Structure of cDNA clones coding for human type II procollagen

The $\alpha 1(\text{II})$ chain is more similar to the $\alpha 1(\text{I})$ chain than two other α chains of fibrillar collagens

Clinton T. BALDWIN,* Anthony M. REGINATO,*† Carol SMITH,* Sergio A. JIMENEZ*† and Darwin J. PROCKOP*†

*Department of Biochemistry and Molecular Biology and †Department of Medicine, Jefferson Institute of Molecular Medicine, Jefferson Medical College, Thomas Jefferson University, Philadelphia, PA 19107, U.S.A.

Overlapping cDNA clones were isolated for human type II procollagen. Nucleotide sequencing of the clones provided over 2.5 kb of new coding sequences for the human pro $\alpha 1(\text{II})$ gene and the first complete amino acid sequence of type II procollagen from any species. Comparison with published data for cDNA clones covering the entire lengths of the human type I and type III procollagens made it possible to compare in detail the coding sequences and primary structures of the three most abundant human fibrillar collagens. The results indicated that the marked preference in the third base codons for glycine, proline and alanine previously seen in other fibrillar collagens was maintained in type II procollagen. The domains of the pro $\alpha 1(\text{II})$ chain are about the same size as the same domains of the pro α chains of type I and type III procollagens. However, the major triple-helical domain is 15 amino acid residues less than the triple-helical domain of type III procollagen. Comparison of hydropathy profiles indicated that the α chain domain of type II procollagen is more similar to the α chain domain of the pro $\alpha 1(\text{I})$ chain than to the pro $\alpha 2(\text{I})$ chain or the pro $\alpha 1(\text{III})$ chain. The results therefore suggest that selective pressure in the evolution of the pro $\alpha 1(\text{II})$ and pro $\alpha 1(\text{I})$ genes is more similar than the selective pressure in the evolution of the pro $\alpha 2(\text{I})$ and pro $\alpha 1(\text{III})$ genes.

INTRODUCTION

Type II collagen is the most abundant protein in cartilage and a member of the family of fibrillar collagens (see Piez & Reddi, 1984; Prockop & Kivirikko, 1984). In cartilage it forms an arcade that provides a scaffolding for proteoglycans and other constituents of the cartilage matrix. The protein is also found in the vitreous humour and related structures of the eye (von der Mark *et al.*, 1979). In addition, it appears to have an important role in the embryonic development of many tissues. There are preliminary indications that mutations in the gene for type II procollagen (*COL2A1*), the precursor of type II collagen, are the cause of several human diseases. For example, restriction-fragment-length polymorphism data suggest that mutations in the type II procollagen gene are the cause of hereditary arthro-ophthalmopathy (Stickler's syndrome) in some families (Francomano *et al.*, 1987; Knowlton *et al.*, 1989a) and the cause of primary generalized osteoarthritis in other families (Peltonen *et al.*, 1988; Knowlton 1989b).

To date, the structures of the type II collagen and of the type II procollagen gene have not been fully identified. About half of the 1014 amino acid residues of the $\alpha 1(\text{II})$ chain were defined by peptide sequencing of the bovine protein (see Butler *et al.*, 1974, 1976; Francis *et al.*, 1978; Bornstein & Traub, 1979). About 300 codons for the C-terminal region of the pro $\alpha 1(\text{II})$ chain were defined by nucleotide sequencing of cDNAs for the chick protein (Sandell *et al.*, 1983, 1984; Ninomiya *et al.*, 1984; Young *et al.*, 1984). About 120 codons were defined by nucleotide sequencing of genomic clones for the 5'-end of the

pro $\alpha 1(\text{II})$ gene from rat (Kohno *et al.*, 1984, 1985). About 600 codons for the pro $\alpha 1(\text{II})$ chain were defined by nucleotide sequencing for cDNA and genomic clones for the human protein (Strom & Upholt, 1984; Cheah *et al.*, 1985; Elima *et al.*, 1985, 1987; Nunez *et al.*, 1985, 1986; Sangiorgi *et al.*, 1985; Stoker *et al.*, 1985).

