Biochem. J. (1993) **289**, 195–199 (Printed in Great Britain)

Two cysteine substitutions in procollagen I: a glycine replacement near the N-terminus of α1(I) chain causes lethal osteogenesis imperfecta and a glycine replacement in the α2(I) chain markedly destabilizes the triple helix

Andrzej FERTALA, Andrea WESTERHAUSEN, Gregory MORRIS, John E. ROONEY and Darwin J. PROCKOP

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

Cultured skin fibroblasts were examined from two probands with type II (lethal) osteogenesis imperfecta. One proband had a single base mutation which converted the glycine codon at position α1-244 in the α1(I) chain of procollagen I into a cysteine codon whereas the other had a similar mutation that converted the glycine codon at position α2-787 of the α2(I) chain into a cysteine codon. Both mutations produced post-translational overmodification of procollagen I. The Cys α1-244 mutation, however, had a minimal effect on the thermal stability or secretion of the protein whereas the Cys α2-787 mutation markedly decreased the thermal stability and, apparently as a result, essentially none of the mutated protein was secreted. The results provide clear exceptions to two previous generalizations about the position-specificity of glycine substitutions in procollagen I.

INTRODUCTION

Mutations that change the primary structure of procollagen I have been shown to cause osteogenesis imperfecta (OI), a heritable disorder characterized by extreme brittleness of bone (Prockop, 1990; Byers, 1990; Kuivaniemi et al., 1991). With only a few exceptions, unrelated probands have different mutations in the genes for the proα(I) or proα2(I) chains of procollagen I (*COL1A1* and *COL1A2*). Also, most of the severe forms of the disease are caused by mutations that cause synthesis of structurally abnormal but partially functioning proα1(I) or proα2(I) chains. By far the most common mutations are single-base mutations which substitute amino acids with bulkier side chains for the glycine residues that normally appear as every third residue in the repeating -Gly-Xaa-Yaa sequence of the triple-helical domain of procollagen I (see Piez, 1984). The glycine substitutions, however, differ widely in their effects. Some that cause lethal variants of OI prevent folding of the protein into a triple helix and cause degradation of triple-stranded monomers containing normal and abnormal proα1(I) and proα2(I) chains (Williams and Prockop, 1983; Prockop, 1990). However, similar glycine substitutions in different positions in the same chains have little effect on the thermal stability of the protein. Instead, they allow folding but they alter the conformation of the molecule sufficiently to interfere with the normal self-assembly of the protein into collagen fibrils (Vogel et al., 1988; Torre-Blanco et al., 1992). Several hypotheses have been advanced to explain why some glycine substitutions have larger effects than others on the functional properties of procollagen and cause more severe phenotypes (Byers, 1990; Kuivaniemi et al., 1991). One hypothesis is that glycine substitutions near the C-terminus of the protein are more severe, because they more markedly interfere with the folding of the triple helix from the C-terminus to the N-terminus of the protein. Another hypothesis is that glycine substitutions in the proα1(I) chains are more devastating in their effects than substitutions in the proα2(I) chain because the proα1(I) chains are more critical for the normal function of the protein.

Here we have compared two substitutions of cysteine for glycine that cause lethal OI. One is a substitution of cysteine for glycine position 244 in the α1(I) chain and the other a substitution of cysteine for glycine position 787 in the α2(I) chain.

MATERIALS AND METHODS

The probands

The first proband (JIMM 14; Cys α1-244) was the first-born child of healthy unrelated parents. There was no family history of symptoms of OI or related disorders. Both parents had normal sclerae. Uncomplicated delivery occurred at term. The child was 44 cm long and weighed 2900 g. The sclera were normal. The limbs were markedly shortened and deformed, and the cranial vault was soft. Many areas of crepitation were perceived on palpation of the skull, the rib cage and the limbs. Bone X-ray examination revealed numerous fractures, recent and partially healed, involving the limbs and the ribs. Both clinical and X-ray findings supported the diagnosis of type II OI (Sillence et al., 1979; Sillence, 1982; Byers et al., 1991). The child died from bronchopneumonia at 8 weeks of age.

