THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 278, No. 12, Issue of March 21, pp. 10006–10012, 2003
Printed in U.S.A.

Type I Collagen Triplet Duplication Mutation in Lethal Osteogenesis Imperfecta Shifts Register of α Chains throughout the Helix and Disrupts Incorporation of Mutant Helices into Fibrils and Extracellular Matrix*

Received for publication, December 9, 2002, and in revised form, January 6, 2003
Published, JBC Papers in Press, January 20, 2003, DOI 10.1074/jbc.M212523200

Wayne A. Cabral‡, Marianna V. Mertts§, Elena Makareeva§, Alain Colige¶, Mustafa Tekin‖, Arti Pandya‖, Sergey Leikin§, and Joan C. Marini‡**

From the ‡Section on Connective Tissue Disorders, HDB/NICHD, National Institutes of Health, Bethesda, Maryland 20892, §Section on Physical Biochemistry, NICHD, National Institutes of Health, Bethesda, Maryland 20892, ¶Laboratory of Connective Tissues Research, University of Liege, Liege, Belgium, and ‖Department of Human Genetics, MCV/VCU, Richmond, Virginia 23298

The majority of collagen mutations causing osteogenesis imperfecta (OI) are glycine substitutions that disrupt formation of the triple helix. A rare type of collagen mutation consists of a duplication or deletion of one or two Gly-*X*-*Y* triplets. These mutations shift the register of collagen chains with respect to each other in the helix but do not interrupt the triplet sequence, yet they have severe clinical consequences. We investigated the effect of shifting the register of the collagen helix by a single Gly-*X*-*Y* triplet on collagen assembly, stability, and incorporation into fibrils and matrix. These studies utilized a triplet duplication in COL1A1 exon 44 that occurred in the cDNA and gDNA of two siblings with lethal OI. The normal allele encodes three identical Gly-Ala-Hyp triplets at aa 868–876, whereas the mutant allele encodes four. The register shift delays helix formation, causing overmodification. Differential scanning calorimetry yielded a decrease in T_m of 2 °C for helices with one mutant chain and a 6 °C decrease in helices with two mutant chains. An *in vitro* binary co-processing assay of N-proteinase cleavage demonstrated that procollagen with the triplet duplication has slower N-propeptide cleavage than in normal controls or procollagen with proα1(I) G832S, G898S, or G997S substitutions, showing that the register shift persists through the entire helix. The register shift disrupts incorporation of mutant collagen into fibrils and matrix. Proband fibrils formed inefficiently *in vitro* and contained only normal helices and helices with a single mutant chain. Helices with two mutant chains and a significant portion of helices with one mutant chain did not form fibrils. In matrix deposited by proband fibroblasts, mutant chains were abundant in the immaturely cross-linked fraction but constituted a minor fraction of maturely cross-linked chains. The profound effects of shifting the collagen triplet register on chain interactions in the helix and on fibril formation correlate with the severe clinical consequences.

Osteogenesis imperfecta (OI)[1] is an autosomal dominant disorder of connective tissue. Its most significant clinical feature is skeletal fragility, causing the bones of affected individuals to be susceptible to fracture from minimal trauma or nontraumatic impact (1). Other symptoms of OI include short stature, blue sclerae, joint laxity, dentinogenesis imperfecta, and hearing loss (2). The severity of OI varies widely, ranging from perinatal lethal to barely detectable, as delineated by the Sillence classification (3).

The full clinical spectrum of OI is caused by defects in the structure or synthesis of type I collagen, the most abundant protein of the extracellular matrix of bone, skin, and tendon (4, 5). Defects in COL1A1 that result in the synthesis of half the normal amount of collagen cause the mildest form of the disease (OI type I) (6). The clinically significant forms of OI (OI types II, III, and IV) are caused by structural defects in either the α1(I) or α2(I) chains. Over 250 such mutations have now been delineated in individuals with OI (7, 8). The overwhelming majority (about 85%) are point mutations that result in the substitution of the glycine residue in a typical collagen tripeptide, Gly-*X*-*Y*, by another amino acid. Substitution mutations are thought to exert a detrimental effect on collagen function, because their side chains are larger than that of glycine and cause local interference with the folding of the triple helix (9). A smaller fraction (about 10%) of collagen mutations result in single exon skipping. These mutations maintain the Gly-*X*-*Y* triplet pattern but may cause local looping out of chains in the triple helix (4). An even less common set of structural mutations is located in the C-terminal propeptide. Since the C-propeptide is cleaved from the mature collagen molecule before incorporation into fibrils, the mutant region of the chain is not incorporated into matrix. Instead, they are thought to exert their effect by delaying the incorporation of the mutant chains into collagen trimer (10).

An additional rare and interesting group of mutations consists of deletions or duplications of the codons for one or two Gly-*X*-*Y* triplets. Only 10 cases of single triplet deletion or duplication have been reported (11–14), including five deletions and three duplications in the α1(I) chain and 2 deletions in the α2(I) chain. There are also four cases involving deletion or duplication of two triplets, all in the α2(I) chain (14, 15). These mutations are of special interest because they must disrupt

* 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.

** To whom correspondence and reprint requests should be addressed: Heritable Disorders Branch, Bldg. 10; Rm. 9s241, 9000 Rockville Pike, Bethesda, MD 20892. Tel.: 301-496-6683; Fax: 301-402-0234; E-mail: oidoc@helix.nih.gov.

[1] The abbreviations used are: OI, osteogenesis imperfecta; RT-PCR, reverse transcription-PCR; nt, nucleotides; DSC, differential scanning calorimetry; aa, amino acids; gDNA, genomic DNA.

collagen functioning by a mechanism quite different from that initiated by glycine substitutions. *A priori*, one might have expected mild functional defects from mutations that shift the Gly-*X*-*Y* register of the collagen helix by a single triplet unit rather than interrupt helix folding in the manner of a glycine substitution. In fact, small register shifts cause a lethal or severe phenotype. Determinations of helix stability and procollagen processing were reported for two lethal cases with deletion of one of the three Gly-Ala-Hyp triplets at aa 868–876 in α1(I) exon 44 (11, 12). These deletions decreased collagen helix T_m by only 0–1 °C. Processing of proband collagen by pericellular enzymes and purified N-proteinase was indistinguishable from normal, as was cleavage by vertebral collagenase. The processing data led investigators to propose that there was limited propagation of the register shift toward the N-terminal end of the procollagen trimer.

We report here a single triplet duplication in α1(I)E44 in siblings with lethal type II OI. Determinations of thermal stability and N-protease cleavage indicate that the register shift is propagated the full length of the collagen helix. *In vitro* fibrillogenesis and matrix deposition studies demonstrate that presence of the register shift impairs incorporation into fibrils and cross-linking into matrix. These studies provide new insight into the mechanisms of register shift mutations in collagen disorders.