Here we describe the isolation and analysis of a series of overlapping cDNA clones for human type II procollagen. The results provide over 2.5 kb of new coding sequences for the human pro $\alpha 1(\text{II})$ gene and the first complete amino acid sequence of type II procollagen from any species. Because cDNA clones covering the entire lengths of human type I (Kuivaniemi *et al.*, 1988; Tromp *et al.*, 1988) and type III (Ala-Kokko *et al.*, 1989) procollagens were recently analysed, the data made it possible to compare in detail the coding sequences and primary structures of the three most abundant human fibrillar collagens. The availability of the cDNA clones will greatly facilitate detection of mutations in the human type II procollagen gene that may cause heritable disorders of cartilage.

MATERIALS AND METHODS

To prepare the cDNA libraries, total RNA was isolated from a pooled sample prepared by dissecting cartilage from knee condyles, femoral heads and tibiae from a spontaneous 24-week human abortion. The tissue was obtained from the International Institute for the Advancement of Medicine, Philadelphia, PA, U.S.A., with protocols reviewed and approved by the institutional

† To whom correspondence should be addressed.

These sequence data have been submitted to EMBL/GenBank Data Libraries.

review committee in accordance with the National Organ Transplant Act and Pennsylvania Organ Transplant Act. To remove adherent tissues from the cartilage, the tissues were incubated for 1 h at 37 °C in Hanks medium containing 2 mg of trypsin/ml and 2 mg of bacterial collagenase/ml (Cheung & Ryan, 1981). The medium was discarded. The tissues were then minced and digested overnight at 37 °C in Dulbecco's modified Eagle's medium (GIBCO) containing 10% (v/v) fetal bovine serum and 0.5 mg of bacterial collagenase/ml. The cells released by the digestion were filtered through a nylon membrane and washed four times with medium not containing collagenase. Half the cells, about 3.8×10^8 , were taken directly for extraction of total RNA. The other half were first plated in a 175 cm² plastic culture flask and maintained at 37 °C for 4 days in Dulbecco's modified Eagle's medium containing 10% (v/v) bovine calf serum before isolation of RNA.

Total RNA was isolated from either the freshly isolated cells or the cultured cells by extraction with guanidinium isothiocyanate/CsCl and density-gradient centrifugation (Maniatis *et al.*, 1982). About 15 µg of total RNA was used to synthesize cDNA (Gubler & Hoffman, 1983) by using oligo(dT) and Maloney murine-leukaemia virus reverse transcriptase (Bethesda Research Laboratories). *EcoRI* linkers were attached to the double-stranded cDNA, ligated to a lambda vector (lambda-ZAP; Stratagene), and packaged *in vitro* by using a commercial package extract (Gigapack; Stratagene). A bacteriophage library was obtained that contained about 10^6 independent clones. The library was screened with a *PstI/EcoRI* 0.4 kb fragment from the 5'-end of pHCAR3, a cDNA clone for the 3'-end of the human type II procollagen message kindly provided by Dr. E. Vuorio (Elima *et al.*, 1987). The probe was labelled with ³²P by nick translation to a specific radioactivity to about 10^8 c.p.m./µg of DNA and the library was screened by plaque hybridization. Positive clones were plaque-purified and inserts in the vector were automatically converted into a phagemid (Bluescript) with a helper virus (R408) under conditions recommended by the supplier (Stratagene). Two clones (AR5959 and AR5960 in Fig. 1) were isolated from the library. A second library was prepared by using a double primer consisting of oligo(dT) together with an 18-mer coding for a 5'-region of AR5960. A third cDNA library was prepared by using a double primer consisting of oligo(dT) and an 18-mer coding for the 5'-region of CS2 (see Fig. 1). The clone PCR15 was prepared by using the polymerase chain reaction (Saiki *et al.*, 1985) and one primer for the 5'-region of CS2 (Fig. 1) and another primer coding for the 5'-end of exon 1 as defined by nucleotide sequencing of genomic clones (Sangiorgi *et al.*, 1985; Nunez *et al.*, 1986).

