THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1988 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 263, No. 32, Issue of November 15, pp. 17159-17166, 1988
Printed in U.S.A.

Cloning and Sequencing of Pro-α1(XI) Collagen cDNA Demonstrates That Type XI Belongs to the Fibrillar Class of Collagens and Reveals That the Expression of the Gene Is Not Restricted to Cartilagenous Tissue*

(Received for publication, February 26, 1988)

Michael Bernard‡, Hidekatsu Yoshioka‡, Elida Rodriguez§, Michel van der Rest§¶, Tomoatsu Kimura‖, Yoshifumi Ninomiya‖, Bjorn R. Olsen‖, and Francesco Ramirez‡

From the ‡Department of Microbiology and Immunology, Morse Institute of Molecular Genetics, State University of New York, Health Science Center at Brooklyn, Brooklyn, New York, 11203, the §Genetics Unit, Shriners Hospital, Montreal, Quebec, Canada H3G1A6, and the ‖Department of Anatomy and Cellular Biology, Harvard Medical School, Boston, Massachusetts 02115

We have isolated several overlapping cDNA clones encoding α1(XI) collagen chains from human and rat cDNA libraries. Together the human cDNAs code for 335 uninterrupted Gly-X-Y triplets, and a 264-amino acid C-propeptide, while the rat cDNAs cover the entire C-propeptide and about a third of the triple-helical domain. Comparison of the human and rodent nucleotide sequences showed a 95% sequence similarity. The identification of the clones as α1(XI) cDNAs was based on the complete identity between the amino acid sequences of three human α1(XI) cyanogen bromide peptides and the cDNA-derived sequence. Examination of the cDNA-derived amino acid sequence showed a variety of structural features characteristic of fibrillar-forming collagens. In addition, nucleotide sequence analysis of a selected portion of the corresponding human gene revealed the characteristic 54-base pair exon motif. We conclude therefore that pro-α1(XI) collagen belongs to the group of fibrillar collagen genes. We also suggest that the expression of this gene is not restricted to cartilage, as previously thought, since the cDNA libraries from which the clones were isolated, originated from both cartilagenous and noncartilaginous tissues.

Burgeson and Hollister (1979) first reported the identification of a minor collagen component in cartilage which upon SDS-PAGE[1] fractionation yielded three distinct chains, provisionally named 1α, 2α, and 3α. More recent data suggest that the three chains are assembled into one parental triple-helical molecule, called Type XI collagen (Morris and Bachinger, 1987). The 1α, 2α, and 3α chains have therefore been re-named α1(XI), α2(XI), and α3(XI), respectively (Eyre and Wu, 1987). The genetic identity of the α3(XI) chain, however, remains controversial, because on peptide mapping this chain exhibits a pattern similar, if not identical, to that of the α1 chain of Type II collagen (Burgeson *et al.*, 1982). The α1(XI) and α2(XI) chains, on the other hand, are distinct gene products which are closely related to the α1(V) and α2(V) chains, respectively (Burgeson *et al.*, 1982; Eyre and Wu, 1987).

Functionally, Type XI collagen has been implicated in regulation of the diameter of Type II collagen fibrils and in collagen:proteoglycan interactions (Sheren *et al.*, 1986; Smith *et al.*, 1985). These and other possible functions may be limited to the pericellular region of chondrocytes, albeit such a conclusion rests on immunological studies possibly hampered by the limited epitope accessibility in the matrix, as observed for Type V collagen (Fitch *et al.*, 1984; Ricard-Blum *et al.*, 1982).

The concomitant expression of Types II and XI in all cartilage tissues closely resembles that of Types I and V in non-cartilage matrices (Eyre and Wu, 1987). However, unlike Type XI, which has never been detected in a tissue lacking Type II, Type V collagen is not exclusively associated with Type I. There are in fact, at least two cases in which Type V collagen is found associated with Type II rather than Type I (Eyre and Wu, 1987). Types V and XI are therefore believed to be closely related collagens not only structurally but also functionally, in that both are minor matrix components exhibiting functions that are ancillary to those of the major tissue collagens, Types I and II (Eyre and Wu, 1987).

The isolation of pro-α1(XI) clones from two mammalian species has now enabled us to obtain novel and important information regarding the structure, evolution, and expression of Type XI collagen. In addition to providing the amino acid sequence of the C-propeptide and most of the human α1(XI) chain, our studies confirm the fibrillar nature of Type XI collagen and suggest that the expression of COL11A1 is not restricted to cartilage. Finally, the availability of Type XI probes should help ascertaining the potential role of this collagen in human cartilage disorders, such as chondrodystrophies and osteoarthritis.

MATERIALS AND METHODS

cDNA and Genomic Clones—Three human and one rodent cDNA libraries were used in these studies. One of the human libraries (a

* This work was supported by National Institutes of Health Grants AR-38648, R01 HD 22657, AR 36819, and AR 36820, March of Dimes, Birth Defects Foundation Grant 1-1042, Canadian MRC Grant MT-7796, and the Shriners of North America. 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™/EMBL Data Bank with accession number(s) J04177.

¶ Senior scholar from the Fonds de la Recherche en Santé du Quebec.

[1] The abbreviations used are: SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; bp, base pairs; C-propeptide, carboxyl-terminal propeptide; C-protease, the enzyme that specifically cleaves the carboxyl-terminal propeptide; COL1A1, COL1A2, COL2A1, COL3A1, COL5A2 and COL11A1, pro-α1(I), pro-α2(I), pro-α1(II), pro-α1(III), pro-α2(V), and pro-α1(XI) collagen genes, respectively; HPLC, high performance liquid chromatography.

generous gift of Drs. P. Berg and H. Okayama, Stanford University) was generated from RNA purified from an SV40-transformed fibroblast cell line (GM 637) (Okayama, and Berg, 1983). The second was constructed in λgt11, using as template poly(A$^+$) RNA from short term cultured costal chondrocytes as described.[2] The third was a λgt11 placental tissue cDNA library purchased from Clontech (Palo Alto, CA). The RNA from a Swarm rat chondrosarcoma was the source of the rodent cDNA library which was constructed in the λgt11 vector as described.[2]

The positive-negative screening protocol which led to the identification of the first Type XI clone, OK4, from the GM 637 cDNA library has been detailed elsewhere (Weil *et al.*, 1987). It should be noted that, lacking conclusive evidence, in that report we referred to OK4 as a Type V-related cDNA. The screenings of the cDNA libraries, as well as the processing and analysis of the positive clones were essentially as described (Weil *et al.*, 1987).[2]

The genomic library was constructed in the Charon 4A vector and it contained human DNA partially digested with *Eco*RI (Ramirez *et al.*, 1979). Screening, isolation, and purification of the phage clones were performed as described previously (Chu *et al.*, 1984).

