Defective splicing of mRNA from one COL1A1 allele of type I collagen in nondeforming (type I) osteogenesis imperfecta.

M L Stover, … , M B McKinstry, D W Rowe

J Clin Invest. 1993;92(4):1994-2002. https://doi.org/10.1172/JCI116794.

Research Article

Osteogenesis imperfecta (OI) type I is the mildest form of heritable bone fragility resulting from mutations within the COL1A1 gene. We studied fibroblasts established from a child with OI type I and demonstrated underproduction of alpha 1 (I) collagen chains and alpha 1 (I) mRNA. Indirect RNase protection suggested two species of alpha 1 (I) mRNA, one of which was not collinear with fully spliced alpha 1 (I) mRNA. The noncollinear population was confined to the nuclear compartment of the cell, and contained the entire sequence of intron 26 and a G-->A transition in the first position of the intron donor site. The G-->A transition was also identified in the genomic DNA. The retained intron contained an in-frame stop codon and introduced an out-of-frame insertion within the collagen mRNA producing stop codons downstream of the insertion. These changes probably account for the failure of the mutant RNA to appear in the cytoplasm. Unlike other splice site mutations within collagen mRNA that resulted in exon skipping and a truncated but inframe RNA transcript, this mutation did not result in production of a defective collagen pro alpha 1 (I) chain. Instead, the mild nature of the disease in this case reflects failure to process the defective mRNA and thus the absence of a protein product from the mutant allele.

Defective Splicing of mRNA from One COL1A1 Allele of Type I Collagen in Nondeforming (Type I) Osteogenesis Imperfecta

M. L. Stover, D. Primorac, S. C. Liu,* M. B. McKinstry, and D. W. Rowe
*From the Departments of Pediatrics and *Biostructure and Function, University of Connecticut Health Center, Farmington, Connecticut 06030*

Abstract

Osteogenesis imperfecta (OI) type I is the mildest form of heritable bone fragility resulting from mutations within the COL1A1 gene. We studied fibroblasts established from a child with OI type I and demonstrated underproduction of α1(I) collagen chains and α1(I) mRNA. Indirect RNase protection suggested two species of α1(I) mRNA, one of which was not collinear with fully spliced α1(I) mRNA. The noncollinear population was confined to the nuclear compartment of the cell, and contained the entire sequence of intron 26 and a G → A transition in the first position of the intron donor site. The G → A transition was also identified in the genomic DNA. The retained intron contained an in-frame stop codon and introduced an out-of-frame insertion within the collagen mRNA producing stop codons downstream of the insertion. These changes probably account for the failure of the mutant RNA to appear in the cytoplasm. Unlike other splice site mutations within collagen mRNA that resulted in exon skipping and a truncated but in-frame RNA transcript, this mutation did not result in production of a defective collagen proα1(I) chain. Instead, the mild nature of the disease in this case reflects failure to process the defective mRNA and thus the absence of a protein product from the mutant allele. (*J. Clin. Invest.* 1993. 92:1994–2002.) Key words: osteogenesis imperfecta • RNA splicing • nuclear RNA transport • RNase protection • premature stop codon

Introduction

Osteogenesis imperfecta (OI)[1] is a heritable disorder that results in bone fragility. Identification of mutations within the COL1A1 and COL1A2 genes that encode the chains of type I collagen has begun to provide a detailed understanding of the structural requirements for the type I collagen molecule (1, 2). Mutations that disrupt the triple-helical configuration of collagen, including substitutions for glycine in the gly-x-y triplet (3), partial gene deletions (4), and exon skipping (5), produce, depending on the location of the mutation, a range of disease severity that extends from lethal (OI type II) to severely deforming (OI type III) to mildly deforming (OI type IV) (6).