For nucleotide sequencing, double-stranded DNAs from the phagemid were sequenced with the dideoxy method with T7 DNA polymerase (Sanger *et al.*, 1977).

RESULTS AND DISCUSSION

Nucleotide sequences of the cDNAs

A series of four overlapping cDNAs were isolated from three separate cDNA libraries (Fig. 1). The overlapping cDNAs were sequenced so as to generate all of the coding sequences of type II procollagen mRNA except for 40 nucleotide residues at the 5'-end. The last 40

Fig. 1. Partial restriction map of the cloned cDNAs for the pro α 1(II) chain of human type II procollagen

Key: N, N-propeptide; C, C-propeptide; ATG, start site for translation; X, *XhoI*-cleavage sites; E, *EcoRI*-cleavage sites; TAA, end of translation; AR113, CS2, AR5960 and AR5959, the four cDNA clones isolated here; PCR115, the polymerase chain product prepared here. Arrows indicate the sequencing strategy.

nucleotide residues were obtained by cloning a product obtained with the polymerase chain reaction with human cDNA as the template, one primer for the 5'-end of CS2 and a second primer for the 5'-end of exon 1 (Sangiorgi *et al.*, 1985; Nunez *et al.*, 1986). Except for regions that were previously reported by others, all the sequences were confirmed by sequencing both strands of the double-stranded cDNAs. The results provided over 2.5 kb of new coding sequences for the human pro α 1(II) gene.

Previously published data made it possible to locate most of the cDNA nucleotides in specific exons of the type II procollagen gene. Ten nucleotide sequences at the 3'-end of the cDNAs were assigned to the last 11 exons of the gene as defined by Cheah *et al.* (1985). Nucleotides from the 5'-end of the cDNAs were assigned to the first three exons on the basis of the data of Sangiorgi *et al.* (1985) and Nunez *et al.* (1986). Nucleotides were assigned to three internal exons (exons 24, 34 and 39) on the basis of data provided by Strom & Upholt (1984) and Sangiorgi *et al.* (1985). Nucleotides were assigned to most of the remaining exons by alignment with exons in the chick type II procollagen gene as defined by Upholt & Sandell (1986). To align the nucleotide sequences defined here with the known gene structures for fibrillar collagens, we elected to define the junction exon containing the procollagen N-proteinase-cleavage site as exon 6 and the junction exon containing the procollagen C-proteinase-cleavage site as exon 49. As a result, it was possible to assign all the nucleotides to specific exons except for the sequences between the 5'-end of exon 3 and the 3'-end of exon 6 (Fig. 2).

Since the data provided all the coding sequences for human type II procollagen, it was possible to examine conservation of amino acid sequences between the human protein and the regions of the bovine protein previously defined by peptide sequencing (Butler *et al.*, 1974, 1976; Francis *et al.*, 1978; Bornstein & Traub, 1979). There was a high degree of conservation in that there were only nine differences between the human and bovine proteins in about 500 amino acid residues. Similarly, there was extensive conservation of amino acid sequences in the N-propeptide between the human sequences and the previously defined rat sequences (Kohn *et al.*, 1984). Only ten of the 121 amino acid residues in the N-propeptide were different.

