THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 265, No. 11, Issue of April 15, pp. 6423-6426, 1990
Printed in U.S.A.

Pro-α1(XI) Collagen

STRUCTURE OF THE AMINO-TERMINAL PROPEPTIDE AND EXPRESSION OF THE GENE IN TUMOR CELL LINES*

(Received for publication, November 21, 1989)

Hidekatsu Yoshioka‡ and Francesco Ramirez§

From the Department of Microbiology and Immunology, Morse Institute of Molecular Genetics, State University of New York, Health Science Center, Brooklyn, New York 11203

We have determined the nucleotide sequence of several overlapping cDNA clones encoding the amino-terminal portion of human α1(XI) procollagen. These experiments have revealed that this domain of the pro-α(XI) chain displays structural features common to other fibrillar procollagen molecules, such as a putative amino-terminal proteinase cleavage site and an interrupted collagenous segment. In the latter, structural similarities were noted when α1(XI) was compared with α1(II) and α2(V) procollagens. Overall, however, the amino-terminal region of pro-α1(XI) differs greatly in composition and size from that of other fibrillar chains. Nearly three-fourths of this domain is in fact composed of a 383-amino acid globular region in which a 3-cysteine cluster signals the transition to a long and highly acidic carboxyl-terminal segment. Finally, the unrestricted expression of this cartilage-specific collagen gene has been confirmed by the finding of high levels of pro-α1(XI) mRNA in two human rhabdomyosarcoma cell lines.

Type XI collagen is a minor extracellular component of hyaline cartilage (1). This macromolecule is composed of three distinct subunits, termed α1(XI), α2(XI), and α3(XI) collagens (2). The genes that code for the precursor procollagen form of α1(XI) and α2(XI) chains (COL11A1 and COL11A2)[1] have been recently cloned and localized on human chromosomes 1 and 6, respectively (3–5). In contrast, the α3(XI) chain is believed to be a post-translational variation of the major cartilage collagen, α1(II) (6). Aside from inferring some of the primary structure of pro-α1(XI) and pro-α2(XI) chains, the cloning experiments have also indicated that type XI collagen belongs to the fibril-forming group of molecules which, in addition, includes types I, II, III, and V (3, 5). Furthermore, these studies have revealed that, unlike COL11A2, COL11A1 is also expressed in noncartilagenous tissues, for pro-α1(XI) clones were isolated from cDNA libraries prepared from placental tissue and a SV40-transformed fibroblast cell line (3, 5).

Emerging biochemical evidence suggests that collagen XI may play an important role in fibrillogenesis by controlling lateral growth of collagen II fibrils (2, 7). In non-cartilage tissue, a comparable function, with respect to collagen I fibrils, has been proposed for the minor collagen V (8). This, and additional structural and metabolic similarities between collagens V and XI, have lent support to the notion that the relationship between collagens II and XI in cartilage is analogous to that of collagens I and V in other tissues (9). Moreover, the recent finding in adult bovine bone of α1(XI), α1(V), and α2(V) chains in the ratio of 1:1:1 has suggested the possible formation of cross-type heterotrimers between some of the subunits of collagens V and XI (10).

To gather more information about the structure and expression of this minor cartilage collagen type, we have here extended our cloning experiments on pro-α1(XI). This study has inferred the organization of the N-propeptide and thus completed the determination of the amino acid sequence of the human chain. Moreover, analysis of a large panel of cell lines has confirmed the unrestricted expression of the gene by documenting high levels of pro-α1(XI) mRNA in two human tumor cells of noncartilagenous origin.

MATERIALS AND METHODS

cDNA Cloning and DNA Sequencing—Primer extension cDNA cloning using the λgt10 phage vector and screening of the recombinants, as well as isolation and purification of the resulting positive clones, were essentially as previously described (11). Sequencing was carried out according to the dideoxy chain termination procedure on double-stranded DNA (12). Multiple sequencing of both strands was performed for the sequences of this report. Computer analysis was done using the sequence analysis program developed by Mount and Conrad (13).