Nucleotide Sequence Analysis—DNA sequencing was carried out with the dideoxy chain termination technique (Sanger *et al.*, 1977) either on single stranded DNA in M13 vectors using a standard protocol (Vieira and Messing, 1982), or on double stranded DNA in pUC vectors following modifications of the procedures described by Hattori and Sakaki (1986) and Zagursky *et al.* (1986). Sequencing of both strands was performed for the majority of the sequences presented here.

Collagen Purification—Cartilage collagens were prepared by a modification of the method of Burgeson and Hollister (1979). Human articular cartilages were obtained at autopsy from infants (<4 years of age) who died from disorders unrelated to connective tissues. The insoluble residues, left after a 4 M guanidine hydrochloride extraction of the proteoglycans (Roughley and White, 1980), were powderized at liquid nitrogen temperature in a freezer-mill (SPEX Industries, Metuchen, NJ). The powder was resuspended (300 mg/ml) in cold 4 M guanidine hydrochloride, 50 mM Tris-HCl, pH 7.4, and extracted for 48 h at 4 °C under constant stirring. The insoluble residue was collected by centrifugation (12,000 × *g*, 10 min), washed four times with cold distilled water, and resuspended (150 mg/ml) in a cold 0.5 M acetic acid, 0.2 M NaCl solution, and treated with pepsin (1 mg/ml) for 72 h. The pH was then raised to 8.1 with NaOH to denature pepsin. The viscous solution was clarified by centrifugation (24,000 × *g*, 30 min), and the supernatant was dialyzed for 72 h against three changes of a 0.5 M acetic acid, 0.9 M NaCl solution. The precipitated Type II collagen was pelleted by centrifugation (24,000 × *g*, 30 min). The supernatant was filtered through Whatman No. 1 paper and dialyzed for 72 h against three changes of a 0.5 M acetic acid, 1.2 M NaCl solution. The precipitated Type XI collagen was collected by centrifugation as above. The same procedure was further applied to precipitate the Type IX collagen at 2.0 M NaCl in 0.5 M acetic acid. The purified collagens were dialyzed against 0.5 M acetic acid and lyophilized.

Separation of the Type XI Chains by CM-cellulose Chromatography—Type XI collagen (8 mg) was dissolved in 8 ml of 0.04 M sodium acetate, pH 4.8, buffer containing 6 M urea, denatured by heating at 42 °C for 10 min, and loaded on a 1.6 × 14-cm CM-cellulose (Whatman CM52) column, heated to 42°C, and equilibrated with the same buffer. Elution was performed with a NaCl gradient (0–0.09 M, 500 ml) in the same buffer and monitored with a Schoeffel 770 monitor set at 230 nm.

Gel Electrophoresis—SDS-PAGE was performed in 7 cm × 8 cm × 0.75-mm (length × width × depth) gels in a mini Protean II cell (Bio-Rad) according to Laemmli (1970).

High Performance Liquid Chromatography (HPLC)—Separation of CNBr-derived peptides was done as described previously (van der Rest and Fietzek, 1982) on a Vydac TP 201 (4.6 × 250 mm) reversed-phase column (The Separation Group). The initial separation was done using 10 mM heptafluorobutyric acid as ion-pairing agent, and the peptides were further purified by a second chromatography step with 9 mM trifluoroacetic acid. The equipment used was from Beckman and consisted of a Model 334 chromatograph, a Model 160 UV monitor equipped with a zinc lamp, and a CR-1B data system.

Amino Acid Sequence Analysis—Amino acid sequences were determined by automated Edman degradation in an Applied Biosystems Model 470A sequenator, using the trifluoroacetic acid conversion program. The phenylthiohydantoin derivatives of the amino acids were identified by HPLC as described by Lazure *et al.* (1983).

Cleavage with Cyanogen Bromide—The purified chains were dissolved in 70% formic acid containing 12 mg/ml CNBr, and the solutions were flushed with nitrogen. After a 16-h digestion at room temperature, the solution was diluted four times with water and freeze-dried.

RESULTS

Isolation and Characterization of Type V-related cDNA Clones—We have already documented the isolation of a distinct collagen-coding recombinant, OK4, from the GM 637 library (Weil *et al.*, 1987). The insert of this clone is nearly 4.2 kilobases in length and contains a 579-nucleotide long 3′-untranslated region which includes a short poly(A) tail, a 792-bp portion coding for a cysteine-containing globular domain, and a 2812-bp portion coding for an uninterrupted collagenous sequence (Fig. 1). Based on the similarity between the amino acid composition of OK4 and that of the α1(V) chain, we provisionally referred to this clone as coding for a Type V-related gene product (Weil *et al.*, 1987). In order to obtain further 5′ sequences, an appropriate subfragment of OK4 was utilized to screen 5 × 10^4 phage plaques of the λgt11 human placental tissue cDNA library. This resulted in the identification of two positive phages, of which one, Hpl 1, was found to extend an additional 202 bp beyond the 5′ end of OK4 (Fig. 1). At the same time, and independently of these experiments, some of us (T. K., Y. N., and B. R. O.) isolated from human and rat cartilage-specific libraries several clones whose sequences were found to be identical and homologous to OK4, respectively. The restriction map of the largest human chondrocyte clone, KTh 98, is shown in Fig. 1, whereas Fig. 2 shows a compositus of the nucleotide sequence and conceptual amino acid translation of the three overlapping human clones. Fig. 3 shows a comparison between the common coding sequences of the human and the largest rat clone, pKT 1201. This analysis showed a level of amino acid divergence of only 6% in the C-propeptide and 5% in the triple helical domain.

Collectively, these data suggested that the Type V-related gene product encoded by OK4 is expressed in at least four different tissues, of which two are cartilageneous. This raised the strong possibility that the cDNAs may code for one of the Type XI collagen chains. Hence, biochemical analyses were performed to prove the validity of this hypothesis.