The second proband (JIMM 6; Cys α2-787) was the second child of normal parents with no family history of OI or related disorders. Ultrasound examination at $5\frac{1}{2}$ months of gestation showed abnormal intracranial structures. The fetus was small with a large head/trunk ratio and short lower extremities. Onset of labour was spontaneous at 32 weeks of gestation. The proband was in breech presentation and stillborn. The child was small in that it was 31.5 cm long and weighed 960 g. There were extensive injuries of the neck with partial avulsion. The skull was poorly calcified and partially collapsed. The bitemporal distance was broad. The sclerae were blue or bluish black. There were 12 very

Abbreviations used: OI, osteogenesis imperfecta; 1 × SSC, 0.15 M NaCl, 0.015 M sodium citrate, pH 6.8.

Table 1 Primers for PCR amplification

Underlinings indicate restriction sites that are convenient for cloning of the PCR products.

Template	Target sequence	Primer sequence
cDNA	α1 83–96	GTGTCTAGAATGAAGGGACACAGAGGTTTCAGTGGTTTGGG
cDNA	α1 404–413	GGGTGAATTCACACCTCGCTCTCCAGCAGCACCTTTAGGTCCAGGGAA
cDNA	α1 215–224	CATGGATCCCAGCCTGGTGCTAAAGGTGCCAATGGTGCT
cDNA	α1 404–413	GGGTGAATTCACACCTCGCTCTCCAGCAGCACCTTTAGGTCCAGGGAA
gDNA	α1 215–224	CAGCCTGGTGCTAAAGGTGC
gDNA	α1, Intron 19	GCTCAGTTGGCTCTGGCGTC
cDNA	α1 497–503	TTACCTGCAGGCAAGAGGCGAGAGAGGTTT
cDNA	α1 −203 to −222	TTTCTAGAAAGGTTGAATGCACTTTTG
cDNA	α2 757–764*	AAAAGGGATCCCTCTGGAGAGGCTGGTACT
cDNA	α2 797–803*	GAGGAATTCACCCACAGCACCAGCAAC

* To use genomic DNA (gDNA) as a template, the same set of primers was used.

short ribs and short thickened long bones with sharp bowing of the humeri, tibiae and the femurs. Chromosome analysis revealed a normal 46 XX karyotype.

Cell culture and biochemical studies

Skin fibroblasts were grown under standard conditions (Steinmann et al., 1984; Pack et al., 1989). To examine synthesis of the procollagens, the cells were incubated with [^{3}H]proline or [^{14}C]proline for 4 h at 37 °C as described previously (Steinmann et al., 1984; Superti-Furga and Steinmann, 1988; Pack et al., 1989). The procollagens were examined either directly or after digestion to collagen by pepsin (Steinmann et al., 1984; Pack et al., 1989). The thermal stability of the pepsin-purified collagens was assayed by digestion with trypsin and chymotrypsin as described by Bruckner and Prockop (1981). In brief, the samples of collagen were dissolved in 0.4 M NaCl/0.1 M Tris/HCl buffer, pH 7.4. The samples were preincubated at the temperatures indicated for 5 min and then digested for 2 min with 100 μg/ml trypsin and 250 μg/ml chymotrypsin. The digestion was terminated by addition of 0.1 vol. of 5 mg/ml soya bean trypsin inhibitor. The samples were dissolved in hot electrophoresis sample buffer, heated at 100 °C for 2 min and analysed by SDS/PAGE. The electrophoresis was performed as described previously (Laemmli, 1970; Pack et al., 1989) and the radiolabelled proteins were visualized either by autoradiofluorography (Laskey and Mills, 1975) or with a phosphor storage imager (PhosphorImager).