MATERIALS AND METHODS

Clinical Cases—The probands were the male and female offspring of a 22-year-old gravida II para1 (G2P1) mother and a 25-year-old father, born at 32 and 37 weeks gestation, respectively. Prenatal ultrasound at 18–22 weeks gestation detected a short-limbed skeletal dysplasia in each child. The male child was delivered vaginally with forceps due to breech presentation. Weight was appropriate for age (2013 g), but crown to heel length was 38 cm (50% for 28 weeks gestation). At delivery, he had a soft skull with an anterior laceration, draining blood and cerebrospinal fluid, a narrow chest, and bowed extremities. He died 1 h after birth. The female child was born by spontaneous vaginal delivery. Birth weight was appropriate for age (2770 g), but length was short (43 cm; 50% for 32 weeks gestation). Deformities noted at birth included a soft cranium with mineralized bone only on lateral portions of the skull, blue sclerae, a high arched palate, and a narrow chest. Extremities had rhizomelic shortening and bowing and were abducted into an extreme frog-legged position. Radiograms showed multiple fractures of ribs and all long bones. The infant died at age 1 month of respiratory insufficiency.

Fibroblast Culture—Cultures established from dermal biopsies were grown in Dulbecco's modified Eagle's medium, with 10% serum and 2 mM glutamine. For large scale procollagen preparation, confluent cells of Proband 1, father, control, and α1(I) G832S, G898S, and G997S were treated with 50 μg/ml ascorbate in serum-free Dulbecco's modified Eagle's medium. Medium was harvested at 24-h intervals for 2 days and brought to 100 mM Tris-HCl, pH 7.4, 250 mM EDTA, 0.2% NaN_3, 1 mM phenylmethylsulfonyl fluoride, 5 mM benzamidine, and 10 mM *N*-ethylmaleimide. Ammonium sulfate-precipitated procollagen was collected by centrifugation. To label steady state procollagens, confluent cells were incubated without serum for 2 h with 50 μg/ml ascorbic acid and then 16 h with 260 μCi/ml 3.96 TBq/mmol L-[2,3,4,5-^{3}H]proline. Procollagens were harvested, and collagen was prepared as described (16). Isolated α1(I) chains were digested with cyanogen bromide (17).

Mutation Detection and Sequencing—Fibroblast RNA was isolated using Tri-Reagent (Molecular Research Center, Cincinnati, OH) (18). α1(I) exon 41–49 cDNA was obtained by reverse transcription polymerase chain reaction (RT-PCR) using 1 μg of RNA, 20 units of murine leukemia virus reverse transcriptase, and oligo(dT). The cDNA was amplified by PCR (19) using a sense primer corresponding to nt 2961–2990 (5′-ACTCCCGGGCCTCAAGGTATTGCTGGACAG-3′) and an antisense primer complementary to cDNA nt 3696–3725 (5′-GGG CAG GAA GCT GAA GTC GAA ACC AGC GCT-3′) and 2.5 units of Amplitaq (Invitrogen). Cycling conditions were 94 °C for 5 min; then 35 cycles of 94 °C for 1 min, 65 °C for 1 min, and 72 °C for 1 min; and finally 72 °C for 7 min. ^{32}P-Labeled antisense riboprobe was synthesized for the aa 787–1173 region of α1(I) collagen (20). Riboprobe and exon 41–49 cDNA were mixed for DNA:RNA hybrid analysis, as described (20).

For RT-PCR analysis, cDNA amplification used a sense primer in exon 43 (5′-CCTGGACGAGACGGTTCTCCTGGCGCCAAG-3′) and an antisense primer complementary to exon 45 (5′-GCCGACAGGACCG-GCGGGACCAGCAGGACC-3′). PCR conditions were 94 °C for 2 min; 30 cycles of 94 °C for 1 min, 69 °C for 30 s, and 72 °C for 30 s; and finally 72 °C for 7 min.

For genomic PCR, DNA was isolated from parental leukocytes and proband fibroblasts with the Puregene DNA Isolation Kit (Gentra Systems, Minneapolis, MN). Reactions used 500 ng of DNA and 1.0 units of Amplitaq. The sense primer corresponded to nt 13240–13269 of COL1A1 intron 43 genomic sequence (5′-TGACCCATATTCCCCTGCT-CTCCCCGCCAG-3′), and the antisense primer was complementary to nt 13378–13407 (5′-GGTACAGGGAACTGGAGCCCAGCTACTTAC-3′) in intron 44. PCR conditions were 94 °C for 5 min; then 30 cycles of 94 °C for 1 min, 65 °C for 30 s, and 72 °C for 30 s; and finally 72 °C for 7 min. RT-PCR and genomic PCR products were electrophoresed on 2% agarose gels and then visualized with ethidium bromide.

The exon 43–45 RT-PCR and intron 43–44 gDNA PCR products were subcloned and sequenced by the dideoxy chain termination method (21) with the Sequenase 2.0 kit (Amersham Biosciences). The sequencing primer for cDNA subclones corresponded to cDNA nt 3189–3218 in α1(I) exon 43 (5′-CCTGGACGAGACGGTTCTCCTGGCGCCAAG-3′). The sequencing primer for genomic DNA subclones corresponded to COL1A1 nt 13240–13269 in intron 43 (5′-TGACCCATATTCCC-CTGCTCTCCCCGCCAG-3′).

Matrix Deposition—Confluent fibroblasts were stimulated every other day for 9 days with 100 μg/ml ascorbic acid and then incubated for 24 h with 260 μCi/ml [^{3}H]proline in serum-free medium. Procollagens in media were precipitated with ammonium sulfate. Matrix collagens were serially extracted, as described (22). In brief, newly synthesized collagens were extracted for 24 h with neutral salt (0.15 M NaCl in 50 mM Tris-HCl, pH 7.5). Collagens with acid-labile cross-links were extracted for 24 h with 0.5 M acetic acid. Collagens with mature cross-links were extracted by pepsin digestion (0.1 mg/ml) for 24 h. All matrix fractions were precipitated with 2 M NaCl.

Matrix Chase—Confluent fibroblasts were stimulated every other day for 9 days with 100 μg/ml ascorbic acid, incubated for 48 h with 260 μCi/ml [^{3}H]proline in serum-free medium, and then chased with fresh Dulbecco's modified Eagle's medium containing 10% fetal bovine serum and 10 mM nonradioactive proline. Individual cultures were harvested at 24-h intervals for 5 days, and the matrix layer was processed with protease inhibitors, as described (22). Matrix extracts were resuspended in 0.5 M acetic acid and digested overnight with pepsin. Collagens were precipitated with 2 M NaCl.