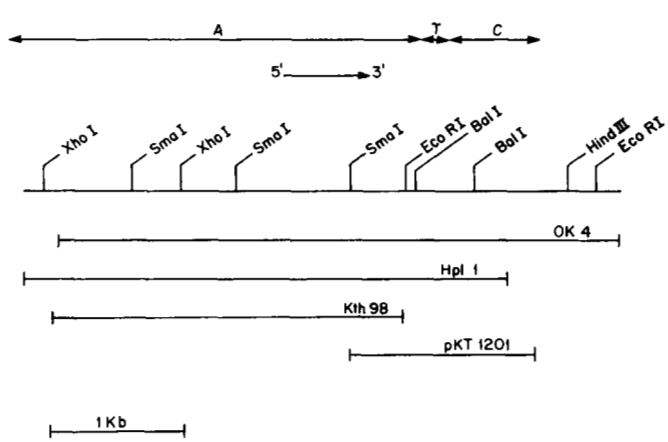

FIG. 1. **Restriction endonuclease map of the mammalian COL11A1 cDNA clones.** On *top* is the relative location of the three protein domains encoded by the cDNAs, namely C-propeptide (*C*), telopeptide (*T*), and αchain (*A*). *Underneath* is the restriction map of the human clones, whose relative location is depicted along with that of the rat cDNA, pKT 1201. *kb*, kilobase.

[2] T. Kimura, M. G. Mattei, J. W. Stevens, M. B. Goldring, Y. Ninomiya, and B. R. Olsen, manuscript submitted for publication.

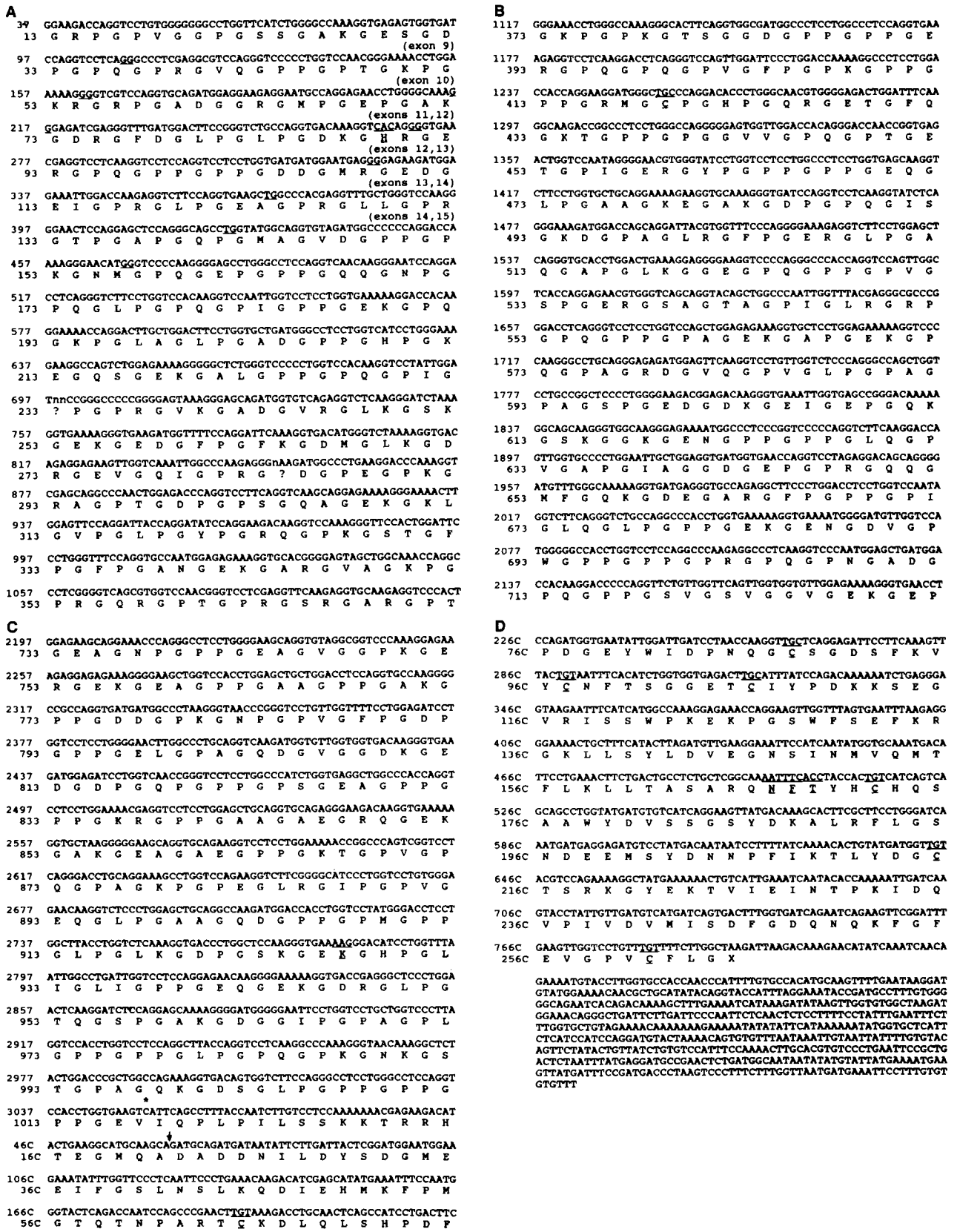

FIG. 2. **Nucleotide and amino acid sequences of the human COL11A1.** In the *first line*, on the *left side* of the sequence, is the *numbering* of the nucleotides starting from the beginning of the α-chain, while the *numbers* in the *second line* refer to the encoded amino acid residues. Note that in the C-propeptide the numbering re-initiates and it is followed by the letter *C*. The demarcations of the sequenced exons are indicated by the *underlined* dinucleotides, and the exons are numbered 5′ to 3′ on the *right side* of the sequence. *Underlined* are also the putative cross-linking sites, the C-propeptide and α-chain cystenyl residues, and the C-propeptide carbohydrate attachment site. The *arrow* signifies the putative C-protease cleavage site, while the *asterisk* indicates the end of the α-chain.