RNA Isolation and Northern Blot Hybridization—For Northern blots, poly(A)$^+$ RNA was purified from cultured cells using the isothiocyanate method followed by elution through an oligo(dT)-cellulose (Collaborative Research) column (14). Approximately 1 μg of poly(A)$^+$ RNA was electrophoresed in 0.8% gel in the presence of formamide/formaldehyde, transferred onto nitrocellulose filters (Millipore), and hybridized to nick-translated DNA as described (15).

The 14 cell lines used for the Northern blot experiments are listed below along with the source from which they were obtained. American Type Culture Collection: HT-1080 (CCL 121), human fibrosarcoma; RD (CCL 136), human embryonal rhabdomyosarcoma; A-673 (CRL 1598) and A-204 (HTB 82), human rhabdomyosarcomas; Hs 578T (HTB 126), human breast ductal carcinoma; KATO III (HT 103), human gastric carcinoma; A-549 (CCL 185), human lung carcinoma; Hs 696 (HTB 151), primary human adenocarcinoma metastatic to bone sacrum; CRL 1183, human skin fibroblasts; AV_3 (CCL 21),

* This work was supported from Grants AR-38648 and HD-22657 from the National Institutes of Health. 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) J05407.

‡ On leave of absence from Kumamoto Medical School, Kumamoto (Japan).

§ To whom correspondence and reprint requests should be addressed. Present address: Brookdale Center for Molecular Biology, Mount Sinai School of Medicine, 1 Gustave Levy Place, New York, NY 10029.

[1] The abbreviations used are: COL11A1, the gene encoding pro-α1(XI) collagen; COL11A2, the gene encoding pro-α2(XI) collagen; N-propeptide, the amino-terminal propeptide of the procollagen molecule; N-proteinase, the enzyme that cleaves the N-propeptide.

human amnion cells. National Institute of General Medical Sciences: GM00637D, SV40-transformed human adult fibroblasts. Dr. M. Young (National Institute of Dental Research, Bone Research Branch): 563 and 564, bovine fetal long bone; 590, human adult bone.

RESULTS AND DISCUSSION

Oligomer-specific cDNA Cloning—We have previously isolated and sequenced two overlapping human cDNA clones, OK4 and Hpl 1, coding a 264-amino acid C-propeptide and 1005 amino acid residues of the triple helical domain of human pro-α(XI) collagen (3). To complete the determination of the primary structure of pro-α1(XI), we have now extended our sequence analysis by oligomer-specific cDNA cloning.

To this end, a synthetic oligomer, termed 1A, complementary to a sequence located near the 5′ end of clone Hpl 1, was used to prime reverse transcription of pro-α1(XI) mRNA (Figs. 1 and 2). To ensure specificity, the resulting clones were screened with an oligomer probe, termed 1B, positioned 7 nucleotides 5′ of the sequence of oligomer 1A (Figs. 1 and 2). As a result, four positive recombinants were identified and characterized. Upon sequence inspection, the largest of them, HY3, was found to contain an open reading frame whose conceptual translation overlaps with the sequence of Hpl 1 and extends into a noncollagenous domain (Fig. 1). Hence, a second cloning experiment was performed using oligomer 2A for RNA priming and oligomer 2B for screening of cDNAs (Figs. 1 and 2). This led to the isolation of two short clones the longest of which, HY93, extends nearly 150 base pairs 5′ of HY3 (Fig. 1). Moreover, parallel probing of the oligomer 1A-primed cDNA library with oligomer 2B identified a 1.3-kilobase pair cDNA, HY477, which overlaps with Hpl 1 and extends an additional 280 base pairs 5′ of HY93 (Fig. 1). Since HY477 also bears a continuous open reading frame, oligomers 3A and 3B were synthesized and used for primer extension cDNA cloning and for screening of the recombinants, respectively (Figs. 1 and 2). Sequencing of two of the resulting positive clones, HY81 and HY83, established their identity and, more importantly, revealed the untranslated nature of the first 161 base pairs of the longest clone, HY83