In OI type I fractures are usually limited to the prepubertal years and are not associated with bone deformity. At times the fracture frequency is not recognized as a distinct abnormality and the disorder only comes to medical attention when early-onset osteoporosis develops in adult family members. Thus, mild forms of OI and "familial osteoporosis" often coexist, in part because the underlying pathogenesis may be similar. For mutations in the α1(I) chain (COL1A1), there appears to be a positional effect in which those at the COOH terminus are more likely to result in severe disease than those at the NH_2 terminus (7). In fact, substitutions for glycine have been identified near the NH_2-terminal end of the triple helix that result in nondeforming bone disease (type I OI) and heritable osteoporosis (8). Mutations within the α2(I) chain (COL1A2 gene) that produce mild OI or osteoporosis (9) do not always occur at the NH_2-terminal end of the chain (10).

The most common mechanism for OI type I appears to be decreased synthesis of normal type I collagen molecules. The synthesis of proα1(1) collagen chains (11) and the steady-state level of α1(I) mRNA appears to be reduced by ~ 50% (12, 13), suggesting that one of the two COL1A1 alleles is "null." We previously showed that dermal fibroblasts from some individuals with type I OI contain normal or elevated amounts of α1(I) mRNA in the nucleus (14). Furthermore, a novel species of α1(I) collagen mRNA present in the nuclear compartment of cells from one such child was not collinear with a cDNA probe (15). In this paper we show that a mutation within a splice donor site results in inclusion of the entire succeeding intron in the mature mRNA that accumulates in the nuclear compartment. Apparently because no abnormal proα1(I) chains are synthesized from the mutant allele, the clinical phenotype of this child is mild.

Methods

Description of patient. CF was a 13-yr-old female at the time of the dermal biopsy (cell line 053). She had been followed at the Newington Childrens Hospital (NCH) between 10 and 15 yr of age. She had sustained seven fractures before her first NCH visit, all of which healed without deformity. Subsequently, the fracture rate diminished and her primary problems were ligament strains and knee discomfort secondary to patellar subluxation. Her height was at the 50–75 percentile. She had blue sclera, ligamentous laxity, a mild thoracic scoliosis (< 10%), and normal dentition. Her current age is 28 yr, and she and her family, including an affected brother, have been lost to follow-up.

Cell culture. A fibroblast strain was derived from a skin explant taken from the inner aspect of the upper arm. Fibroblasts were maintained as previously described (12). Confluent cultures were harvested for protein or RNA after 48 h in fresh culture media supplemented daily with 25 μg/ml ascorbic acid.

This work was a poster presentation at the Americal Society of Bone and Mineral Research, Minneapolis, MN, September 30–October 4, 1992.

Address correspondence to Dr. David W. Rowe, Department of Pediatrics, School of Medicine, University of Connecticut Health Center, 263 Farmington Avenue, Farminton, CT 06030-1515.

D. Primorac's permanent address is the University of Zagreb, School of Medicine at Split, 58000 Split, Croatia.

Received for publication 2 February 1993 and in revised form 30 April 1993.

1. *Abbreviations used in this paper:* nt, nucleotide; OI, osteogenesis imperfecta.

J. Clin. Invest.

0021-9738/93/10/1994/09 $2.00
Volume 92, October 1993, 1994–2002

Collagen analysis. For experiments analyzing collagen synthesis, the cultured cells were radiolabeled with [2,3-^{3}H]-L-proline (DuPont Co., NEN Research Products, Boston, MA) as previously described (12). Collagen synthesis was determined from collagenase susceptible counts in an aliquot of the combined medium and cell extract (16). The relative proportion of collagen α1(I), α2(I), and α1(III) chains was determined by interrupted SDS-PAGE electrophoresis of pepsin-treated protein extracts (17).

Cell fractionation and nucleic acid extraction. Total RNA was extracted from the confluent fibroblasts cell layer using the SDS proteinase K method (18). The nuclear and cytoplasmic compartments of the cells were isolated as previously described (14). Briefly, confluent fibroblast cultures (12–100-mm petri plates) were scraped from the plate in 10 ml of cold PBS, centrifuged, and resuspended in 15 ml of lysis buffer (10 mM Tris, pH 7.5, 10 mM KCl, and 1.5 mM $MgCl_2$) containing 0.25% Triton X-100. The pellet was disrupted by dounce homogenization with a tight-fitting pestle. The suspension was centrifuged for 10 min at 2,200 *g* and the supernatant was transferred to a 30-ml corex tube (cytoplasmic extract) containing 2.0 ml of a 10× extraction buffer (10% SDS, 0.10 M Tris, pH 7.5, 0.05 M EDTA, and 500 μg/ml of proteinase K). The nuclear pellet was reextracted a second time in 5 ml of lysis buffer with Dounce homogenization. After centrifugation, the supernatant was combined with the first extract, producing a final concentration of the extraction buffer of 1×. The nuclear pellet was extracted in 15 ml of 1× extraction buffer. Subsequent steps of RNA isolation followed the SDS proteinase K protocol (18).