FIG. 3. **Comparison between the human and rat sequences.** Numbering of the human sequence is as in Fig. 2. Only the nucleotide (*above*) and amino acid (*below*) differences with the rat sequence are indicated. The *asterisk* signifies the end of the α-chain, while the *arrow* indicates the putative C-protease cleavage site.

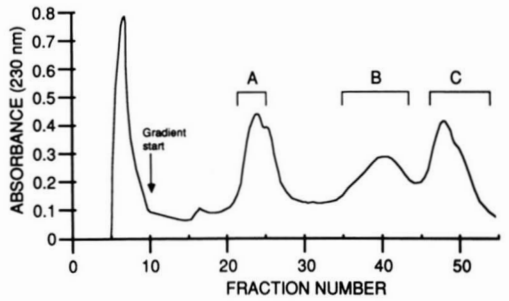

FIG. 4. **CM-cellulose chromatography of Type XI heat-denatured chains.** *Brackets* indicate the fractions pooled for further analysis (*A*, *B*, and *C*). The start of the NaCl gradient is signified by the *arrow*.

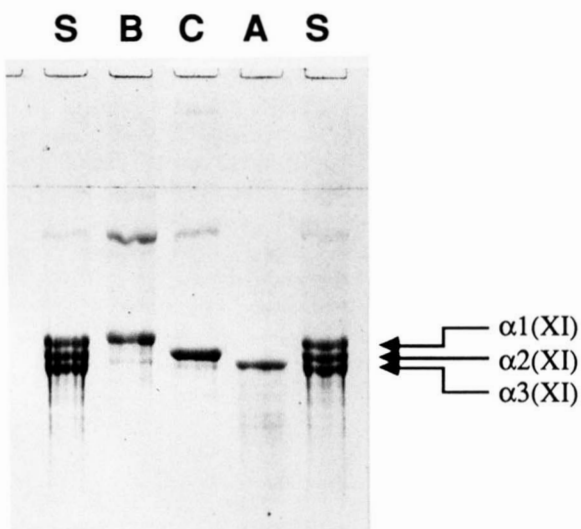

FIG. 5. **6% SDS-PAGE of the Type XI chains.** The purified Type XI collagen is in the *lanes* labeled *S* whereas the separated chains, pooled after CM-cellulose chromatography, are labeled as in Fig. 4 as *A*, *B*, and *C*.

Biochemical Analyses of the Human α1(XI) Collagen Chain—Identification of the collagen precipitating at 1.2 M NaCl as Type XI collagen was accomplished by gel electrophoresis, in which the α1(XI), α2(XI), and α3(XI) chains appeared as bands of approximately equal intensities (Fig. 5), and by the chromatographic elution of the chains on a CM-cellulose column (Fig. 4). The separated chains were further characterized by amino acid analyses (data not shown), and electrophoresis of their CNBr derived peptides (Fig. 6). By all these criteria, our preparation was identical to the one originally described by Burgeson and Hollister (1979). The CNBr-derived fragments of the α1(XI) chain were then separated by HPLC (Fig. 7). All peptide-containing fractions were analyzed by gel electrophoresis and amino acid analysis. Two fractions containing short fragments undetectable in 15% acrylamide gels (*A* and *B* in Fig. 7), and one fraction containing a 35-kDa peptide (*C* in Fig. 7) were purified by a second HPLC run. The fractions indicated by *arrows* in Fig. 8 were then selected for sequencing. The amino acid sequences of peptides A, B, and C are shown in Fig. 9 together with the translated amino acid sequence of the cDNA clones. This comparison conclusively proved that the clones code for the pro-α1 chain of Type XI collagen. Incidentally, similar analyses using α1(V) CNBr peptides and tryptic peptides encompassing a total of 136 amino acid residues revealed a 86% sequence similarity with the α1(XI) chain (data not shown).

Genomic Cloning—One of the most distinctive features of

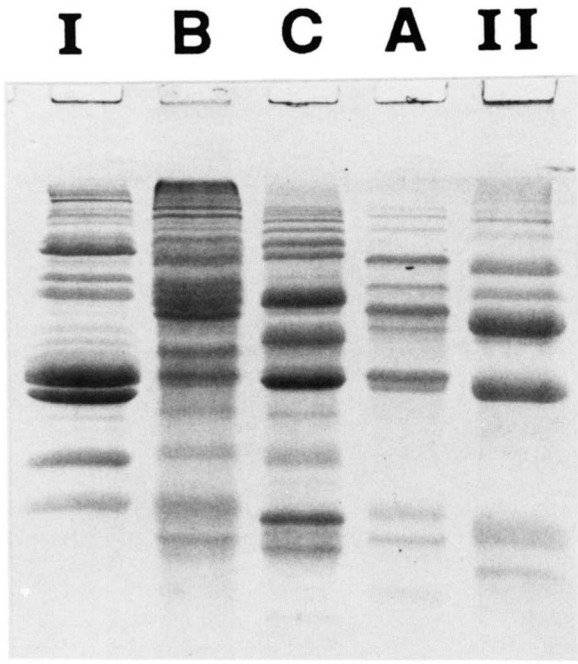

FIG. 6. **15% SDS-PAGE of the CNBr derived peptides of the purified Type XI chains.** *Lane I*, Type I collagen $\{\alpha1(I)\}_2\,\alpha2(I)$; *lane B*, α1(XI); *lane C*, α2(XI); *lane A*, α3(XI); *lane II*, α1(II).

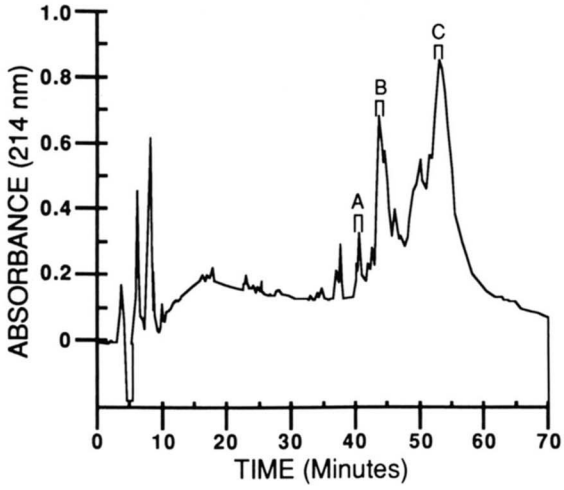

FIG. 7. **Reversed-phase HPLC of the CNBr-derived peptides from the α1(XI) chain.** Peptides derived from 1 mg of α1(XI) chain were eluted with an aqueous gradient of acetonitrile (12.4–44.4%) over 60 min, in the presence of 10 mM heptafluorobutyric acid. *Brackets* indicate the fractions that were used for further analysis (*A*, *B*, and *C*).