One striking difference between the human and the rat nucleotide sequences was that the rat sequence (Kohn

[illegible]

Met Pro

Fig. 2. Nucleotide and amino acid sequences of cDNA clones for the pro α 1(II) chain

The nucleotide residues are numbered from the start for transcription and the amino acid residues are numbered from the start site of translation (Nunez *et al.*, 1986). Top line: rat, nucleotide sequences from rat cDNA (Kohno *et al.*, 1984, 1985). Second line: human g, nucleotide sequences obtained from genomic clones for human type II procollagen (Strom & Upholt, 1984; Cheah *et al.*, 1985; Sangiorgi *et al.*, 1985; Nunez *et al.*, 1985, 1986). Third and fourth lines: human, nucleotide sequences and encoded amino acid sequences obtained here. Fifth line: human g, amino acid sequences obtained by selective sequencing of human genomic DNA (Strom & Upholt, 1984; Cheah *et al.*, 1985; Sangiorgi *et al.*, 1985; Nunez *et al.*, 1985, 1986). Sixth line: rat, rat amino acid sequences derived from cloned cDNA sequences (Kohno *et al.*, 1984, 1985). An additional line is entered for the amino acid sequences from bovine type II collagen where these were defined (see Butler *et al.*, 1974, 1976; Francis *et al.*, 1978; Bornstein & Traub, 1979). Also, some sequences are indicated for the chicken protein. These are amino acid residues 1039–1130 sequenced by W. Butler as reported by Cheah *et al.* (1985) and codons 1131–1160 as reported by Sandell *et al.* (1984). Key: ···, nucleotide sequences that are identical with the human; - , amino acid sequences that are identical with the human; E1 etc., exon boundaries defined as described in the text; ↓, three possible sites for cleavage of the signal peptide (Kohno *et al.*, 1984); ↑, cleavage site for procollagen *N*-proteinase; ▼, beginning and end of the major triple-helical domain of the $\alpha 1(\text{II})$ chain.

Data for type I collagen are from Bernard *et al.* (1983), Kuivaniemi *et al.* (1988) and Tromp *et al.* (1988). Data from type III collagen are from Ala-Kokko *et al.* (1989)

Amino acid	$\alpha 1(I)$	$\alpha 2(I)$	$\alpha 1(III)$	$\alpha 1(II)$	Third base
Glycine	0.05	0.51	0.45	0.44	U
	0.28	0.22	0.15	0.27	C
	0.18	0.22	0.33	0.23	A
	0.03	0.05	0.07	0.06	G
Codons examined	342	342	362	342	
Proline (total)	0.60	0.62	0.58	0.60	U
	0.38	0.21	0.12	0.24	C
	0.02	0.16	0.30	0.13	A
	0	0.01	0	0.03	G
Codons examined	235	199	240	227	
Proline (Yaa position)	0.84	0.73	0.56	0.73	U
	0.13	0.09	0.06	0.08	C
	0.03	0.16	0.36	0.15	A
	0	0.02	0.02	0.04	G
Codons examined	117	91	143	117	
Alanine	0.75	0.76	0.68	0.57	U
	0.20	0.15	0.15	0.24	C
	0.05	0.08	0.16	0.14	A
	0	0	0.01	0.03	G
Codons examined	118	107	89	97	

Table 2. Comparison of the sizes of the domains of human type II procollagen with the same domains of type I and type III procollagens

Data for type I procollagen are from Bernard *et al.* (1983), Kuivaniemi *et al.* (1988) and Tromp *et al.* (1988). Data from type III collagen are from Ala-Kokko *et al.* (1989).

Domain	No. of amino acid residues			
	Pro α 1(I)	Pro α 2(I)	Pro α 1(III)	Pro α 1(II)
Signal peptide	22	22	24	21, 23 or 25
N-propeptide	139	57	129	112
Cys in N-propeptide	10	2	11	1
Triple helix in N-propeptide	48	42	36	16+63
N-telopeptide	17	14	14	19
Triple helix	1014	1014	1029	1014
C-telopeptide	26	11	25	27
C-propeptide	246	247	246	246

et al., 1984) contained a codon for glycine between the end of the triple helix of the N-propeptide and the N-proteinase-cleavage site that was not found in the human sequence (after residue 108 in Fig. 2).