Preparation of Fluorescent Labeled Procollagen—Ammonium sulfate protein precipitates were redissolved in 0.1 M sodium carbonate, 0.5 M NaCl (pH 9.3) at 0.2 mg/ml collagen concentration. Cy2 and monoreactive Cy5 dyes (Amersham Biosciences) (23, 24) were dissolved in 1 ml of anhydrous dimethylformamide, according to product directions. Procollagen was added to lyophilized dye aliquots, 100 μl of protein solution/10 μl of dye aliquot for Cy5, and 100 μl of protein solution/50 μl of dye aliquot for Cy2 labeling, shaken for 30 min at room temperature, analyzed by SDS-PAGE for labeling efficiency, frozen on dry ice, and stored at −80 °C.

N-proteinase Cleavage—Labeled procollagen was transferred into 50 mM Tris, 0.5 M NaCl, 4 mM $CaCl_2$, 0.5 mM phenylmethylsulfonyl fluoride, 2.5 mM *N*-ethylmaleimide, 0.02% Brij 35 (pH 8) on AutoSeq G-50 microspin columns (Amersham Biosciences). Collagen concentration was adjusted to 0.1 mg/ml. All possible binary mixtures of Cy2- and Cy5-labeled procollagens were prepared (*e.g.* C-Cy2/C-Cy5, C-Cy2/OI-INS-Cy5, C-Cy5/OI-INS-Cy2, and OI-INS-Cy2/OI-INS-Cy5 were made) to compare cleavage kinetics of C and OI-INS. N-proteinase (25) was added on ice, and mixtures were placed at 34 °C. Sample aliquots were collected at different times after the start of the reaction, mixed with a lithium dodecyl sulfate gel sample buffer (Invitrogen) with added dithiothreitol and EDTA, and rapidly frozen. Samples were analyzed on precast 6% Tris/glycine and 7% Tris acetate minigels (Invitrogen). The gels were scanned on a FLA3000 fluorescence scanner (Fuji Medical Systems, Stamford, CT) at 50 × 50-μm resolution. Fractions of the cleaved proteins were determined from band intensities using PeakFit software (SPSS Inc., Chicago, IL), corrected for the cleaved protein in the initial mix and for the effect of the fluorescent label on the cleavage by using control mixtures (*e.g.* C-Cy2/C-Cy5 and OI-INS-Cy2/OI-INS-Cy5 for the C *versus* OI-INS experiment). The effect of the fluorescent label did not exceed 10%. Experiments were repeated in triplicate.

Preparation of Full-length Collagen by N- and C-proteinase Cleavage—Ammonium sulfate procollagen precipitate was doped with 10% Cy5-labeled procollagen. The mixture was chromatographed on two

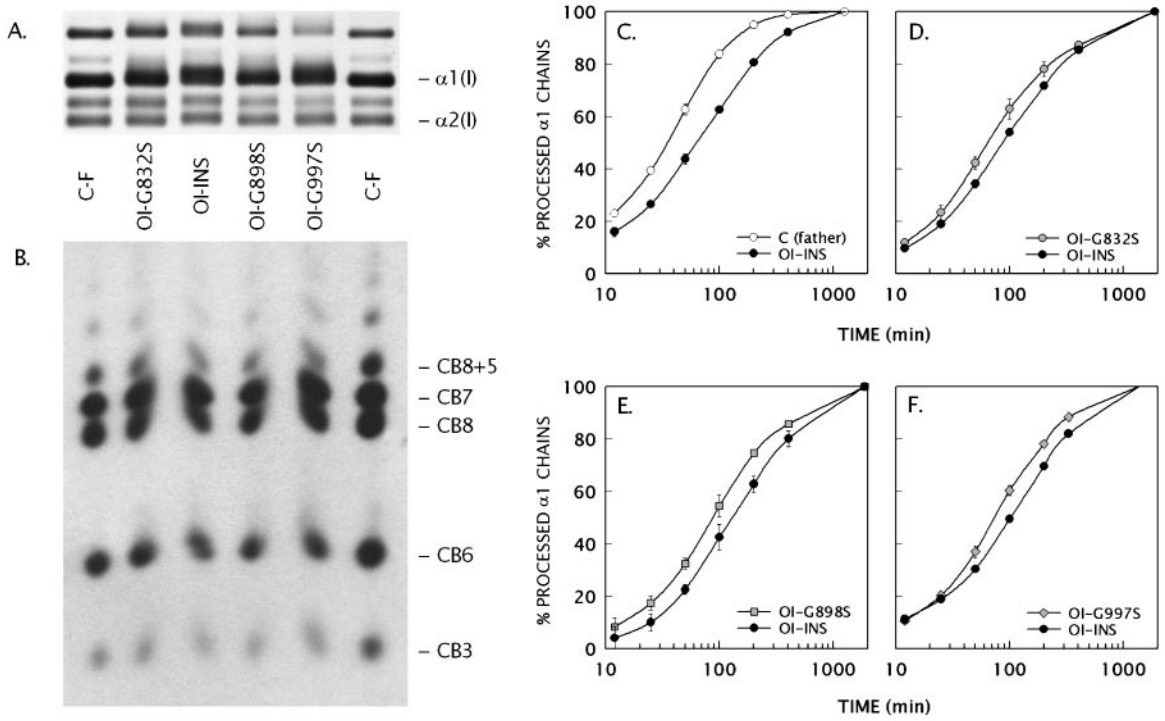

FIG. 1. **Proband type I collagen and kinetics of *in vitro* N-propeptide cleavage by N-protease.** *A*, a similar extent of posttranslational overmodification was observed in the insertion mutation (OI-INS) and glycine substitution mutations in the region of α1(I) surrounding the insertion (OI-G832S, OI-G898S, and OI-G997S), as indicated by broad α1(I) bands on SDS-urea-PAGE. *B*, SDS-urea-PAGE of α1(I) CNBr peptides of these mutant collagens showed closely matched overmodification, especially for the N-terminal ends of OI-INS and OI-G997 (note CB8 and CB8 + 5). *C–F*, binary co-processing assays yielded slower N-propeptide cleavage in OI-INS than in control and glycine substitution mutations. Each point on the curve is an average of two co-processing experiments with inverted fluorescent label (*e.g.* C-F show the average of data for C-F-Cy2/OI-INS-Cy5, and C-F-Cy5/OI-INS-Cy2 binary mixtures). *Error bars* indicate S.D. in these experiments.