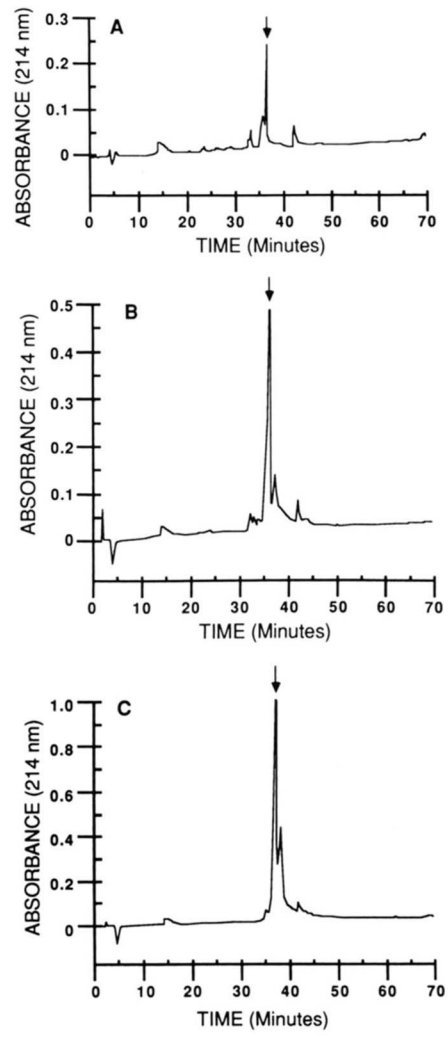

FIG. 8. **Second reversed-phase HPLC of the CNBr-derived peptides from the α1(XI) chains.** The fractions labeled *A*, *B*, and *C* in Fig. 7 were eluted with the same gradient as in Fig. 7, except that 9 mM trifluoroacetic acid was used. The *arrows* indicate the fractions that eventually were sequenced.

```
PEPTIDE A: RGEDGEIGPRGLPGEAGPRGLLGPRGXPGAPGQ
     cDNA: RGEDGEIGPRGLPGEAGPRGLLGPRGTPGAPGQ
           108                              140

PEPTIDE B: GPQGEPGPPGQQGNPGPQGLPGPQGPIGPPGEXGPQGKPG
     cDNA: GPQGEPGPPGQQGNPGPQGLPGPQGPIGPPGEKGPQGKPG
           157                                    196

PEPTIDE C: GLXGDRGEVGQIGPRG
     cDNA: GLKGDRGEVGQIGPRG
           268           283
```

FIG. 9. **Sequence comparison between three α1(XI) peptides and the translation of the human cDNAs.** Numbering of the amino acid residues is as in Figs. 2 and 3. Undetermined amino acids are labeled *X*. Note that prolines in the third, *Y*, position are hydroxylated.

the collagen genes is the pattern of distribution of the exons coding for the triple helical domains (for a review see Ramirez *et al.*, 1985). This feature can in fact be used to divide the relatively large group of collagen genes into distinct subfamilies with identical exon/intron arrangements. Accordingly, characterization of the Type I, II, III, and V collagen genes has provided a basis for categorizing these collagens in a distinct subfamily, the fibrillar collagens (Ramirez *et al.*, 1985). To examine whether also COL11A1 belongs to this group, we isolated several genomic clones from a human library and analyzed the structure of one of them, Pen 3. Because of the size and complexity of COL11A1, rather than detailing the organization of the entire clone, we decided to characterize a selected portion by shot-gun subcloning and sequencing. To this end, the *Eco*RI-digested DNA of Pen 3 was hybridized to the 5′-most *Xho*I subfragment of OK4 (Fig. 1). The positive 5.5-kilobase *Eco*RI fragment of Pen 3 was inserted into pUC18, and in turn used to generate a collection of *Sau*3A subclones which were then screened with the original *Xho*I probe of OK4. Twenty-four of the positively hybridizing *Sau*3A subclones were randomly chosen and sequenced. This led to the identification of seven triple-helical domain-coding exons, whose relative location with respect to the cDNA sequence is shown in Fig. 2. Several points should be noted. Firstly, the exons cover a continuous stretch of 360

TABLE I

Sequence divergence between fibrillar collagen genes and COL1A1

	Amino acids	Nucleotides
	%	
COL2A1	33	28
COL1A2	39	35
COL3A1	40	40
COL5A2	48	42
COL11A1	62	52

TABLE II

Codon usage in the triple helical domains of the fibrillar collagen genes

Computations of the data were based on the entire sequence of the triple-helical domains of COL1A1 (Bernard *et al.*, 1983a; D'Alessio *et al.*, 1988), COL1A2 (Bernard *et al.*, 1983b; de Wet *et al.*, 1987), COL5A2 (Weil *et al.*, 1987; D. Woodbury, V. Benson-Chanda, and F. Ramirez, manuscript in preparation), COL3A1 (Chu *et al.*, 1985; R. Janeczko and F. Ramirez, manuscript in preparation), and COL2A1 (Sangiorgi *et al.*, 1985; M. W. Su, V. Benson-Chanda, B. Lee, H. Vissing, and F. Ramirez, manuscript in preparation).