Analysis of mRNA-derived cDNA and genomic DNA

To prepare cDNA for use as a template for the PCR (Saiki et al., 1985), total RNA was isolated with a procedure involving extraction with guanidinium isothiocyanate and centrifugation on a caesium chloride gradient (Maniatis et al., 1982). The RNA was used to synthesize double-stranded cDNA as described by Gubler and Hoffman (1983). To isolate genomic DNA templates for the PCR, DNA was extracted from 175 cm^2 flasks of skin fibroblasts with proteinase K and SDS (Maniatis et al., 1982).

With the first proband (Cys α1-244), preliminary protein analysis (see below) suggested the presence of a cysteine substitution in the CNBr fragment CB8 of the α1(I) chain (amino acids 124–402). Therefore cDNA was prepared from fibroblast mRNA and an appropriate region of the cDNA was amplified by PCR with two sets of primers (Table 1). The PCR products were cloned into M13 and sequenced with the dideoxynucleotide chain-termination method (Sanger et al., 1977). To amplify an appropriate region of genomic DNA by PCR, the 5′-primer was complementary to codons of amino acids 215–224 of the α1(I) chain. The 3′-primer contained sequences of intron 19 of the *COL1A1* gene (Table 1).

With the second proband (Cys α2-787), preliminary protein data (see below) suggested a mutation near the collagenase cleavage site in either the α1(I) or α2(I) chain. Therefore corresponding regions of cDNAs for both chains were amplified by PCR. For the proα1(I) chain, the 5′-primer was complementary to the codons for amino acids 497–503 and the 3′-primer contained the nucleotides −203 to −222 in the 3′-untranslated region of the proα1(I) cDNA (Table 1). The PCR products were cloned into the plasmid Bluescript (Stratagene). The two alleles were distinguished by means of a polymorphic site at position −88 (Westerhausen et al., 1990). One clone for each allele was digested with exonuclease III to generate a deletion library (Erase-a-Base, Promega) and clones from the library were sequenced. To amplify the cDNA for the proα2(I) chain by PCR, the 5′-primer was complementary to the nucleotide sequence of amino acids 757–764 and the 3′-primer contained codons for amino acids 797–803 (Table 1). The PCR products were cloned into filamentous bacteriophage M13 mp18 or m13 mp19 and sequenced. To amplify the corresponding region of the proα2(I) gene, the same primers were used with a genomic DNA template.

PCR products were also used for dot blotting on nitrocellulose filters with allele-specific oligonucelotides (Studencki and Wallace, 1984). The oligonucleotides were 19-mers spanning the codons containing the mutations. The sequence of the oligonucleotides used for the α1-244 mutation was GGACCCCAGG/TCCCCGGCG. For the α2-787 mutation the sequence of the oligonucleotides was GGTCTCCC-TG/TGCTCGAGAG. The blots were prehybridized for 2 h in 6 × SSC (1 × SSC = 0.15 M NaCl and 0.015 M sodium citrate, pH 6.8), 5 × Denhardt's solution, 0.5% SDS and 0.25 mg/ml denatured salmon sperm DNA. Hybridization was performed at 60 °C for 4 h after the addition of γ-^{32}P-labelled oligonucleotides to the same solution. The filters were washed four times in 2 × SSC/0.1% SDS at room temperature for 15 min. The samples were washed in 2 × SSC at 1 °C below the calculated melting temperatures for 2 min (Maniatis et al., 1982; Studencki and Wallace, 1984).

RESULTS

Analysis of procollagen synthesized by the probands' fibroblasts

Fibroblasts from the two probands were incubated with either [^{3}H]proline or [^{14}C]proline and the procollagens synthesized were analysed by SDS/PAGE either directly or after pepsin digestion.