1.6 × 5-cm columns of DEAE-cellulose (DE52; Whatman) as described (26–28). First, the mixture was loaded in 2 M urea, 0.15 M NaCl, 0.1 M Tris-HCl (pH 7.4) and eluted with the same buffer. Second, the mixture was loaded in 2 M urea, 0.1 M Tris-HCl (pH 8.6) and eluted with an NaCl gradient. Fractions containing procollagen on SDS-PAGE were pooled and concentrated by pressure ultrafiltration through an Amicon YM30 membrane. Procollagen was transferred into 50 mM Tris, 0.5 M NaCl, 4 mM $CaCl_2$, 0.5 mM phenylmethylsulfonyl fluoride, 2.5 mM *N*-ethylmaleimide, 0.02% Brij 35 (pH 8) and simultaneously digested by N- and C-proteinase at 32 °C. C-proteinase was the generous gift of Prof. K. E. Kadler (University of Manchester, Manchester, UK). Cleavage was monitored by SDS-PAGE. Complete digestion was observed at 40–70 h, and the reaction was stopped by the addition of EDTA to a final concentration of 20 mM. Collagen was precipitated twice by 0.6 M NaCl in 0.5 M acetic acid and analyzed for purity by SDS-PAGE.

Differential Scanning Calorimetry—DSC scans from 10 to 50 °C were performed at 0.125 to 1 °C/min heating rates in a Nano II DSC instrument (Calorimetry Sciences Corp., American Fork, UT) as described (29). Pepsin digestion (~1:10 pepsin/collagen) of ammonium sulfate precipitates in 0.5 M acetic acid at 4 °C overnight resulted in complete removal of N- and C-propeptides. To prevent fibrillogenesis, 0.1–1.2 mg/ml procollagen, full-length collagen, or pepsin-treated collagen solutions in either 2 mM HCl (pH 2.7) or 0.2 M sodium phosphate, 0.5 M glycerol (pH 7.4) were used. The denaturation temperature (T_m) in phosphate/glycerol buffers was used to extrapolate T_m to physiological conditions (29), but better resolution of mutant collagen forms was achieved in 2 mM HCl.

In Vitro Fibrillogenesis—At 5 °C, full-length collagen in 2 mM HCl (pH 2.7) was mixed 1:1 with 0.26 M NaCl, 6 mM sodium phosphate (final pH 6.9), transferred into a precooled quartz cuvette (1-cm path length), and placed into a V-560 spectrophotometer (Jasco Inc., Easton, MD) equipped with a thermoelectric temperature controller. Fibrillogenesis was initiated by a temperature jump to 36.6 °C and monitored by turbidity measurement (optical density at 350 nm, A_{350}) (30). After A_{350} reached saturation, fibers were pelleted (5 min at 10,000–14,000 × *g*) and redissolved in 2 mM HCl for subsequent analysis by DSC. Supernatant was dialyzed against 2 mM HCl for DSC analysis. Collagen concentration before fibrillogenesis and in supernatant after fibrillogenesis were measured by Sircol assay (Biocolor Ltd., Belfast, Northern Ireland). Aliquots from fibers and supernatant were labeled with Cy5 and analyzed on 3–8% Tris acetate or 4–12% Bis-Tris minigels (Invitrogen).

RESULTS

Collagen Protein Analysis—The type I collagen produced by the cultured dermal fibroblasts of both probands displayed identical electrophoretic abnormalities on SDS-urea-PAGE. The α1(I) chain was doubled in width, consisting of a normal and an electrophoretically delayed form (Fig. 1*A*). The α2(I) band was slightly broadened with a delayed base line. The abnormal collagen was secreted from the cell as efficiently as normal collagen. All CNBr peptides from the α1(I) chain of the proband showed both normal and electrophoretically delayed forms (Fig. 1*B*). Since the α1(I) chains were overmodified along their full length, the mutation was localized to the COOH-terminal quarter of either α chain. The mother had a lighter, less tightly resolved band just above the α1(I) band in cell layer collagen (not shown). This pattern is characteristic of a low percentage of α1(I) chains with excess posttranslational modification. Given the clinical information that the unaffected parents have two children with a dominant genetic disorder, this finding prompted further investigation of the mother as a likely mosaic carrier of the collagen mutation.

Mutation Identification in cDNA and gDNA—The α1(I) and α2(I) cDNA coding for the C-terminal quarter of the collagen helix were examined by RNA:DNA hybrid analysis (20). Using RNase A, we detected a mismatch in the cDNA coding for exons 41–49 of the α1(I) chain.

RT-PCR screening of α1(I) cDNA localized the mismatch to exon 44 (Fig. 2*A*). Both normal and more slowly migrating products were detected in the probands' cDNA. The more slowly migrating product was faintly visible in the mother's sample. This electrophoretically slower product was shown to

be a heteroduplex of normal and mutant fragments. The small fraction of mutant α1(I) transcripts in maternal cells can be easily visualized because of the sensitivity of heteroduplex analysis for structurally distinct products. The localization of the collagen mutation to exon 44 was confirmed by PCR amplification of genomic DNA (Fig. 2*B*).

Sequencing of subclones of proband cDNA exons 43–45 and proband and mother gDNA intron 43-intron 44 revealed the same relatively unusual type of collagen mutation, confirming the mother as a mosaic carrier. The mutant allele has a 9-bp insertion (5′-GGT GCT CCT-3′), coding exactly for an extra Gly-Ala-Hyp triplet (Fig. 3). The insertion is a duplication in a highly repetitive region. The normal allele has two identical 9-bp sequences coding for aa 868–873 and an adjacent 9 bp that differs by only one nucleotide. The mutant allele has three of the identical 9-bp units.

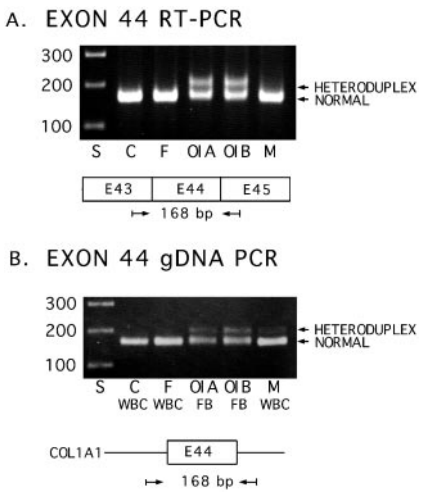

FIG. 2. **PCR screening of proband and parental cDNA and genomic DNA.** *A*, cDNA screening of proband and parental fibroblasts. Normal and heteroduplex products were observed in α1(I)E44 in the samples of both probands (*lanes 4* and *5*) and, to a lesser extent, in the samples of the mother (*lane 6*). *B*, screening of genomic DNA from control and parental leukocytes and proband fibroblasts. Normal and heteroduplex products were obtained from the probands (*lanes 4* and *5*) and mother (*lane 6*).