Amino acid	Third base	COL-1A1	COL-2A1	COL-1A2	COL-3A1	COL-5A2	COL-11A1
		%					
Glycine	T	52	43	54	44	37	45
	C	27	28	20	12	16	13
	A	18	24	22	37	34	28
	G	3	5	4	7	13	14
Proline	T	59	61	63	55	51	51
	C	38	28	21	13	9	7
	A	3	11	15	31	33	39
	G	0	0	1	1	7	3
Alanine	T	75	59	77	61	50	44
	C	20	29	15	14	6	11
	A	5	12	8	23	33	45
	G	0	0	0	2	11	0

nucleotides with a base pair size pattern, 54-54-54-54-45-54-45, clearly indicative of a fibrillar collagen motif. Furthermore, this particular pattern is found only in one specific region of the Type I genes, namely between exons 9 and 15 (as counted from the 5′ to 3′ end) (de Wet *et al.*, 1987; D'Alessio *et al.*, 1988). These exons code for amino acid residues 37–156. Within this segment of the triple helical domain lies a cross-linking site, whose sequence, Gly-*X*-Hyl-Gly-His-Arg, and relative location, between residues 85 and 90, is highly conserved among the fibrillar collagen molecules. Moreover, this canonical element is the last hexapeptide encoded by exon 11, which is 54 bp in size (de Wet *et al.*, 1987; D'Alessio *et al.*, 1988). Inspection of the translation product of the α1(XI) sequence identified the hexapeptide Gly-Asp-Lys-Gly-His-Arg within the last 18 bp of a 54-bp exon (Fig. 2). Hence, by analogy to other fibrillar collagen genes, this exon was numbered as 11 (Fig. 2). This conclusion enabled us to extrapolate that our sequence begins at the 13th amino acid residue of the triple helical domain and that the α1(XI) chain is 1017 residues long (Fig. 2).

Sequence Comparison between the Pro-α1(XI) Chain and the Other Fibrillar Collagens—Having categorized the pro-α1(XI) chain as a member of the fibrillar-forming collagens, we then attempted to more firmly place this gene within the group. To this end, we analyzed the human C-propeptide and triple-helical domain sequences. Of these two segments, the former, which lacks the repetitive Gly-*X*-*Y* triplet structure, has been used for ascertaining the level of sequence similarity between each of the collagen genes and COL1A1. On the other hand, the abundance of glycine, proline, and alanine residues in the triple-helical domain has served as a basis for establishing the patterns of codon usage, an indicator of kinship among related genes.

Although pairwise comparison of the C-propeptide sequences showed minimal degree of homology, with the exception of few nucleotides around the conserved cysteinyl residues (for a review, see Dion and Myers, 1987), it enabled us to position COL1A11 closer to COL5A2 and at the highest point of divergence from COL1A1 (Table I).

The analysis of the wobble base conservation in the triple-helical domain codons established a kinship among COL3A1, COL5A2, and COL11A1. This conclusion rests on the finding that in COL1A1, COL1A2, and COL2A1 there is a strong preference in the wobble position first for T and secondly for C, while in COL3A1, COL5A2, and COL11A1 the first preference is for T, and secondly for A (Table II).

DISCUSSION

Identification of the Pro-α1(XI) Collagen cDNA Clones—The isolation of several cDNA clones from two mammalian species provides the first detailed study of the primary structure of a Type XI collagen chain. The overall length of the human cDNAs is nearly 5 kilobases, and, in addition to the 3′-untranslated region, it comprises the entire C-terminal propeptide and, in our estimate, 99% of the triple-helical domain. The identity of the recombinant molecules was established by direct comparison with the amino acid sequence of three α1(XI) CNBr-derived peptides determined by Edman degradation (Fig. 9). Surprisingly, our experiments indicate that COL11A1 expression, albeit at extremely low levels, is not confined to cartilagenous tissues. It remains to be demonstrated whether such a finding is indicative of post-transcriptional controls that restrict Type XI biosynthesis in some non-cartilagenous tissue, such as placenta, or whether, given the low representation of the mRNA, the protein is undetectable by conventional methods. Notwithstanding this problem, these studies have generated a number of important pieces of information related to the structure and evolution of two of the major domains of the procollagen molecule, the C-propeptide and the triple helix.

C-propeptide—It is now widely accepted that intracellular folding of the procollagen chains into the triple helix begins at the C-propeptides with the formation of inter- and intrachain disulfide bonds (Uitto and Prockop, 1974; Rosenbloom *et al.*, 1976). Less understood is how this process is selectively guided in those instances where either several different collagen types are co-expressed in the same tissue, or the same type displays a molecular heterogeneity in different tissues and/or at different times of development (Miller and Gay, 1987). With the increasing acquisition of structural information, several investigators have begun to analyze the data in search of evolutionary conserved features that may shed new light on this problem (Showalter *et al.*, 1980; Fuller and Boedtker 1981; Bernard *et al.*, 1983a, 1983b; Yamada *et al.*, 1983; Ninomiya *et al.*, 1984; Sandell *et al.*, 1984; Sangiorgi *et al.*, 1985; Myers *et al.*, 1985; Chu *et al.*, 1985). One such feature is the number and location of the cysteinyl residues in the C-propeptides of the fibrillar collagens (Dion and Myers, 1987). Two distinct patterns emerge. In polypeptides that form homotrimers (α1(II), α1(III)) or both homo- and heterotrimers (α1(I)), 8 cysteinyl residues are present, whereas in procollagen chains that only form part of heterotrimers (α2(I) and α2(V)), 7 of these residues are found. Based on studies of the avian α1(I) and α1(II) collagens, cysteines 5–8 (numbering from 5′ to 3′) are involved in intrachain linkages and are invariable in number (Olsen, 1982). In point of fact, in a naturally occurring variant a mutation affecting the last of

the α2(I) cysteinyl residues prevents heterotrimer assembly (Pihlajaniemi *et al.*, 1984). On the other hand, the absence of either cysteine number 2 (α2(I)) or number 3 (α2(V)) may be a consequence of chain-specific requirements for those chains that associate exclusively in heterotrimeric forms. This idea seems to be further supported by the finding that in both the human and rat pro-α1(XI) cysteine number 2 is missing, and that, like in α2(I) and α2(V), this is due to a single nucleotide change from a cysteine to a serine codon (Fig. 2). It would be of interest to determine whether the absence of a potential interchain disulfide linkage is indeed a feature of all heterotrimeric collagen chains (α2(XI) and α3(V)), and conversely whether the presence of all 4 cysteines (1 to 4) is always related to homotrimer-forming chains (α1(V)). The idea is also consistent with the hypothesis that the α3(XI) and α1(II) chains are products of the same gene, since α1(II), like α1(I), has 8 cysteinyl residues and is capable of both homo- and heterotrimeric assembly.