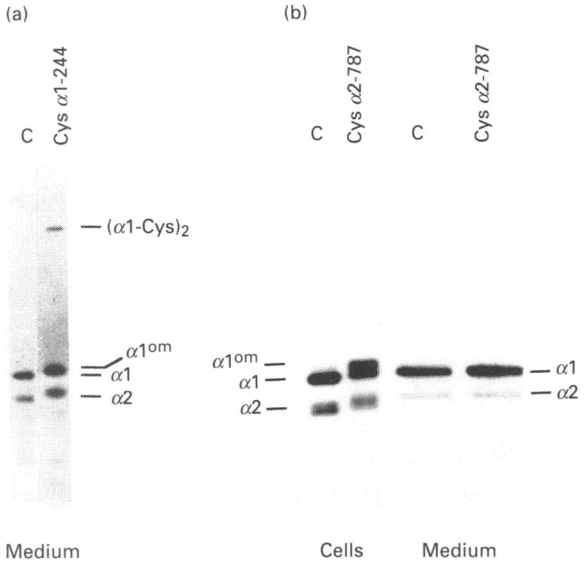

Figure 1 (a) Analysis of pepsin-digested collagen I secreted by fibroblasts from the first proband (Cys α1-244) and (b) pepsin-digested collagen I from cell layer and medium from skin fibroblasts of the second proband (Cys α2-787)

Cultured skin fibroblasts were incubated with [^{3}H]proline, as described in the text, and proteins were digested with pepsin before SDS/PAGE. C, proteins produced by control fibroblasts; $\alpha 1^{om}$, post-translationally overmodified α1(I) chains; $(\alpha 1\text{-Cys})_2$, disulphide dimers of α1(I) chains. The samples were electrophoresed without reduction.

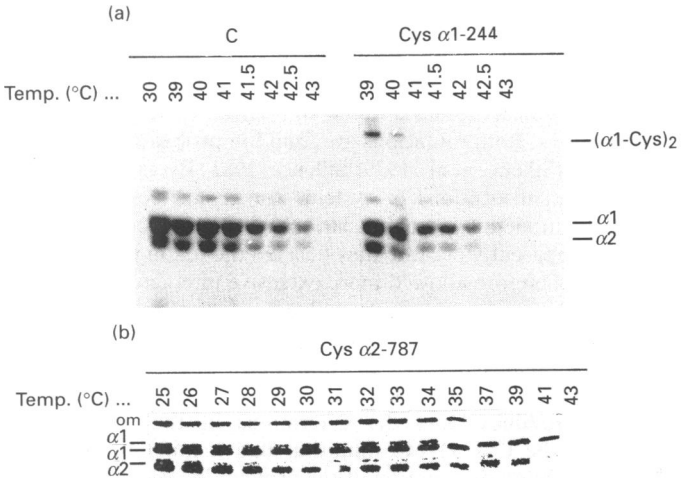

Figure 2 Assay of thermal stability of the triple helix by brief proteinase digestion of collagen from the probands

The proteins were digested with pepsin and then with trypsin and chymotrypsin at the temperature indicated as described. Samples were analysed by SDS/PAGE. C, procollagen secreted by control fibroblasts; Cys α1-244, Cys α2-787, samples from proband's fibroblasts; $(\alpha 1\text{-Cys})_2$, disulphide dimers of α1(I) chains. om, overmodified chains.

As indicated in Figure 1(a), abnormal protein was detected in the medium from the first proband (Cys α1-244). Analysis of the pepsin-digested procollagen indicated that both the α1(I) and α2(I) chains were slowly migrating, an observation suggesting that they were post-translationally overmodified. Also, examination of the protein without reduction revealed a slowly migrating band that was shown by reduction and electrophoresis in a second dimension (not shown) to consist of disulphide-linked α1(I) chains. Similar results were obtained by analysis of the pepsin-digested procollagen in the cell layer of the fibroblasts (not shown). In contrast with the results obtained with the first proband, analyses of procollagen synthesized by fibroblasts from the second proband (Cys α2-787) did not reveal the presence of any abnormal pepsin-digested α chains in the medium (Figure 1b). However, a portion of the intracellular protein was overmodified. There was no evidence of a disulphide-linked α2(I) chain in samples that were examined without reduction.