DNA was harvested from cultured fibroblasts in 1× extraction buffer containing 100 μg/ml of proteinase K. The digest was extracted sequentially in phenol/chloroform and chloroform and adjusted to 0.3 M NaOAc. Two to three volumes of ethanol was layered over the extract, and the DNA was spooled from the interface, washed in 70% ethanol, dried, and resuspended in water.

RNA analysis. The relative content of α1(I) and α2(I) mRNA in the nuclear and cytoplasmic compartments was assessed by dot hybridization as previously described (14). The filters were hybridized with α1(I)- and α2(I)-specific cDNA probes (HF404 [19] and HF32 [20]) after labeling by the random primer method (21). The relative intensity of hybridization was quantitated in a Betascope G03 (Betagen, Waltham, MA) from the slope of the dilution curve of patient and control RNA.

The method for indirect RNase protection of nuclear and cytoplasmic RNA using single-stranded (SS) antisense cDNA to α1(1) mRNA has been previously described (15). Briefly, hybridization of the sscDNA to RNA was carried out at 50°C in 20 μl of hybridization buffer (0.75 M NaCl, 1 mM EDTA, 50 mM Hepes) (*N*-[2-hydroethyl]-piperazine-*N'*-[2-ethanesulfonic acid]) for 3 h. The sample was diluted to 300 μl in RNase buffer (0.2 M NaCl, 20 mM Hepes, pH 7.5, and 2 mM EDTA) and incubated with 1 μg RNase A and 10 μg RNase T1 for 30 min at 37°C. The RNase-resistant hybrids were extracted in SDS proteinase K, ethanol precipitated, and reapplied to 1% agarose gel made up in 2.2 M formaldehyde. The electrophoresis was carried out at room temperature in a 25 mM imidazole, 25 mM 3-*N*-morpholino]-propanesulfonic acid (MOPS), 0.5 mM Na-EDTA buffer, pH 7.0, for 3 h at 100 V. The RNA fragments resistant to the RNase digestion were transferred to a nitrocellulose filter with a Posiblot apparatus (Stratagene, La Jolla, CA) and identified by hybridization with 2.5×10^6 cpm of [$^{32}PO_4$]UTP-labeled cRNA probe of HF404 transcribed from a sp64 vector (22).

The direct RNase protection was performed with a cRNA of a PCR-derived DNA fragment that had been cloned into pBSII-SK$^+$ vector (Stratagene). Nuclear and cytoplasmic RNA (7 μg) were hybridized with 10^4 cpm of the cRNA overnight at 50°C in 50 μl of buffer containing 0.4 M NaCl, 10 mM Hepes, pH 7.3, and 1 mM EDTA. The next day the solution was diluted to 300 μl of T2 buffer (50 mM NaOAc, pH 4.6, 2 mM EDTA, and 0.1 M NaCl) containing 5 μl of crude T2 RNase as previously described (23). The fragments of the protected probe were identified by electrophoresis in a 6.0% denaturing polyacrylamide gel. An RNA ladder of 0.16–1.77 kb (no. 56235A; GIBCO BRL, Gaithersburg, MD) was end labeled with γ-[$^{32}PO_4$]ATP according to the supplier's directions.