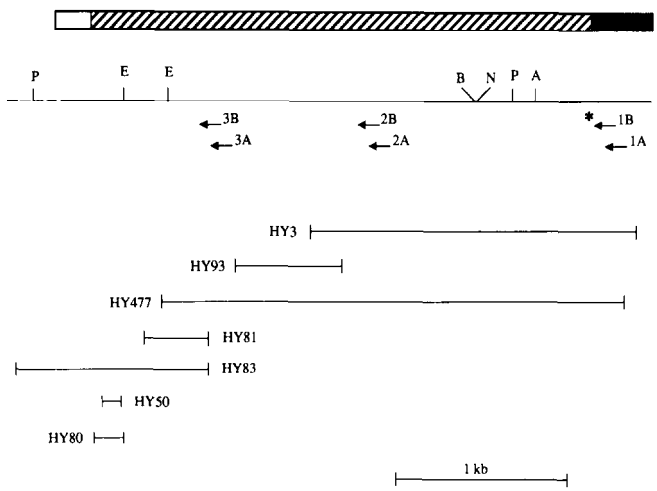

FIG. 1. **Characterization of pro-α1(XI) cDNA clones.** An overall restriction map of the cDNAs generated in this study is shown in relationship to the relative position of the signal peptide (*white box*), N-propeptide (*hatched box*), and helical domain (*black box*) of the protein. The *dotted line* signifies noncoding sequences, the *letters* the following enzymes: *Apa*I (*A*), *Ball* (*B*), *Eco*RI (*E*), *Nco*I (*N*), and *Pst*I (*P*). The *arrows* indicate the positions of the oligomers used for RNA priming and library screening, and the *asterisk* the 5′ foremost position of the previously described clone Hpl 1(3). Below the map are the relative positions of the cDNAs generated in the primer extension cloning experiments.

(Figs. 1 and 2). Confirmatory of these findings was the independent isolation of two very short cDNAs, HY50 and HY80, from the oligomer 1A-primed library probed with the 5′ *Eco*RI fragment of clone HY83 (Fig. 1).

Aside from using distinct oligomers for priming and screening, other potential cloning artifacts were excluded by comparing the Northern blot hybridization pattern of the original pro-α1(XI) cDNA, OK4, with those of two primer-extended recombinants, namely HY93 and the 5′ *Eco*RI subclone of HY83 (Fig. 3).

Sequence Analysis—A composite of the nucleotide sequence of the cDNA clones that cover the 5′ portion of the pro-α1(XI) collagen mRNA is shown in Fig. 2 along with its amino acid translation.

Following the ATG codon is a stretch of 36 amino acids that displays some of the features of a signal peptide (16). Specifically, a positively charged residue, Arg, precedes a hydrophobic core that is in turn followed by a large polar amino acid, Gln, positioned 8 residues to the left of a putative cleavage site, Ala-Ala (Fig. 2).

Four-hundred ninety-two amino acids separate the putative signal peptidase cleavage site from the beginning of the triple helical domain of the α1(XI) chain. This represents the longest N-propeptide hitherto described in a fibrillar procollagen. Moreover, the structure of the α1(XI) N-propeptide is highly divergent from those of other fibrillar molecules. In broad terms, a fibrillar procollagen N-propeptide can be generalized as displaying either a two-domain or a three-domain configuration (17). The latter consists of a globular region with 10 characteristically spaced cysteines that are believed to be involved in intrachain bonding, a continuous or interrupted collagenous stretch that participates in triple helical formation, and a short nonhelical segment that connects the amino peptide to the central triple helical domain. All but the globular region are present in the alternative two-domain structure (17).

Upon coarse examination, the pro-α1(XI) N-propeptide appears to be divided into three consecutive elements: a long globular domain, an interrupted collagenous region, and a short nonhelical segment. The globular region is extremely long, for it encompasses 383 amino acids or nearly three-fourths of the N-propeptide, and displays a unique arrangement of sequences. To be precise, 3 cysteines positioned almost in the middle of the globular region demarcate a sharp transition from a slightly basic into a highly acidic region with theoretical pI values of 8.98 and 3.61, respectively (Fig. 2). Although the significance of such an unusual architecture cannot be inferred from the sequencing data, one may speculate that it might be involved in the slow, and probably partial, proteolysis of type XI procollagen to collagen and/or in interactions with other matrix components (2). Antibodies produced on the basis of the data gathered from this study are currently being utilized to elucidate this and other important questions that pertain to collagen XI metabolism and function.