In further studies, the thermal stability of the procollagens from the two probands was examined by brief digestion with high concentrations of trypsin and chymotrypsin. The collagen from the first proband (Cys α1-244) demonstrated only a slight decrease in thermal stability of about 0.5 °C (Figure 2a). As also noted in Figure 2, the disulphide-linked α1(I) chains had a lower stability with a melting temperature of approx. 40 °C. In contrast with the results obtained with the first proband, a large fraction of the pepsin-digested collagen from the second proband (Cys α2-787) showed a markedly lowered thermal stability. As indicated in Figure 2(b), about half of the protein was digested at about 35 °C. The remainder of the protein had a normal melting temperature of about 41 °C.

Still further studies provided preliminary indications of the locations of the mutations. With procollagen from the first proband (Cys α1-244), two-dimensional gel electrophoresis of CNBr fragments with and without reduction indicated that a large proportion of α1-CB8 was recovered as a disulphide-linked dimer (not shown). The results indicated therefore a cysteine substitution in amino acids 124–402 of the α1(I) chain. With procollagen from the second proband (Cys α2-787), digestion with vertebrate collagenase followed by digestion with trypsin and chymotrypsin (Pack et al., 1989) indicated that the collagenase B fragment had a decreased thermal stability (not shown). The results therefore suggested a mutation in residues 776–1014 of either the α1(I) or α2(I) chain.

Nucleotide sequences of the mutations

To define the mutations, cDNA was prepared from fibroblast mRNAs, the cDNA was used as template for the PCR, and the PCR products were cloned and sequenced. Because the protein data on procollagen from the first proband suggested a cysteine substitution in amino acids 124–402 of the α1(I) chain, regions of the cDNA containing those codons were examined. Analysis of 12 clones demonstrated that six had a single-base mutation that converted the codon of -GGC- for glycine at position α1-244 into -TGC-, a codon for a cysteine (Figure 3). With procollagen from the second proband, regions containing codons for the collagenase B fragment were analysed in a similar manner. Therefore corresponding regions of the cDNAs for both α1(I) and α2(I) chains were analysed. The two alleles for the proα1(I) chain were distinguished by a polymorphism at position −88 (Westerhausen et al., 1990). No mutations were found in either allele in sequences coding for amino acids 497–1285 and in about 180 bp of the 3′-non-translated region. Therefore clones with sequences coding for amino acids 757–803 of the proα2(I) chain were analysed. The results revealed that 11 of 26 M13 clones had

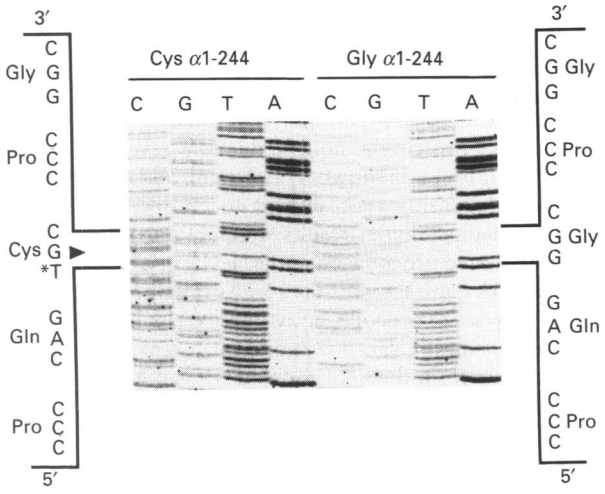

Figure 3 Nucleotide sequence of cDNA from the first proband (Cys α1-244)

cDNAs prepared with mRNA from skin fibroblasts, and the cDNAs were amplified by PCR. The PCR products were cloned and sequenced as described in the text. Cys α1-244, sequence from the first proband's cDNA; Gly α1-244, normal clone from the proband.