A second feature that places the pro-α1(XI) chain in the group of fibrillar collagens is the presence of a tripeptide (Asn-Phe-Thr), which is a potential acceptor for *N*-asparaginyl-linked carbohydrate attachment in the C-propeptide (Struck and Lennarz, 1980) (Fig. 2).

Finally, in the human sequence two possible C-proteinase cleavage sites were identified at C21-C22 and C23-C24 (Dion and Myers, 1987). Because the rat chain lacks the latter, we have assigned the cleavage site to the Ala-Asp sequence of the former (Figs. 2 and 3). Assuming that our premise is correct, the C-terminal telopeptide is then 21 residues long.

Helical Domain—The fibrillar nature of the pro-α1(XI) chain has been confirmed by the inspection of the triple-helical domain sequence, whose long uninterrupted motif of Gly-*X*-*Y* repeats is the most obvious evidence. In addition, the presence of the cross-linking sites further strengthens this argument. In the fibrillar collagens the lysines involved in these intermolecular cross-links are located at positions 87 and 930 of the triple helix and in the two telopeptide domains (Eyre *et al.*, 1984). In the α1(XI) sequence, three of the conserved lysines were identified. Two of them lie within the helix, one at position 87 and the other at 927. The latter represents a positional variation as compared to the majority of the other fibrillar chains. Such an exception is not the first one, as this cross-linking site is located at position 933 in the α2 chain of Type I collagen (Eyre *et al.*, 1984; de Wet *et al.*, 1987). The C-terminal telopeptide in all but two of the collagens (α2(I) and α2(V)) contains the forth cross-linking lysine, 9 residues upstream from the C-protease cleavage site (Eyre *et al.*, 1984; Weil *et al.*, 1987; de Wet *et al.*, 1987). In the α1(XI) two consecutive lysines are seen in both mammalian species, 11 and 12 amino acid residues upstream from the C-terminal end of the telopeptide. Whether these two lysyl residues do indeed participate in the cross-linking processes must await experimental confirmation.

Collagen content in extracellular matrices is regulated by a delicate balance between biosynthesis and degradation. Within the helical domains of Types I–III, at residue positions 775–776, one finds the vertebrate collagenase cleavage site (Kuhn, 1987). Type XI, which is resistant to vertebrate collagenase cleavage, lacks the collagenase cleavage sequence at positions 775–776. A similar finding was also reported for Type V collagen (Myers *et al.*, 1985), albeit at a closer inspection an alternative potential site was found between 667 and 668 in the α2(V) chain (Weil *et al.*, 1987). Similarly, in the α1(XI) a potential cleavage sequence is seen between residues 175 and 176. It is therefore conceivable that alternative cleavage sites may exist in Type V and XI collagens. This in turn, could reflect the existence of different collagenase specificities. Such a notion is supported not only by the isolation of a Type V-specific collagenase (Mainardi *et al.*, 1980), but also by the work of Fields *et al.* (1987) which showed other possible cleavage sequences for the enzyme, and emphasized the importance of secondary and tertiary structural considerations in addition to sequence specificity.

A final feature, which seems unique to the α1(XI) chain, is the presence of an unusual cysteinyl residue located approximately in the middle of the triple-helical domain. This residue is not utilized in the formation of an interchain disulfide bond, because reduction of the pepsin-treated molecule does not affect the migration of Type XI chains on SDS-PAGE (Burgeson and Hollister, 1979; Kielty *et al.*, 1984). Therefore, we speculate that this residue may be involved in bonding to other components in the extracellular matrix.

Type XI, a Fibrillar Collagen—Our structural findings clearly define Type XI as a fibrillar collagen. This conclusion confirms the previous classification of the molecule on the basis of protein data (Miller and Gay, 1987). Although limited to a very small portion of the gene, our characterization of COL11A1 strongly suggests that its organization is identical to that of other fibrillar collagen genes.

Analyses of the C-propeptide and helical domain sequences showed that Type XI exhibits a close kinship with Types III and V. The kinship with Type V was not surprising because of the already suggested structural and, possibly, functional similarities between the two minor collagens. In addition, and in line with what has been seen for most of the fibrillar collagen genes, the COL11A1 offers another example of chromosomal dispersion, for it resides on the short arm of chromosome 1 (Henry *et al.*, 1988).

Although the data presented here quite convincingly place Type XI among the fibrillar collagens, it will be important to further detail the organization of COL11A1 in order to identify possible deviations from the prototypical structure of a fibrillar collagen gene. Fusion of two triple-helical domain coding exons in COL1A1 and variations in the number of N-propeptide encoding exons in COL2A1 and COL3A1 have in fact been observed (Chu *et al.*, 1984, 1985).[3,4] These findings have been helpful in tracing the evolutionary history of the genes, as well as delineating some of the selective pressures exerted upon their products. Finally, and relevant to the surprising finding of pro-α1(XI) mRNA in cells of non-cartilagenous origin, the analysis of potential regulatory elements should enhance our understanding of Type XI function in physiological and pathological conditions.

Acknowledgments—We wish to thank Drs. Berg and Okayama for providing us with the cDNA library, Dr. B. Brodsky for many helpful discussions and suggestions during the course of this work, E. de Miguel and E. Wan for their skillful technical assistance, and Roseann Lingeza for preparing and typing the manuscript.

REFERENCES

Bernard, M., Chu, M. L., Myers, J. C., Ramirez, F., Eikenberry, E. F., and Prockop, D. J. (1983a) *Biochemistry* **22,** 5213–5223

Bernard, M., Myers, J. C., Chu, M. L., Ramirez, F., Eikenberry, E. F., and Prockop, D. J. (1983b) *Biochemistry* **22,** 1140–1145

Burgeson, R. E., and Hollister, D. (1979) *Biochem. Biophys. Res. Commun.* **87,** 1124–1131

Burgeson, R. E., Hebda, P., Morris, N., and Hollister, D. (1982) *J. Biol. Chem.* **257,** 7852–7856

Chu, M. L., de Wet, W., Bernard, M., Ding, J. F., Morabito, M.,

[3] V. Benson-Chanda, M. W. Su, D. Weil, M. L. Chu, and F. Ramirez, manuscript in preparation.

[4] M. W. Su, V. Benson-Chanda, B. Lee, H. Vissing, and F. Ramirez, manuscript in preparation.