Twenty-four Gly-*X*-*Y* repeated triplets constitute the collagenous segment of the pro-α1(XI) N-propeptide. Two noncollagenous interruptions divide the repeated motif into three clusters of 4, 17, and 3 Gly-*X*-*Y* triplets (Fig. 2). We argue that, because of its length and the low imino acid content, it is most unlikely that the last set of repeats can participate in triple helical assembly. With the exception of the length of the noncollagenous interruptions, 10 *versus* 4 residues, the organization of the first two collagenous clusters of pro-α1(XI) resembles somewhat those of the pro-α1(II) and pro-α2(V) N-propeptides (18, 19). More interestingly, the three

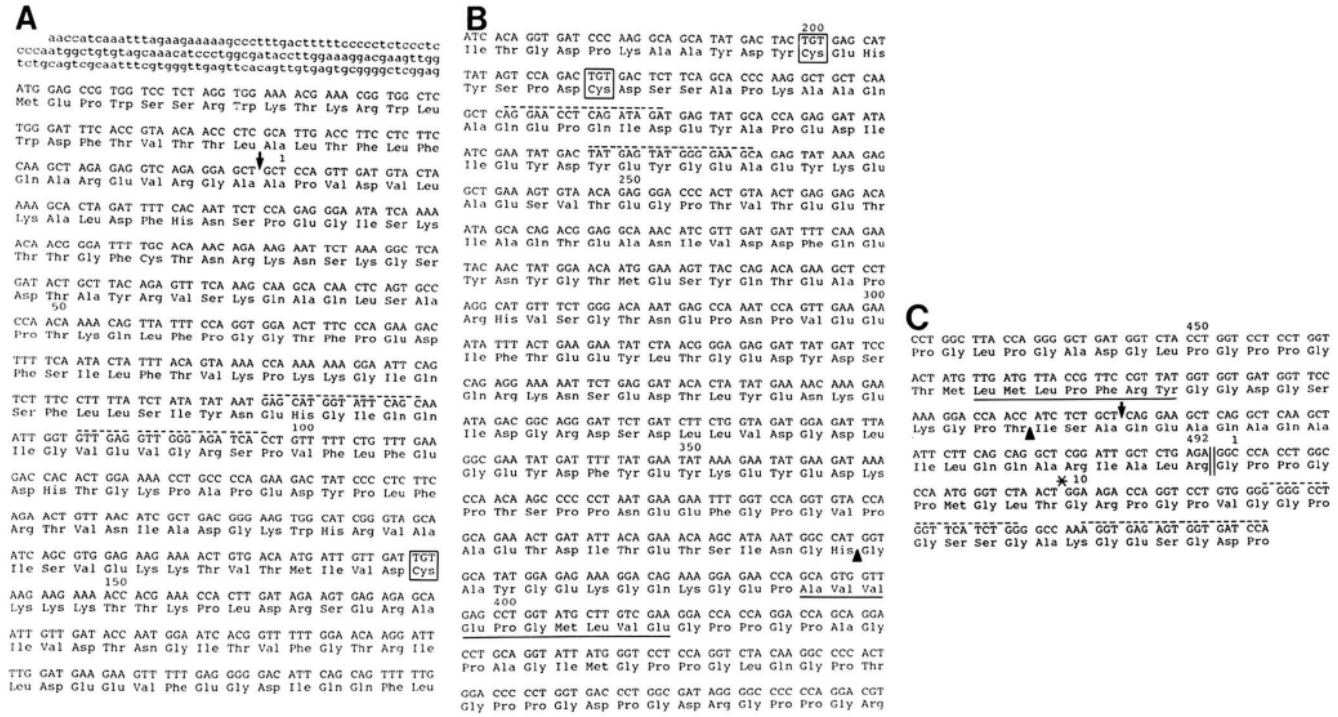

FIG. 2. **Nucleotide and deduced amino acid sequences of the 5′ portion of pro-α(XI) collagen.** *Capital letters* indicate the coding sequences, whereas *lower case letters* signify the 5′ noncoding sequences. Amino acid residues are *numbered* above the sequences. Numbering of the N-propeptide sequence begins immediately after the putative signal peptide cleavage site and ends at the beginning of the triple helical domain (*double vertical bar*). The triple helical domain sequences are numbered from the first Gly and extend to the sequence of Hpl 1 (*asterisk*) used for the generation of the first sets of synthetic oligomers. *Dotted lines* above the nucleotide sequence indicate the positions of the synthetic oligomers, which are from 3′ to 5′: 1A, 1B, 2A, 2B, 3A, and 3B. Cysteines of the globular region are *boxed.* Borders of the collagenous segment are signified by the symbol (▲), while its interruptions are *underlined. Arrows* indicate the putative signal peptidase and N-proteinase cleavage sites.

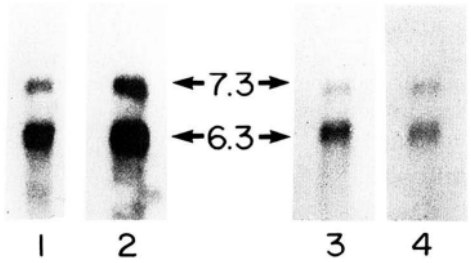

FIG. 3. **Northern blot hybridizations of pro-α1(XI) cDNAs.** The probes used are OK4 (*lanes 1* and *3*), HY93 (*lane 2*), and the 5′ *Eco*RI fragment of HY83 (*lane 4*). Approximately 1 μg of A-204 poly(A)$^+$ RNA was used in each of the two Northern blots. Sizes of the pro-α1(XI) mRNA species are indicated in kilobases.