Effect of Register Shift on Collagen Thermal Stability—Differential scanning calorimetry thermograms of proband collagen (OI-INS) were done at acidic and neutral pH (Fig. 4). Since mutant collagen has three species ($\alpha1_2\alpha2$, $\alpha1(ins)\alpha1\alpha2$, and $\alpha1(ins)_2\alpha2$), up to three different peaks may be expected on thermograms. All three peaks are clearly visible at acidic pH, compared with the single normal peak of the control (father (C-F), Fig. 4). Based on relative intensity ratios determined by Gaussian deconvolution, 0.25 (35 °C):0.5 (39 °C):0.25 (41 °C), these peaks correspond to molecules with two, one, and no mutant α1 chains, respectively. Only two peaks with the melting temperature (T_m) difference of ~2 °C and a long low temperature tail can be resolved at neutral pH. The difference between thermograms at neutral and acidic pH and the change in the overall T_m of the triple helix suggest that the register shift propagates through the triple helix.

Propagation of Register Shift Affects N-propeptide Processing of Mutant Procollagen—We compared the kinetics of N-propeptide cleavage in the insertion mutation (OI-INS), paternal control collagen (C-F), unrelated control collagen (C), and three-glycine substitution mutations in the region of α1(I) around the insertion, G832S (31), G898S, and G997S.[2] These Gly → Ser substitutions should have post-translational modification similar to that of the triplet duplication. By high pressure liquid chromatography, we determined that lysine hydroxylation of all mutant collagens was about double that of control: 21 ± 0.5% in C-F, 43 ± 1.5% in OI-INS, 40 ± 1.0% in G832S, 39 ± 0.5% in G898S, and 38 ± 1.0% in G997S. Gel migration of α1(I) chains (Fig. 1*A*) indicated that all mutations had a similar extent of lysine glycosylation. Most importantly, the N-terminal CB 8 + 5 peptide of OI-INS and OI-G997S had identical

[2] W. A. Cabral, H. Nishioka, and J. C. Marini, unpublished data.

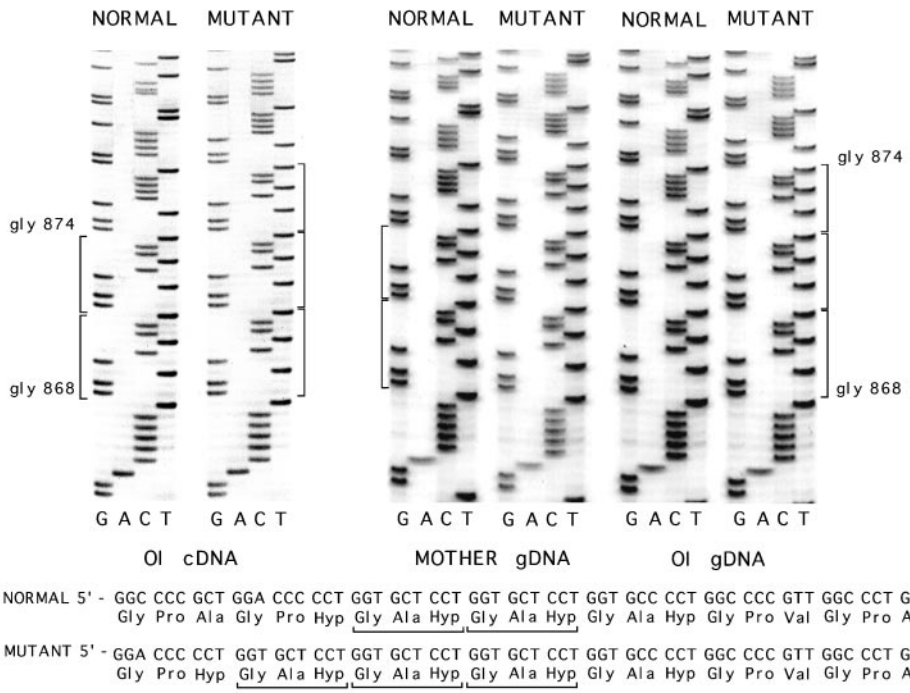

FIG. 3. **Sequencing of normal and mutant cDNA and genomic DNA alleles.** Sequencing of cDNA and genomic DNA revealed a 9-bp duplication in COL1A1, which is also present in DNA from the probands' mother. The normal allele encodes two identical Gly-*X*-*Y* triplets (GGT GCT CCT) at amino acids 868–873, whereas the mutant allele contains three.

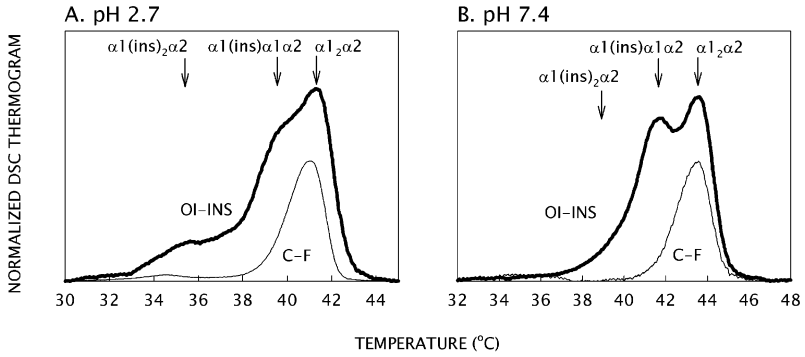

FIG. 4. **Normalized DSC thermograms.** *A*, DSC of proband collagen (OI-INS (*thick line*)) and normal control protein from his father (C-F (*thin line*)) in 2 mM HCl, pH 2.7. *B*, DSC in 0.2 M sodium phosphate, 0.5 M glycerol, pH 7.4. Scans were performed at 0.125 °C/min. Phosphate/glycerol buffer was used to prevent fibrillogenesis at neutral pH. Collagen T_m in this buffer is 1.7 °C higher than in physiological solution (29).

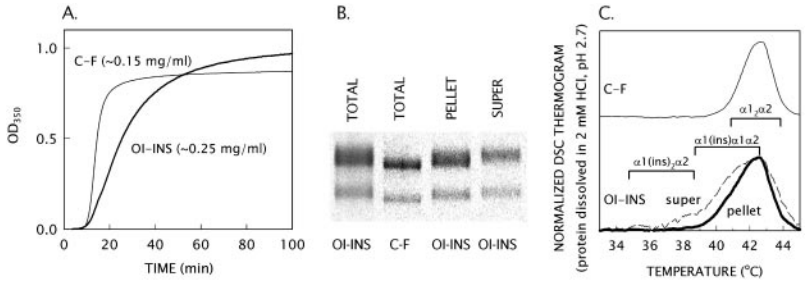

FIG. 5. ***In vitro* fibrillogenesis.** *In vitro* fibrillogenesis of proband collagen (OI-INS) and normal control protein from his father (C-F). *A*, kinetics of fiber formation by collagen prepared with N- and C-proteinase monitored by turbidity measurement. *B*, SDS-PAGE of total protein and fibrillogenesis-capable (*PELLET*) and -incapable (*SUPER*) fractions of OI-INS and C-F collagen. *C*, DSC thermograms of redissolved pellets and supernatant after *in vitro* fibrillogenesis. Each solution was dialyzed against 2 mM HCl (pH 2.7). The scans were performed in 2 mM HCl at 0.3 °C/min heating rate. *Brackets* indicate expected positions of denaturation peaks for triple helices containing no, one, and two mutant α1(I) chains, based on DSC thermogram shown in Fig. 4*A*. The slightly higher T_m values than seen in Fig. 4*A* relate to faster scanning in this thermogram (29).

glycosylation (Fig. 1*B*), so the effect of overmodification on cleavage of N-propeptide from these mutant procollagens should be the same.

