of the Nuffield Foundation (Oliver Bird Trust), the Medical Research Council and the Arthritis and Rheumatism Council is gratefully acknowledged.

REFERENCES

- Edwards, Y. H., Parker, M., Povey, S., West, L. F., Farrington, F. M. & Solomon, E. (1985) *Ann. Hum. Genet.* **49**, 101–109
- Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13
- Gross, S. J. & Harris, H. (1975) *Nature (London)* **255**, 680–683
- Harper, M. E. & Saunders, G. F. (1981) *Chromosoma* **83**, 431–439
- Hoyland, J., Thomas, J. T., Marriott, A., Ayad, S., Boot-Handford, R. P., Grant, M. E. & Freemont, A. J. (1991) *Bone Miner.*, in the press
- Jones, E. A., Goodfellow, P. N., Kennet, R. H. & Bodmer, W. F. (1976) *Somatic Cell Mol. Genet.* **2**, 483–496
- Kietly, C. M., Povey, S. & Hopkinson, D. A. (1982) *Ann. Hum. Genet.* **46**, 307–327
- Kimura, T., Mattei, M.-G., Stevens, J. W., Goldring, M. B., Ninomiya, Y. & Olsen, B. R. (1989) *Eur. J. Biochem.* **179**, 71–78

- Kwan, A. P. L., Freemont, A. J. & Grant, M. E. (1986) *Biosci. Rep.* **6**, 155–162
- Kwan, A. P. L., Dickson, I. R., Freemont, A. J. & Grant, M. E. (1989) *J. Cell Biol.* **109**, 1849–1856
- Kwan, A. P. L., Cummings, C. E., Chapman, J. A. & Grant, M. E. (1991) *J. Cell Biol.* **114**, 597–604
- Law, M. L., Chan, S. D. H., Berger, R., Jones, C., Kao, F. T., Solomon, E. & Cheah, K. S. E. (1990) *Ann. Hum. Genet.* **54**, 23–29
- LuValle, P., Ninomiya, Y., Rosenblum, N. D. & Olsen, B. R. (1988) *J. Biol. Chem.* **263**, 18378–18385
- Marriott, A., Ayad, S. & Grant, M. E. (1991) *J. Cell Sci.* **99**, 641–649
- Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276
- Ninomiya, Y., Showalter, A. M., van der Rest, M., Seidah, N. G., Chretien, M. & Olsen, B. R. (1984) *Biochemistry* **23**, 617–624
- O'Brien, S. (1990) *Genetic Maps*, Book 5, Cold Spring Harbor Laboratory, Cold Spring Harbor
- Phillips, J. R., Shephard, E. A., Povey, S., Davies, M. B., Kelsey, G., Monteiro, M. & West, L. F. (1985) *Ann. Hum. Genet.* **49**, 267–274
- Povey, S., Jeremiah, S. J., Barker, R. F., Hopkinson, D. A., Robson, E. B., Cook, P. J. L. & Solomon, E. (1980) *Ann. Hum. Genet.* **43**, 241–248
- Reed, K. C. & Mann, D. A. (1985) *Nucleic Acids Res.* **13**, 7207–7221
- Remington, M. C., Bashey, R. I., Brighton, C. T. & Jimenez, S. A. (1983) *Collagen Relat. Res.* **3**, 271–278
- 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
- Solomon, E., Hiorns, L. R., Spurr, N., Kurkinen, M., Barlow, D., Hogan, B. L. M. & Dagleish, R. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3330–3334
- Thomas, J. T., Boot-Handford, R. P. & Grant, M. E. (1990) *J. Cell Sci.* **95**, 639–648
- Thomas, J. T., Kwan, A. P. L., Grant, M. E. & Boot-Handford, R. P. (1991) *Biochem. J.* **273**, 141–148
- Vieira, J. & Messing, J. (1987) *Methods Enzymol.* **153**, 3–11
- Watt, J. L. & Stephens, G. S. (1986) in *Human Cytogenetics: A Practical Approach* (Rooney, D. E. & Czepulowski, B. H., eds.), pp. 39–55, IRL Press, Oxford
- Willson, K., Kelly, A., Dudley, K., Goodfellow, P., Spurr, N., Groves, V., Gormon, P., Sheer, D. & Trowsdale, J. (1987) *EMBO J.* **6**, 1967–1974
- Wolff, S. & Perry, P. (1974) *Chromosoma* **48**, 341–353
- Wong, Z., Wilson, V., Patel, I., Povey, S. & Jeffreys, A. G. (1987) *Ann. Hum. Genet.* **51**, 269–288
- Yamaguchi, N., Benya, P. D., van der Rest, M. & Ninomiya, Y. (1989) *J. Biol. Chem.* **264**, 16022–16029
- Yamaguchi, N., Mayne, R. & Ninomiya, Y. (1991) *J. Biol. Chem.* **266**, 4508–4513
- Zabel, B. U., Naylor, S. L., Sakaguchi, A. Y., Bell, G. I. & Shows, T. B. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6932–6936
- Ziegler, A., Field, L. L. & Sakaguchi, A. Y. (1990) *Cytogenet. Cell Genet.* **55**, 118–121

Received 22 April 1991/3 June 1991; accepted 10 June 1991