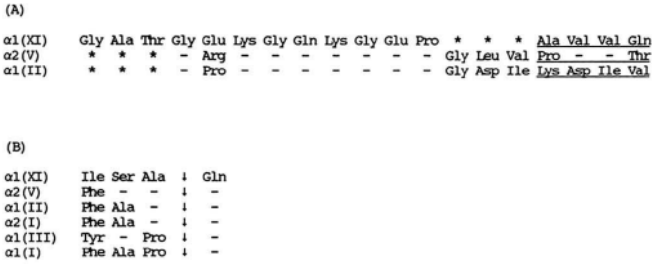

FIG. 4. **Sequence comparison of selected portions of human fibrillar collagen N-propeptides.** *Panel A*, collagenous segments with the adjacent collagenous interruptions are *underlined. Panel B*, sequence around the N-proteinase cleavage site (*arrow*). *Dashes* show no change from the pro-α1(XI) sequence, whereas *asterisks* indicate gaps inserted to give best alignment.

Gly-*X*-*Y* repeats to the left of the first interruption score a nearly perfect match when the three human chains are aligned together (Fig. 4*A*). We believe that these structural similarities are consonant with biochemical evidence strongly suggesting the formation of cross-type heterotrimers between the subunits of types V and XI, as well as with the notion of a common subunit between collagens II and XI (6, 7).

Two important structures are usually found in the short nonhelical segment which connects the collagenous region of the N-propeptide to the α-chain: the N-proteinase cleavage site and the cross-linking lysine residue (17, 20). While the latter is absent in pro-α1(XI), a potential N-proteinase cleavage site, Ile-Ser-Ala↓Gln, is recognized on the basis of its similarity to the moderately conserved sequence of fibrillar procollagen chains (Fig. 4*B*) (19). Assuming that our premise is correct, the α1(XI) chain then contains a 17-amino acid amino-terminal telopeptide which, unique among the fibrillar collagens, lacks the cross-linking site.

COL1A11 Expression in Noncollagenous Tissues—The isolation of human pro-α1(XI) clones from placental tissue and SV40-transformed fibroblast cDNA libraries has previously inferred an unrestricted expression of COL11A1 (3). In contrast, Northern blot analysis has more recently shown that COL11A2 is not expressed in human adult liver, skin, and tendon (5).