In order to detect small differences in cleavage kinetics, we analyzed all possible binary combinations in a co-processing assay. In each pair, two proteins with different fluorescent labels were processed in the same test tube by N-protease under identical conditions, co-electrophoresed on SDS-PAGE, and distinguished by fluorescence scanning. This assay yielded reproducible detection of differences in cleavage kinetics as small as 5–10%. The results of N-propeptide co-processing of different mutant procollagen pairs is shown in Fig. 1, *C–F*. The triplet insertion causes substantially slower α1 N-propeptide cleavage than from control or glycine substitutions with equivalent overmodification. Thus, the decrease in kinetics of N-propeptide cleavage in the insertion mutation is related to the register shift *per se*. This supports the interpretation of the DSC thermograms that the register shift persists through the entire helix.

In Vitro Fibrillogenesis of Mutant Collagen—Classical fibrillogenesis kinetics, a lag phase followed by rapid fiber growth (30), was observed for both mutant and wild type collagens (Fig. 5*A*). Mutant collagen formed fibrils more slowly than control and required a higher initial concentration to achieve a similar extent of fiber formation. Virtually all control collagen formed fibers, since collagen was not detected in the fibrillogenesis supernatant. In contrast, over 20% of molecules remained in the supernatant after fibrillogenesis of proband collagen. The fraction of mutant collagen that did not form fibrils contained helices with a greater extent of posttranslational overmodification (Fig. 5*B*) and, therefore, a higher content of molecules containing mutant chains.

DSC thermograms (Fig. 5*C*) demonstrated that only molecules with no or one mutant chain were incorporated into fibers *in vitro*. The melting peak corresponding to helices with two mutant chains is totally missing from thermograms of fibers; the shoulder corresponding to helices with a single mutant chain is substantially reduced. All molecules with two mutant chains and a significant fraction of molecules with one mutant chain remained in the supernatant, explaining the higher extent of posttranslational modification in the supernatant collagen. By circular dichroism,[3] we found that many of the molecules with two mutant chains were irreversibly denatured during fibrillogenesis because of their extreme instability. The main peak of the supernatant thermogram is a mixture composed mostly of molecules with a single mutant chain and a small fraction of molecules with no mutant chain.

Matrix Deposition—The incorporation of proband and control collagen into matrix was compared by serial extractions of the matrix deposited by cultured cells (Fig. 6). The proband overmodified chains are equally abundant in the media and the neutral salt extract (fraction 1), the later containing helices that are not cross-linked with other matrix molecules. The mutant chains are also efficiently incorporated into the immaturely cross-linked fraction (fraction 2). However, they are substantially less abundant in the maturely cross-linked fraction extracted with pepsin (fraction 3), which has predominantly normally migrating α1(I) chains. Since helices with two mutant chains are resistant to fibril incorporation in the *in vitro* assay, the overmodified α1(I) chains in the pepsin extracts are more likely derived from helices with one mutant α1(I) chain.

Matrix Chase—A pulse-chase experiment examined the stability of collagen deposited in matrix by cultured proband and control cells. Matrix stability was not significantly altered (Fig. 7). The probands' normal and overmodified α1(I) chains could not be quantitated separately, but an equivalent proportion of overmodified α1(I) chains is visible in each proband sample.

[3] E. Makareeva and S. Leikin, unpublished data.

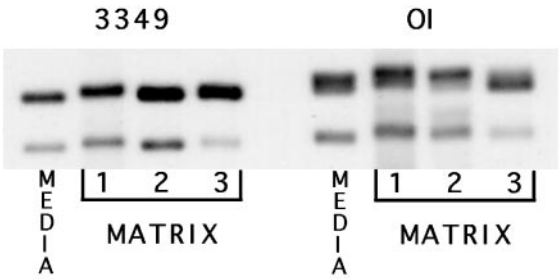

FIG. 6. **Incorporation of proband collagen into extracellular matrix.** Sequential extraction of type I collagen incorporated in matrix by control (*3349*) and proband fibroblasts. Media collagen was digested with pepsin. Matrix was extracted serially, first with NaCl to isolate newly incorporated collagen without cross-links (matrix fraction 1) and then with acetic acid for immaturely cross-linked collagen (matrix fraction 2) and finally with pepsin to release fully cross-linked collagen (matrix fraction 3). Fractions were analyzed by 6% SDS-urea-PAGE.

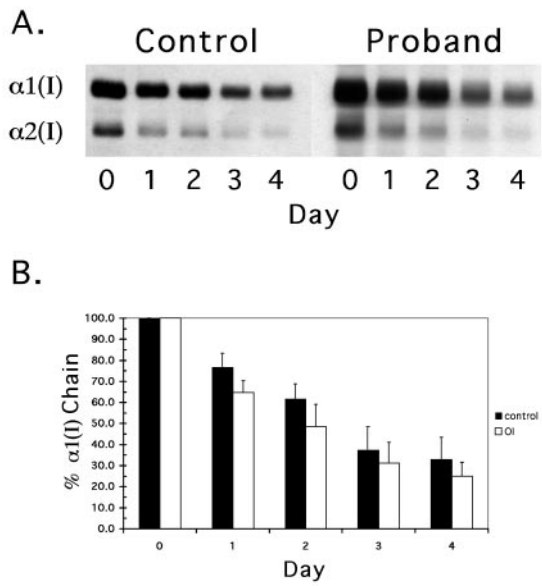

FIG. 7. **Turnover of collagen incorporated into matrix in culture.** *A*, labeled collagen was allowed to incorporate into extracellular matrix deposited in culture by postconfluent control and proband fibroblasts. Matrices were collected at 24-h intervals, and collagens were extracted by pepsin digestion. Samples were analyzed by 6% SDS-urea-PAGE. *B*, the experiment was done in triplicate and quantitated by densitometry of autoradiograms.