Comparison with previously published nucleotide sequences for the human type II procollagen (Strom & Upholt, 1984; Cheah *et al.*, 1985; Sangiorgi *et al.*, 1985) indicated differences in codons for four amino acids (residues 841, 842, 968 and 988 in Fig. 2). In addition, there were 12 base differences that did not change the amino acid sequence. One difference in the third base position was found in exon 23 (the codon for residue 488), three in exon 24 (residues 507, 508 and 509), one in exon 34 (residue 728), three in exon 29 (residues 834, 837 and 840), one in exon 42 (residue 965), two in exon 44 (residues 991 and 992) and one in exon 49 (residue 1137).

Comparison of codon usage among human cDNAs for fibrillar collagens

Nucleotide sequences for the complete human mRNAs from the pro α 1(I) (Tromp *et al.*, 1988) and pro α 2(I) chains (Kuivaniemi *et al.*, 1988) of type I procollagen and the pro α 1(III) chain of type III procollagen were recently defined (Ala-Kokko *et al.*, 1989). The results obtained here made it possible, not only to define the third base codon usage for type II procollagen, but also to compare it with codon usage for the other fibrillar collagens. As indicated in Table 1, the marked preference previously seen in the other fibrillar collagens in the third base codons for glycine, proline and alanine was maintained in type II procollagen. G was used in 6% or less of the codons. U was the most commonly used base in the codons. C was rarely used as the third base codon for proline in the Yaa position of the repeating Gly-Xaa-Yaa sequence of the protein, apparently because C-G sequences are prone to mutations (Bird, 1986; Brown & Bird, 1986).

Comparison of the domains of type II procollagen with type I and type III procollagens

The data developed here together with the recently obtained data on human type I and type III procollagens also made it possible to compare the domains of the proteins (Table 2). Because of ambiguity as to the cleavage site for the signal peptide (Kohn *et al.*, 1984), it is not clear whether the signal peptide of type II

Fig. 3. Hydropathy profiles of the complete amino acid sequences for the human pro α 1(I) chain, pro α 2(I) chain, pro α 1(III) chain and pro α 1(II) chain

The values were generated with the procedure of Kyte & Doolittle (1982) by using a window of seven amino acid residues. The sequences for the type I and type III procollagen chains were based on the cDNA clones (Bernard *et al.*, 1983; Kuivaniemi *et al.*, 1988; Tromp *et al.*, 1988; Ala-Kokko *et al.*, 1989).

procollagen has 21, 23 or 25 amino acid residues (Fig. 2). It is apparent, however, that the signal peptide of the pro α 1(II) chain is about the same size as the signal peptides of type I procollagen and type III procollagen. The total size of the N-propeptide (112 amino acid residues) is intermediate between the size of the N-propeptide pro α 1 chains of type I and type III collagens and the pro α 2 chain of type I procollagen. The N-propeptide contains less cysteine than the three N-propeptides of the other two procollagens. The triple-helical domain of the N-propeptides of type II procollagen is larger than the triple-helical domains found in the N-propeptides of the other two procollagens. Also, the triple-helical domain in the N-propeptide differs

Fig. 4. Two-dimensional dot-matrix comparisons of the hydropathy plots for human pro α 1(II) chain with the human pro α 1(I) chain, the human pro α 2(I) chain and the human pro α 1(III) chain

Data are taken from the hydropathy plot shown in Fig. 3. A dot indicates that the hydropathy values for the two chains are equal. (a) Pro α 1(II) chain (abscissa) versus pro α 1(I) chain; (b) pro α 1(II) chain versus pro α 2(I) chain; (c) pro α 1(II) chain versus pro α 1(III) chain.

in that it is interrupted by four amino acid residues that are non-collagenous. As a result, there is a sequence of 16 amino acid residues with a Gly-Xaa-Yaa repeat, followed by four non-collagenous amino acid residues, and then followed by 63 amino acid residues with a Gly-Xaa-Yaa repeat. The N-telopeptide is slightly larger in size than the two fibrillar collagens. The triple-helical domain is identical with that of the two chains of type I procollagen, but 15 amino acid residues less than the triple-helical domain of type III procollagen (Ala-Kokko *et al.*, 1989). The remaining domains are similar in size.