To further explore this phenomenon, we analyzed the collagen content in the RNAs from a large panel of normal and tumor cell lines with probes coding for pro-α1(XI), pro-α2(XI), pro-α2(V), pro-α1(II), and pro-α2(I) chains. Of the 14 cell lines analyzed, only 3 gave a positive signal when hybridized to the COL11A1 probe (Fig. 5). To be precise, two

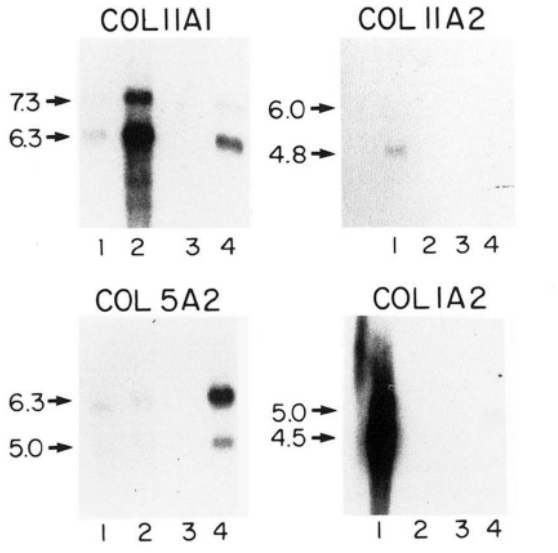

FIG. 5. **Pattern of collagen expression.** In each of the four autoradiograms, the RNAs are from GM00637D (*lane 1*), A-204 (*lane 2*), HT-1080 (*lane 3*), and RD (*lane 4*). Sizes of mRNAs, in kilobases, are indicated on the *left side* of the *panels*. Each autoradiogram was exposed for an identical period of time (48 h); each *lane* contains the same amount of poly(A)$^+$ RNA (1 μg); each probe was used at the same concentration (5 ng/ml) and at a nearly identical specific activity (3×10^8 cpm/μg). Note that the 6.3-kilobase pro-α2(XI) mRNA species (5) cannot be seen in this particular exposure. Furthermore, the COL1A2 band seen with RD RNA is a cross-hybridization product because it does not correspond to the size of a pro-α2(I) mRNA (22).

rhabdomyosarcomas, A204 and RD, were found to express pro-α1(XI) and pro-α2(V) mRNAs and not to have detectable amounts of pro-α2(I) and pro-α2(XI) mRNAs (Fig. 5). In contrast, in the SV40-transformed fibroblast line GM0036D, pro-α1(XI), pro-α2(XI), and pro-α2(V) mRNAs were present at levels significantly lower than pro-α2(I) (Fig. 5). This result is consistent with the relative representation of each of these collagen sequences in the GM0036D cDNA library (3, 21, 22). Relative amounts of pro-α2(V) and pro-α1(XI) mRNAs were also noted to be distinctly different in each of the two rhabdomyosarcomas (Fig. 5). Type XI mRNAs were not expressed to a significant amount in other randomly chosen tumor lines, including an embryonal rhabdomyosarcoma, as well as in fetal or adult bone tissue, and amnion fibroblasts. Finally, in none of the cell lines analyzed in this manner were detectable levels of pro-α1(II) mRNA (data not shown).

In conclusion, the results are consistent with the previous suggestion (3) that COL11A1 transcription occurs in tissues other than cartilage, such as placenta, and some tumor or virally transformed cell lines. It could be argued that ectopic gene expression in transformed cells is not physiologically significant. It is, however, of some interest to note that the independently co-expressed type XI gene, COL11A2, and the major cartilage collagen gene, type II, do not display the same pattern of expression as COL11A1. Hence, the two rhabdomyosarcomas may represent suitable experimental models for elucidating the molecular factors involved in the selective activation of this cartilage-specific collagen gene.