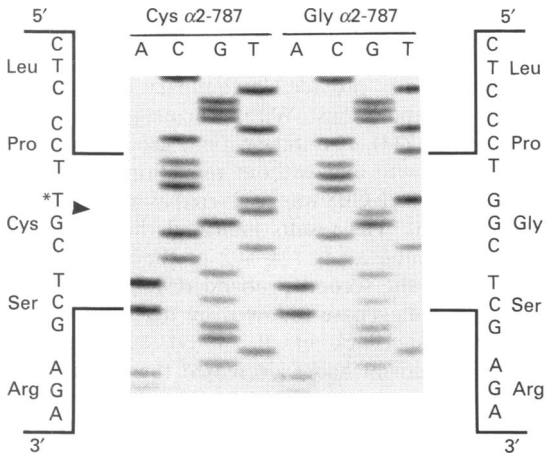

Figure 4 Nucleotide sequence of cDNA from the second proband (Cys α2-787)

DNA was processed and analysed as described in Figure 3 and in the text.

a single-base mutation that converted the codon of -GGC- for glycine at position α2-787 into -TGC-, a codon for cysteine (Figure 4).

To confirm the mutations, genomic DNA was used as a template for PCR, and the PCR products were hybridized with oligonucleotide probes for the normal sequences and abnormal sequences containing the codon for cysteine at α1-244 or the codon for cysteine at α2-787. Genomic DNA from the first proband hybridized with the probe from the normal sequence and the sequence containing the codon for cysteine at position α1-244 (not shown). The second proband genomic DNA hybridized with oligonucleotide probe both for the normal sequence and for a sequence containing the codon for cysteine at position α2-787 (not shown).

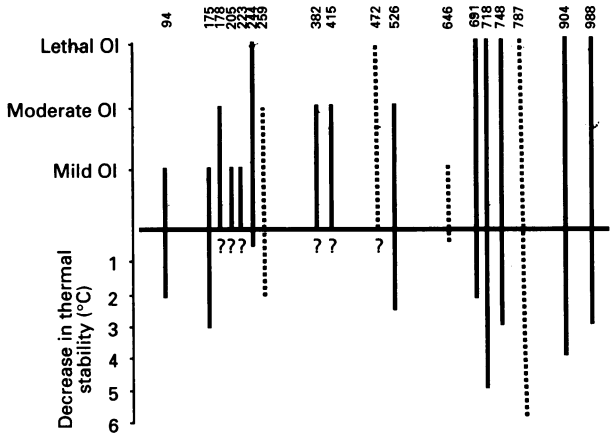

Figure 5 Schematic summary of cysteine-for-glycine substitutions in procollagen I as a cause of OI

For summary of previously reported cysteine substitutions, see Kuivaniemi et al. (1991). Solid bars, cysteine substitutions in the α1(I) chain; broken bars, cysteine-for-glycine substitutions in the α2(I) chain; question marks, mutations for which effects of the thermal stability were not assayed.

DISCUSSION

Well over 70 mutations that change the primary structure of the proα1(I) or proα2(I) chain of procollagen I have been found in probands with OI (Byers, 1990; Kuivaniemi et al., 1991). The most common mutations are single base substitutions that convert one of the 338 glycine codons in either the α1(I) or the α2(I) chain into a codon for an amino acid with a bulkier side chain. Since the collagen triple helix is generally regarded as a rigid rod with a uniform structure, it was surprising to find that the effects of glycine substitutions were highly position-specific. Substitutions of bulkier amino acids at some glycine positions produced mild phenotypes, whereas substitutions of the same amino acids at other glycine positions were lethal. Also, it was surprising to find that substitutions at some glycine positions had little effect on the thermal stability of the triple helix as assayed by proteinase digestion, whereas substitution of the same amino acid at other glycine positions markedly lowered the thermal stability (see Pack et al., 1989; Steinmann et al., 1991).

The two mutations defined here provide clear exceptions to several current hypotheses about the position effects of glycine substitutions. Both mutations are found in probands with lethal type II OI (Sillence et al., 1979; Sillence, 1981; Byers et al., 1991). Both were substitutions of cysteine for glycine residues. Both produced moderate post-translational overmodification of the protein, apparently because they delayed the folding of the triple helix and therefore allowed more extensive interactions with the post-translational enzymes (see Byers, 1990; Prockop, 1990). Both were approximately the same distance from the ends of the molecule in that the Cys α1-244 mutation was 261 residues from the N-terminus of the α1(I) chain and the Cys α2-787 mutation was 238 residues from the C-terminus of the α2(I) chain. However, the Cys α1-244 mutation had a minimal effect on thermal stability or secretion of the protein. In contrast, the Cys α2-787 markedly decreased the thermal stability. Apparently as a result, it delayed secretion and probably produced degradation of the procollagen (Williams and Prockop, 1983). Of the 17 cysteine substitutions previously reported (Kuivaniemi et al., 1991), no lethal phenotypes were noted with glycine replacement in the N-terminal half of the α1(I) chain (Figure 5). Also, no glycine replacement in the α2(I) chain was shown to decrease