Comparison of hydropathy profiles of human type II procollagen with those of human type I and type III procollagens

The data developed here also made it possible to carry out comparisons of the hydropathy profiles (Kyte &

Doolittle, 1982) of human type II procollagen with those of human type I and type III procollagens. Visual inspection of the results (Fig. 3) suggested that the α chain domain of type II procollagen (residues 1–1014 in Fig. 3) was more similar to the α chain domain of the pro α 1(I) chain than to the pro α 2(I) chain or the pro α 1(III) chain. The impression was confirmed by a two-dimensional matrix comparison of the hydropathy values (Fig. 4). A distinct distribution of points along the diagonal was seen in the comparison of the α 1(II) chain with the α 1(I) chain, but not the other two chains for fibrillar collagens. Therefore the results suggest that the selective pressure during evolution of pro α 1(I) and pro α 1(II) genes is more similar than the selective pressure on the pro α 2(I) and pro α 1(III) genes. The selective pressures on the pro α 1(I) and pro α 1(II) chains are similar even though type I procollagen has remained a heterotrimer and type II procollagen a homotrimer.

Visual inspection of the hydropathy plots (Fig. 3) suggests that the region near the N-propeptide cleavage site of type II procollagen is more similar to the N-propeptide cleavage site of type I procollagen than to the N-propeptide cleavage site of type III procollagen, an observation consistent with evidence that type I and type II procollagens are cleaved by the same procollagen N-proteinase, whereas a separate procollagen N-proteinase is required to cleave type III procollagen (Halila & Peltonen, 1986; Tuderman *et al.*, 1987; Dombrowski & Prockop, 1988). There were no striking differences among the hydropathy plots for the C-propeptides of the three fibrillar collagens.

Note added in proof (received 10 July 1989)

Recent results from sequencing a genomic clone indicate that the boundary between exons 6 and 7 is between nucleotides 574 and 575, and not between nucleotides 545 and 546 as indicated in Fig. 2 (L. Ala-Kokko & D. J. Prockop, unpublished work).

The work presented here was supported in part by National Institutes of Health Research Grant AR-39740 and a postdoctoral fellowship grant from the Arthritis Foundation (to A. M. R.).

REFERENCES

- Ala-Kokko, L., Kontusaari, S., Baldwin, C., Kuivaniemi, H. & Prockop, D. J. (1989) *Biochem. J.* **260**, 509–516
- Bernard, M. P., Chu, M.-L., Myers, J. C., Ramirez, F., Eikenberry, E. F. & Prockop, D. M. (1983) *Biochemistry* **22**, 5213–5223
- Bird, A. P. (1986) *Nature (London)* **321**, 209–213
- Bornstein, P. & Traub, W. (1979) in *The Proteins*, 4th edn. (Neurath, H. & Hill, R. L., eds.), pp. 411–631, Academic Press, New York
- Brown, W. A. R. & Bird, A. P. (1986) *Nature (London)* **322**, 477–481
- Butler, W. T., Miller, E. J., Finch, J. E., Jr. & Inagami, T. (1974) *Biochem. Biophys. Res. Commun.* **57**, 190–195
- Butler, W. T., Miller, E. J. & Finch, J. E., Jr. (1976) *Biochemistry* **15**, 3000–3006
- Cheah, K., Stoker, H., Griffin, J., Grosveld, G. & Solomon, E. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2555–2559
- Cheung, H. S. & Ryan, L. M. (1981) *Anal. Biochem.* **116**, 93–97