DISCUSSION

We have described here a novel single triplet duplication in the type I collagen α1(I) chain and its functional consequences for helix formation and fibrillogenesis. The mutation occurs in two siblings with lethal type II osteogenesis imperfecta. Their mother is a mosaic carrier with a low percentage of heterozygous fibroblasts and leukocytes, 10 and 15%, respectively.[4] Her clinical history and physical exam are entirely normal.

The mutant COL1A1 allele has a 9-bp duplication in exon 44, which has a highly repetitive sequence. In the normal allele, there are two consecutive 5′-GGT GCT CCT-3′ units at nt 3255–3272, followed by 9 bp that differ by a single nucleotide, 5′-GGT GCC CCT-3′. These 27 bp code for three consecutive Gly-Ala-Hyp collagen triplets, at aa 868–876. In the mutant allele, a duplication of one 5′-GGT GCT CCT-3′ unit results in an extra Gly-Ala-Hyp triplet. The repetitive sequence of exon 44 has made it a hot spot for single triplet deletions and duplications. Nine of the 11 known triplet deletions or duplications in α1(I) (11–14) have occurred in this region.

Triplet duplication and deletion mutations form a very interesting and relatively unusual set of mutations causing OI. The more prevalent glycine substitution mutations disrupt the otherwise uninterrupted Gly-*X*-*Y* triplet repeats of type I collagen. Accommodation of the substituting amino acid in the internal aspect of the helix delays helix formation. Triplet mutations do not interrupt the Gly-*X*-*Y* sequence; they shift the "register" of the chains with respect to each other. The severe to lethal phenotype of all cases of OI with this type of mutation indicates the significance of the register shift for the structure and interactive functions of the collagen helix. Functional studies on type I collagen helix formation and structure have been published for two single triplet deletions in exon 44 (11, 12). Both cases have equivalent triplet deletions, with two consecutive Gly-Ala-Hyp triplets present, rather than the usual three.

All triplet deletion and duplication cases, including the one described here, are associated with significant overmodification. This suggests that accommodation of shifted register interactions between different *X* and *Y* residues along the helix is causing delay in helix formation. For the duplication presented here, our findings support propagation of the register shift toward the N-terminal end of the helix rather than formation of a loop accommodating the extra triplet.

The global register shift may affect the stability of the entire helix. The T_m of the collagen with a triplet deletion in exon 44 was found to be decreased by 0–1 °C (11, 12). In our triplet duplication mutation, the T_m of helices with one mutant chain was reduced 2 °C, whereas two mutant chains reduced T_m by 6 °C. Although T_m was determined in the deletion cases by the less sensitive trypsin-chymotrypsin digestion method, the 6 °C drop in T_m with two mutant chains should not have been missed and probably represents a more deleterious effect of the duplication. Deletions and duplications result in register shifts of opposite direction and may cause unequal effects on T_m. Alternatively, the deletion mutations may compensate by local looping out. Also, note that the thermal stability of $\alpha1(ins)_2\alpha2$, in which the α2 chain is out of register with two longer α1(ins) chains, is much lower than that of $\alpha1(ins)\alpha1\alpha2$, in which one α1(ins) chain is out of register with the remaining two chains. Proper register of the α2 chain appears to be particularly important for type I collagen stability.

Furthermore, the difference between the thermograms of the triplet duplication mutation at neutral and acidic pH also supports a longer range effect of the duplication on helix folding. The change in relative stability of mutant and normal forms at different pH suggests that intramolecular salt bridges are different in these species. Since the mutation occurs in a stretch containing no salt bridges, this indicates that the register shift propagates at least three triplets beyond the mutation to the first potential salt bridge in the direction of the N-terminal end.

Propagation of the register shift for the entire length of the helix may affect the kinetics of N-propeptide cleavage (32–34). The duplication mutation showed a decrease in cleavage rate with respect to two control collagens. We demonstrated that the delay was not simply the result of overmodification along the length of the helix by comparing the duplication with glycine substitution mutations in the same region of α1(I). Although overmodification may delay propeptide processing, our co-digestion assay demonstrated that propagation of the register shift along the entire helix must also contribute to the slower cleavage kinetics of the insertion mutation.

The triplet duplication also alters fibril formation *in vitro* and in culture. Collagen molecules with two mutant chains are not incorporated into fibrils *in vitro*, and those with one mutant

[4] W. A. Cabral and J. C. Marini, unpublished data.

chain are poorly incorporated. In addition, the critical concentration for fibrillogenesis is higher for proband than for normal collagen. Improper helix melting may be one of the reasons for poor incorporation of mutant molecules into fibrils. At both the 36.6 °C of *in vitro* fibrillogenesis and the 37.5 °C of body temperature, molecules with two mutant α1 chains may have melted and be unavailable for incorporation. Indeed, we observed a substantial decrease in the circular dichroism signal characteristic of a collagen triple helix (35) upon fibrillogenesis of the insertion mutation, in contrast to enhanced circular dichroism upon fibrillogenesis of control protein.

The matrix deposition studies show that collagen molecules with a single mutant chain are incorporated less efficiently than normal helices. Mutant chains are present proportionately in the media and the non-cross-linked fraction of matrix. They are relatively retained in the immaturely cross-linked fraction, with slow progression into the maturely cross-linked pepsin extract. Delayed cross-linking of mutant molecules is probably related to misalignment of residues on opposite molecules, which would be expected from propagation of the register shift. The delay in cross-linking of the triplet duplication is not seen with α2(I)ΔE16 collagen (36), which causes a larger six-triplet register shift and is more likely to realign by "looping out" of the normal chains than to propagate the register shift along the full helix.[4]

Proband mature matrix has a turnover that is comparable with normal, reflecting both its predominantly normal collagen composition and also the stability of the cross-links formed by helices with one mutant chain that have become fully incorporated. Collagen helices containing mutant α chains often have preferential intracellular breakdown and decreased secretion and sometimes have decreased incorporation into matrix. Matrix turnover of mutant collagen has been studied in only a few additional cases and was reported to have a shorter half-life than in control in a case of lethal OI with α1(I) G667R (22) and a half-life comparable with controls in cases of severe type III OI with α1(I) G589S and α2(I) G586V (37).

Although a register shift mutation does not disturb the uninterrupted Gly-*X*-*Y* triplet repeat of collagen, the altered alignments of *X* and *Y* residues along the chain have profound effects. Collagen formation is delayed, and its stability is decreased along the entire length of the helix. Double mutant and most single mutant helices are not incorporated into collagen fibrils, because of a combination of extreme instability and interference of the register shift with mature cross-linking. In addition to the deficiency of matrix that results from decreased fibrillogenesis, it is likely that misalignment of the collagen chains along the full helical region will also disrupt many of the interactions of collagen with noncollagenous molecules in matrix. Recent studies of a short synthetic heterotrimer containing the integrin-binding epitope of type IV collagen showed a strong effect of chain register on helix conformation (38) and integrin binding (39), supporting the proposal that register shifts may alter functionally significant side chain interactions. The combination of changes in helix stability, interchain X and Y position alignment, fibrillogenesis, and cross-linking results in the severe clinical phenotype of these mutations.