PCR of RNA and genomic DNA. 5.0 μg of nuclear or cytoplasmic RNA was reverse transcribed to cDNA using an α1(I)-specific primer directed to sequences within the COOH-terminal propeptide as previously described (8). The reaction was carried out in 50 mM Tris, pH 8.3, 7 mM KCl, 3 mM $MgCl_2$, 10 mM dithiothreitol, and 2.5 μg of the primer. The mixture was heated at 65°C for 3 min, cooled, adjusted to contain 500 μM dNTPs, and 500 U of Moloney murine leukemia virus (M-MLV) reverse transcriptase. The incubation lasted for 1 h at 37°C and was followed by phenol/chloroform extraction and ethanol precipitation. The sample was dissolved in 20 μl water and 1 μl was used for PCR amplification. In the case of genomic DNA, 1 μg of DNA was used in the PCR reaction. The oligonucleotide pairs for each set of amplifications are given in Table I.

The PCR mixture contained 50 mM KCl, 10 mM Tris, pH 8.3, 1.5 mM $MgCl_2$, 0.01% gelatin, and 2.5 U AmpliTaq (Perkin Elmer Corp., Norwalk, CT) in a final reaction volume of 84 μl. The mixture was heated to 94°C for 5 min, after which the 16 μl containing 200 μM dNTPs (hot start [24]) was added and the sample was cycled 25 times with denaturation at 94°C for 1 min, hybridization at 62°C for 1 min, and polymerization at 72°C for 1 min in a DNA thermal cycler (Perkin Elmer-Corp.). The products were extracted in phenol/chloroform and 1:10 of the original reaction volume was applied to a 5% polyacrylamide TBE mini-gel.

For direct DNA sequencing of PCR products, we used the dsDNA Cycle Sequencing System (GIBCO BRL). PCR products were cloned either with G-tailing into C-tailed pUC9 vector (Pharmacia Fine Chemicals, Piscataway, NJ) or by blunt end cloning into Sma site pBSII-SK$^+$. For sequencing of the cloned DNA, we used the Sequenase Reagent Kit (no. 70750; United States Biochem. Corp., Cleveland, OH).

Results

Total collagen synthesis by the fibroblasts derived from the patient (Table II) was 4.0% of total protein synthesis. For comparison, three cell strains derived from patients with severe forms of the disease are shown. We have observed significant variation in the rate and percentage of collagen synthesis in part dependent on the age of the donor. The range of synthesis in non-OI control cells is 5–7% (5). In most of the cell strains obtained from patients with OI type I, a reduced rate of total collagen synthesis is observed.

The relative proportion of type I and type III collagen was determined on pepsin-treated radiolabeled protein extracts of the medium and cell layer (Fig. 1). The ratio was elevated to 0.38 and 0.45 in two experiments (control patients range from 0.15 to 0.25). Because an equal amount of total collagen is loaded into each electrophoresis lane, the elevation in the collagen type III-to-I ratio reflects underproduction of type I collagen. The electrophoretic mobilities of the α1(I) and α2(I) chains from the patient sample were identical to those of the control sample.

The relative content of α1(I) and α2(I) mRNA was determined by dot blot hybridization and compared with the ratio of control cells (Fig. 2). The slope of the line of dot density when the filter was hybridized to an α1(I) or α2(I) cDNA probe was compared with the slope of a control RNA applied to the same filter. In this assay, the α1(I)/α2(I) mRNA ratio of the control is normalized to 1. When the slope of the α1(I) and α2(I) mRNA derived from cell strain 053 was compared with the standard (std), the α1(I)/std was 0.30 and the α2(I)/std was 1.19, making the α1(I)/α2(I) ratio of the patient 0.25. Thus, the amount of α1(I) mRNA present in the total cellular RNA was reduced to less than half of the control. Because the control

Table I. Oligonucleotide Pairs Used to Amplify α1(I) mRNA Encoded by HF404

Oligonucleotide pair/exons amplified	NT* 5′ Oligonucleotide-3′	3′-Oligonucleotide 5′ NT*
1:35–39	2541	AGCAGGACCATCAGCACCAGGGGAT
	CCCGGCCCTGCTGGCTTTGCTGGCC	2948
2:30–36	2113	TTTCGGCCACGACTAACCAGG
	AGACCTTGGCGCCCCTGGCCCCTCT	2621
3:25–31	1800	TTGCACACCACGCTCGCCAGGGAAA
	TCAAGATGGTCGCCCCGGA	2159
4:24–28	1746	AGCAGGGCCTTGTTCACCTCTCTCG
	AGCCCTGGCAGCCCTGGTCCTGATG	2005
5:19–23	1330	GCTTCACCGGGACGACCAGTC
	GTATTGCTGGTGCTCCTGG	1708

* The nucleotide number is determined from the α1(I) collagen mRNA start site.

cells have a 2:1 ratio of α1(I)/α2(I), the amount of α1(I) of the patient sample is reduced to < 1:1.