markedly the thermal stability of the triple helix. The results here, however, provide the following conclusions about the position effects of cysteine for glycine substitutions: (a) substitutions near the N-terminus of the molecule do not necessarily produce milder phenotypes; (b) substitutions in the α2(I) chain do not necessarily produce milder phenotypes; and (c) substitutions in the α2(I) chain do not necessarily have less effect on the thermal stability of the triple helix than substitutions in the α1(I) chain. The results as a whole demonstrate that selective regions of the collagen molecule may have more specific roles in the biological function of the protein than previously realized (see Pack et al., 1989; Prockop, 1990; Steinmann et al., 1991).

This work was supported in part by grants from the March of Dimes-Birth Defects Foundation, NIH grant AR38188 and a grant from the Lucille P. Markey Charitable Trust.

REFERENCES

Bruckner, P. and Prockop, D. J. (1981) Anal. Biochem. **110**, 360–368
Byers, P. H. (1990) Trends Genet. **6**, 293–300
Byers, P. H., Wallis, G. A. and Willing, M. C. (1991) J. Med. Genet. **28**, 433–442
Gubler, U. and Hoffman, B. J. (1983) Gene **25**, 263–269
Kuivaniemi, H., Tromp, G. and Prockop, D. J. (1991) FASEB J. **5**, 2052–2060

Received 13 April 1992/6 July 1992; accepted 13 July 1992

Laemmli, U. K. (1970) Nature (London) **227**, 680–685
Laskey, R. R. and Mills, A. D. (1975) Eur. J. Biochem. **56**, 335–341
Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
Pack, M., Constantinou, C. D., Kalia, K., Nielsen, K. B. and Prockop, D. J. (1989) J. Biol. Chem. **264**, 19694–19699
Piez, K. A. (1984) in Extracellular Matrix (Piez, K. A. and Reddi, A. H., eds.), pp. 1–40, Elsevier, New York
Prockop, D. J. (1990) J. Biol. Chem. **265**, 15349–15352
Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. and Arnmein, N. (1985) Science **230**, 1350–1354
Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
Sillence, D. O. (1981) Clin. Orthop. Relat. Res. **159**, 11–25
Sillence, D. O., Senn, A. and Danks, D. M. (1979) J. Med. Genet. **16**, 101–116
Steinmann, B., Rao, V. H., Vogel, A., Bruckner, P., Gitzelmann, R. and Byers, P. H. (1984) J. Biol. Chem. **259**, 11129–11138
Steinmann, B., Westerhausen, A. I., Constantinou, C. D., Superti-Furga, A. and Prockop, D. J. (1991) Biochem. J. **279**, 747–752
Studencki, A. B. and Wallace, R. B. (1984) DNA **3**, 7–15
Superti-Furga, A. and Steinmann, B. (1988) Biochem. Biophys. Res. Commun. **150**, 140–147
Torre-Blanco, A., Adachi, E., Romanic, A. M. and Prockop, D. J. (1992) J. Biol. Chem. **267**, 4968–4973
Vogel, B. E., Doelz, R., Kadler, K. E., Hojima, Y., Engel, J. and Prockop, D. J. (1988) J. Biol. Chem. **263**, 19249–19255
Westerhausen, A. I., Constantinou, C. D. and Prockop, D. J. (1990) Nucleic Acids Res. **18**, 4968
Williams, C. J. and Prockop, D. J. (1983) J. Biol. Chem. **258**, 7721–7728