Myers, J., Williams, C., and Ramirez, F. (1984) *Nature* **310,** 337–340
Chu, M. L., Weil, D., de Wet, W., Bernard, M., Sippola, M., and Ramirez, F. (1985) *J. Biol. Chem.* **260,** 4357–4363
D'Alessio, M., Bernard, M., Pretorius, P., de Wet, W., and Ramirez, F. (1988) *Gene (Amst.)* **67,** 105–115
de Wet, W., Bernard, M., Benson-Chanda, V., Chu, M. L., Dickson, L., Weil, D., and Ramirez, F. (1987) *J. Biol. Chem.* **262,** 16032–16036
Dion, A. S., and Myers, J. C. (1987) *J. Mol. Biol.* **193,** 127–143
Eyre, D. R., and Wu, J.-J. (1987) in *Structure and Function of Collagen Types* (Mayne, R., and Burgeson, R. E., eds) pp. 261–284, Academic Press, New York
Eyre, D. R., Pay, M. A., and Gallop, P. M. (1984) *Annu. Rev. Biochem.* **53,** 717–748
Fields, G., VanWart, H., and Birkeobal-Hansen, H. (1987) *J. Biol. Chem.* **262,** 6221–6226
Fitch, J. M., Gross, J., Mayne, R., Johnson-Wint, B., and Linsenmayer, T. F. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 2791–2795
Fuller, F., and Boedtker, H. (1981) *Biochemistry* **20,** 996–1006
Hattori, M., and Sakaki, Y. (1986) *Anal. Biochem.* **152,** 232–238
Henry, I., Bernheim, A., Bernard, M., van der Rest, M., Kimura, T., Jeanpierre, C., Barichard, F., Berger, R., Olsen, B. R., and Ramirez, F. (1988) *Genomics* **3,** 87–90
Kielty, C., Hulmes, C., Schor, S., and Grant, M. (1984) *FEBS Lett.* **169,** 179–184
Kuhn, K. (1987) in *Structure and Function of Collagen Types* (Mayne, R., and Burgeson, R. E., eds) pp. 1–42, Academic Press, New York
Laemmli, U. K. (1970) *Nature* **227,** 680–685
Lazure, C., Seidah, N. G., Chretien, M., Lallier, R., and St-Pierre, S. (1983) *Can. J. Biochem.* **61,** 287–292
Mainardi, C., Seyer, J., and Kang, A. (1980) *Biochem. Biophys. Res. Commun.* **87,** 1108–1112
Miller, E. J., and Gay, S. (1987) *Methods Enzymol.* **144,** 3–40
Morris, N. P., and Bachinger, H. P. (1987) *J. Biol. Chem.* **262,** 11345–11350
Myers, J. C., Loidl, H. R., Seyer, J. M., and Dion, A. S. (1985) *J. Biol. Chem.* **260,** 11216–11222
Ninomiya, Y., Showalter, A. M., van der Rest, M., Seidah, N. G., Chretien, M., and Olsen, B. R. (1984) *Biochemistry* **23,** 617–624
Okayama, H., and Berg, P. (1983) *Mol. Cell. Biol.* **3,** 280–289
Olsen, B. R. (1982) in *New Trends in Basement Membrane Research* (Kuhn, K., Shoene, H., and Timpl, R., eds) pp. 225–236, Raven Press, New York
Pihlajaniemi, T., Dickson, L., Pope, F. M., Korkonen, V. R., Nicholls, A., Prockop, D. J., and Myers, J. (1984) *J. Biol. Chem.* **259,** 12941–12944
Ramirez, F., Burns, A. L., Mears, J. G., Spence, S., Starkman, D., and Bank, A. (1979) *Nucleic Acids Res.* **7,** 1147–1162
Ramirez, F., Bernard, M., Chu, M. L., Dickson, L., Sangiorgi, F., Weil, D., de Wet, W., Junien, C., and Sobel, M. E. (1985) *Ann. N.Y. Acad. Sci.* **460,** 117–124
Ricard-Blum, S., Hartmann, D. J., Herbage, D., Payen-Meyran, C., and Villa, G. (1982) *FEBS Lett.* **146,** 343–347
Rosenbloom, J., Endo, R., and Harsch, M. (1976) *J. Biol. Chem.* **252,** 2070–2076
Roughley, P. J., and White, R. J. (1980) *J. Biol. Chem.* **255,** 217–224
Sandell, L. J., Prentice, H. L., Kravis, D., and Upholt, W. B. (1984) *J. Biol. Chem.* **259,** 2207–2225
Sanger, F., Nicklen, S., and Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
Sangiorgi, F., Benson-Chanda, V., de Wet, W., Sobel, M., Tsipouras, P., and Ramirez, F. (1985) *Nucleic Acids Res.* **13,** 2207–2225
Sheren, S., Eikenberry, E. F., Broek, D., van der Rest, M., Doehring, T., Kelly, J., Hardt, T., and Brodsky, B. (1986) *Comp. Biochem. Physiol.* **85,** 5–14
Showalter, A. M., Pesciotta, D. M., Eikenberry, E. F., Yamamoto, T., Pastan, I., de Crombrugghe, B., Fietzek, P. P., and Olsen, B. R. (1980) *FEBS Lett.* **111,** 61–65
Smith, G., Williams, J., and Brandt, K. (1985) *J. Biol. Chem.* **260,** 10761–10767
Struck, D. K., and Lennarz, W. J. (1980) in *The Biochemistry of Glycoproteins and Proteoglycans* (Lennarz, W. J., ed) pp. 35–83, Plenum Publishing Corp., New York
Uitto, J., and Prockop, D. J. (1974) *Biochemistry* **13,** 4586–4591
van der Rest, M., and Fietzek, P. P. (1982) *Eur. J. Biochem.* **125,** 491–496
Vieira, J., and Messing, J. (1982) *Gene (Amst.)* **19,** 259–268
Weil, D., Bernard, M., Gargano, S., and Ramirez, F. (1987) *Nucleic Acids Res.* **15,** 181–198
Yamada, Y., Kuhn, K., and de Crombrugghe, B. (1983) *Nucleic Acids Res.* **11,** 2733–2744
Zagursky, R. J., Berman, M. L., Baumeister, K., and Lomax, N. (1986) *Gene Anal. Tech.* **2,** 89–94