- Dombrowski, K. E. & Prockop, D. J. (1988) *J. Biol. Chem.* **263**, 16545–16552
- Elima, K., Makela, J., Vuorio, T., Kauppinen, S., Knowles, J. & Vuorio, E. (1985) *Biochem. J.* **229**, 183–188
- Elima, K., Vuorio, T. & Vuorio, E. (1987) *Nucleic Acids Res.* **15**, 9499–9504
- Francis, G., Butler, W. T. & Finch, J. E., Jr. (1978) *Biochem. J.* **175**, 921–930
- Francomano, C., Liberfarb, R., Hirose, T., Maumenee, I., Streeten, E., Myers, D. & Pyeritz, R. (1987) *Genomics* **1**, 293–296
- Gubler, U. & Hoffman, B. J. (1981) *Gene* **25**, 263–269
- Halila, R. & Peltonen, L. (1986) *Biochem. J.* **239**, 47–52
- Knowlton, R. G., Weaver, E. J., Struyk, A. F., Knobloch, W. H., King, R. A., Norris, K., Shamban, A., Uitlo, J., Jimenez, S. A. & Prockop, D. J. (1989a) *Am. J. Hum. Genet.*, in the press
- Knowlton, R. G., Katzenstein, P., Moskowitz, R. W., Weaver, E. J., Jimenez, S. A., Pathria, M. N., Malamed, C. J. & Prockop, D. J. (1989b) *Cytogenet. Cell Genet.*, in the press
- Kohno, K., Martin, G. & Yamada, Y. (1984) *J. Biol. Chem.* **259**, 13669–13673
- Kohno, K., Sullivan, M. & Yamada, Y. (1985) *J. Biol. Chem.* **260**, 4441–4447
- Kuivaniemi, H., Tromp, G., Chu, M.-L. & Prockop, D. J. (1988) *Biochem. J.* **252**, 633–640
- Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor
- Ninomiya, Y., Showalter, A., van der Rest, M., Seidah, N., Chretien, M. & Olsen, B. (1984) *Biochemistry* **23**, 617–624
- Nunez, A., Francomano, C., Young, M., Martin, G. & Yamada, Y. (1985) *Biochemistry* **24**, 6343–6348
- Nunez, A., Kohno, K., Martin, G. & Yamada, Y. (1986) *Gene* **44**, 11–16
- Peltonen, L., Palotie, A., Vaisanen, P., Vuorio, E. & Ott, J. (1988) *Collagen Relat. Res.*, **8**, 509–510
- Piez, K. A. & Reddi, A. H. (eds.) (1984) *Extracellular Matrix Biochemistry*, Elsevier, New York
- Prockop, D. J. & Kivirikko, K. I. (1984) *N. Engl. J. Med.* **311**, 376–386
- Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985) *Science* **230**, 1350–1354
- Sandell, L. J., Yamada, Y., Dorfman, A. & Upholt, W. B. (1983) *J. Biochem. (Tokyo)* **258**, 11617–11621
- Sandell, L. J., Prentice, H., Kravis, D. & Upholt, W. B. (1984) *J. Biol. Chem.* **259**, 7826–7834
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Sangiorgi, F., Benson-Chada, V., de Wet, W., Sobel, M., Tsipouras, P. & Ramirez, F. (1985) *Nucleic Acids Res.* **13**, 2207–2225
- Stoker, N., Cheah, K., Griffin, J., Pope, F. & Solomon, E. (1985) *Nucleic Acids Res.* **13**, 4611–4613
- Strom, C. & Upholt, W. B. (1984) *Nucleic Acids Res.* **12**, 1025–1038
- Tromp, G., Kuivaniemi, H., Stacey, A., Shikata, H., Baldwin, C., Jaenisch, R. & Prockop, D. J. (1988) *Biochem. J.* **253**, 919–922
- Tuderman, L., Kivirikko, K. I. & Prockop, D. J. (1987) *Biochemistry* **17**, 2948–2954
- Upholt, W. B. & Sandell, J. L. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2325–2329
- von der Mark, H., von der Mark, K., Timpl, R. & Trelstad, R. (1979) *Dev. Biol.* **59**, 75–85
- Young, M., Vogeli, G., Nunez, A., Fernandez, M., Sullivan, M. & Sobel, M. (1984) *Nucleic Acids Res.* **12**, 4207–4228

Received 24 January 1989/29 March 1989; accepted 12 April 1989