The data also provide the basis for an analysis of whether pro-α1(XI) collagen chains are produced by these tumor lines and, if so, to characterize into which parental trimers they are assembled. Relevant to this, early studies have suggested that the collagen of A-204 cultures consists solely of pro-α1(V) and pro-α2(V) chains that undergo limited processing, in that medium collagen appears to retain large nontriple helical domains (23). Such a finding should now be reconsidered in light of the results presented here, and of recent structural data that are consistent with complete proteolytic processing of the α2(V) propeptides (19, 24).

Acknowledgments—We are very much indebted to Drs. B. Olsen and M. Young for providing us with the pro-α2(XI) probe and bone-derived cell lines, respectively, and to Dr. M. van der Rest for many helpful suggestions and discussions. We thank R. Lingeza for the preparation and typing of the manuscript.

REFERENCES

1. Burgeson, R. E., and Hollister, D. (1979) *Biochem. Biophys. Res. Commun.* **87,** 1124–1131
2. Morris, N. P., and Bächinger, H. P. (1987) *J. Biol. Chem.* **262,** 11345–11350
3. Bernard, M., Yoshioka, H., Rodriguez, E., van der Rest, M., Kimura, T., Ninomiya, H., Olsen, B. R., and Ramirez, F. (1988) *J. Biol. Chem.* **263,** 17159–17166
4. Henry, I., Bernheim, Bernard, M., van der Rest, M., Kimura, T., Jeanpierre, C., Barichard, F., Berger, R., Olsen, B. R., Ramirez, F., and Junien, C. (1988) *Genomics* **3,** 87–90
5. Kimura, T., Chea, K. S. E., Chan, S. D. H., Lui, V. C. H., Mattei, M. G., van der Rest, M., Ono, K., Solomon, E., Ninomiya, Y., and Olsen, B. R. (1989) *J. Biol. Chem.* **264,** 13910–13916
6. Burgeson, R. E., Hedba, P. A., Morris, N. P., and Hollister, D. W. (1982) *J. Biol. Chem.* **257,** 7852–7856
7. Mendler, M., Eich-Bender, S. G., Vaughn, L., Winterhalter, K. H., and Bruckner, P. (1989) *J. Cell Biol.* **108,** 191–197
8. Birk, D. E., Ficht, J. M., Babiarz, J. P., and Linsenmayer, T. F. (1988) *J. Cell Biol.* **106,** 999–1008
9. 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, Orlando, FL
10. Niyibizi, C., and Eyre, D. R. (1989) *FEBS Lett.* **242,** 314–318
11. Weil, D., Bernard, M., Combates, N., Wirtz, M. K., Hollister, D. W., Steinmann, B., and Ramirez, F. (1988) *J. Biol. Chem.* **263,** 8561–8564
12. Hattori, M., and Sakaki, Y. (1986) *Anal. Biochem.* **152,** 232–238
13. Mount, D. W., and Conrad, B. (1987) *Nucleic Acids Res.* **14,** 11–16
14. Chirwing, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299
15. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 5201–5205
16. Perlman, D., and Halvorson, H. O. (1983) *J. Mol. Biol.* **167,** 391–409
17. Timpl, R., and Glanville, R. W. (1981) *Clin. Orthop. Relat. Res.* **158,** 224–242
18. Su, M. W., Benson-Chanda, V., Vissing, H., and Ramirez, F. (1989) *Genomics* **4,** 438–441
19. Woodbury, D., Benson-Chanda, V., and Ramirez, F. (1989) *J. Biol. Chem.* **264,** 2735–2738
20. Eyre, D. R., Paz, M. A., and Gallop, P. M. (1984) *Annu. Rev. Biochem.* **53,** 717–748
21. Weil, D., Bernard, M., Gargano, S., and Ramirez, F. (1987) *Nucleic Acids Res.* **15,** 181–198
22. 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
23. Alitalo, K., Myllyla, R., Sage, H., Pritzl, P., Vaheri, A., and Bornstein, P. (1982) *J. Biol. Chem.* **257,** 9016–9024
24. Myers, J. C., Loidl, H. R., Stolle, C. A., and Seyer, J. M. (1985) *J. Biol. Chem.* **260,** 5533–5541