Acknowledgments—We thank Dr. Karl Kadler for the generous gift of C-proteinase. We acknowledge Michael Bonidie, M.D., and Patricia D. Gonzales, M.D., for their astute insistence that samples be obtained from the probands in their local hospital and for referral of the cases. We are grateful to the probands' mother for interest and continued participation in OI research efforts.

REFERENCES

1. Van der Hoeve, J., and de Kleyn, A. (1918) *Arch. Ophthal.* **95,** 81–93
2. Marini, J. C., and Chernoff, E. J. (2001) in *Management of Genetic Syndromes* (Cassidy, S. B., and Allanson, J. E., eds) pp. 281–300, Wiley-Liss, New York
3. Sillence, D. O., Senn, A., and Danks, D. M. (1979) *J. Med. Genet.* **16,** 101–116
4. Byers, P. H. (1993) in *Connective Tissue and Its Heritable Disorders* (Royce, P. M., and Steinmann, B., eds) pp. 317–350, Wiley-Liss, New York
5. Prockop, D. J., and Kivirikko, K. I. (1995) *Annu. Rev. Biochem.* **64,** 403–434
6. Willing, M. C., Deschenes, S. P., Scott, D. A., Byers, P. H., Slayton, R. L., Pitts, S. H., Arikat, H., and Roberts, E. J. (1994) *Am. J. Hum. Genet.* **55,** 638–647
7. Dalgleish, R. (1997) *Nucleic Acids Res.* **25,** 181–187
8. Dalgleish, R. (1998) *Nucleic Acids Res.* **26,** 253–255
9. Byers, P. H., Wallis, G. A., and Willing, M. C. (1991) *J. Med. Genet.* **28,** 433–442
10. Pace, J. M., Kuslich, C. D., Willing, M. C., and Byers, P. H. (2001) *J. Med. Genet.* **38,** 443–449
11. Hawkins, J. R., Superti-Furga, A., Steinmann, B., and Dalgleish, R. (1991) *J. Biol. Chem.* **266,** 22370–22374
12. Wallis, G. A., Kadler, K. E., Starman, B. J., and Byers, P. H. (1992) *J. Biol. Chem.* **267,** 25529–25534
13. Ward, L. M., Lalic, L., Roughley, P. J., and Glorieux, F. H. (2001) *Hum. Mutat.* **17,** 434
14. Pace, J. M., Atkinson, M., Willing, M. C., Wallis, G. A., and Byers, P. H. (2001) *Hum. Mutat.* **18,** 319–326
15. Horwitz, E. M., Prockop, D. J., Fitzpatrick, L. A., Koo, W. W., Gordon, P. L., Neel, M., Sussman, M., Orchard, P., Marx, J. C., Pyeritz, R. E., and Brenner, M. K. (1999) *Nat. Med.* **5,** 309–313
16. Bonadio, J., Holbrook, K. A., Gelinas, R. E., Jacob, J., and Byers, P. H. (1985) *J. Biol. Chem.* **260,** 1734–1742
17. Nicholls, A. C., Oliver, J., Renouf, D. V., Heath, D. A., and Pope, F. M. (1992) *Hum. Genet.* **88,** 627–633
18. Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159
19. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) *Science* **239,** 487–491
20. Grange, D. K., Gottesman, G. S., Lewis, M. B., and Marini, J. C. (1990) *Nucleic Acids Res.* **18,** 4227–4236
21. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
22. Bateman, J. F., and Golub, S. B. (1994) *Matrix Biol.* **14,** 251–262
23. Southwick, P. L., Ernst, L. A., Tauriello, E. W., Parker, S. R., Mujumdar, R. B., Mujumdar, S. R., Clever, H. A., and Waggoner, A. S. (1990) *Cytometry* **11,** 418–430
24. Mujumdar, R. B., Ernst, L. A., Mujumdar, S. R., Lewis, C. J., and Waggoner, A. S. (1993) *Bioconj. Chem.* **4,** 105–111
25. Colige, A., Beschin, A., Samyn, B., Goebels, Y., Van Beeumen, J., Nusgens, B. V., and Lapiere, C. M. (1995) *J. Biol. Chem.* **270,** 16724–16730
26. Fiedler-Nagy, C., Bruckner, P., Hayashi, T., and Prockop, D. J. (1981) *Arch. Biochem. Biophys.* **212,** 668–677
27. Fiedler-Nagy, C., Bruckner, P., Hayashi, T., Fietzek, P. P., and Prockop, D. J. (1982) *J. Biol. Chem.* **257,** 9181–9188
28. McBride, D. J., Kadler, K. E., Hijima, Y., and Prockop, D. J. (1992) *Matrix* **12,** 256–263
29. Leikina, E., Mertts, M. V., Kuznetsova, N., and Leikin, S. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 1314–1318
30. Williams, B. R., Gelman, R. A., Poppke, D. C., and Piez, K. A. (1978) *J. Biol. Chem.* **253,** 6578–6585
31. Marini, J. C., Grange, D. K., Gottesman, G. S., Lewis, M. B., and Koeplin, D. (1989) *J. Biol. Chem.* **264,** 11893–11900
32. Vogel, B. E., Doelz, R., Kadler, K. E., Hojima, Y., Engel, J., and Prockop, D. J. (1988) *J. Biol. Chem.* **263,** 19249–19255
33. Dombrowski, K. E., and Prockop, D. J. (1988) *J. Biol. Chem.* **263,** 16545–16552
34. Vogel, B. E., Minor, R. R., Freund, M., and Prockop, D. J. (1987) *J. Biol. Chem.* **262,** 14737–14744
35. Tiffany, M. L., and Krimm, S. (1969) *Biopolymers* **8,** 347–359
36. Filie, J. D., Orrison, B. M., Wang, Q., Lewis, M. B., and Marini, J. C. (1993) *Hum. Mutat.* **2,** 380–388
37. Forlino, A., Zolezzi, F., Valli, M., Pignatti, P. F., Cetta, G., Brunelli, P. C., and Mottes, M. (1994) *Hum. Mol. Genet.* **3,** 2201–2206
38. Sacca, B., Renner, C., and Moroder, L. (2002) *J. Mol. Biol.* **324,** 309–318
39. Sacca, B., Sinner, E.-K., Kaiser, J., Lubken, C., Eble, J. A., and Moroder, L. (2002) *Chem. Biochem.* **9,** 904–907