Analysis of this patient and other cell lines with OI type I indicated that the α1(I)/α2(I) collagen mRNA ratio extracted from isolated nuclei approximated or exceeded 2:1 in contrast to the low ratio from cytoplasmic compartment (14). This observation suggested a defect in α1(I) mRNA transport similar to that observed in splicing defects associated with β-thalassemia (25, 26). An indirect RNase protection was developed to detect insertions within collagen mRNA (15). It relies on protection of the mature mRNA with a single-stranded antisense cDNA. The protected or cleaved RNA is detected by Northern hybridization of the RNase-digested products. Nuclear and cytoplasmic RNA was protected with a 1.8-kb cDNA fragment derived from HF404 that encodes exons 19–43 (Fig. 3). In RNA from cell strain 053, the expected 1.8-nucleotide (nt) band and two smaller bands (1.4 and 0.4 nt) were visible in nuclear RNA. Only the fully protected band was present in the cytoplasmic RNA and was similar in size to RNA extracted from normal cells.

To characterize the abnormal RNA species, nuclear RNA was reverse transcribed and then PCR amplified in segments to span the region encoded by HF404 (Fig. 4). Of the five PCR primer sets used (Table I), two (sets 3 and 4) produced a second higher molecular weight (~ 150 bp larger) band from the patient but not the control RNA. A third band well above the 1.3-kb marker was also unique to these oligonucleotide sets and probably represents a heterodimer produced when two complementary strands of different lengths hybridized during the late rounds of amplification. Both primer sets amplify a cDNA segment that has intron 26 in common.

The PCR products of the normal and mutant RNA were cloned. DNA sequencing revealed the presence of the intron 26 (Fig. 5 *A*) in the larger species. Furthermore, the expected first position donor site G was replaced by an A. This mutation was confirmed when the corresponding region of the genomic DNA was isolated by PCR and found to be present in ~ 50% of the clones sequenced (Fig. 5 *B*).

The cloned reverse transcriptase PCR fragments encompassing exons 25–28 derived from oligonucleotide set 3 were used to generate a radiolabeled antisense cRNA probes complementary to the retained intron and the normally spliced transcript. These probes were hybridized to nuclear and cytoplasmic RNA derived from the patient and control cell strain (Fig.

Table II. Production of Total Collagen by Dermal Fibroblasts Derived from Patients with Various Forms of OI

Cell line	Collagen synthesis	Total protein synthesis	Percent collagen
	dpm/mg DNA		
053 (This case)	372	1,729	3.98
118 (Lethal OI)	1,385	4,042	6.34
144 (Type III OI, age 5)	818	2,264	6.69
147 (Type III OI, age 46)	693	2,292	5.59

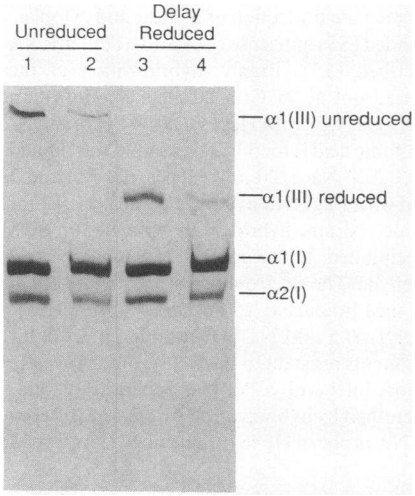

Figure 1. SDS-PAGE of radiolabeled collagen from patient and control fibroblast cultures. The cells were labeled for 24 h and the cell layer and media combined for analysis. The sample was treated with pepsin to destroy the noncollagen proteins before the gel electrophoresis. The type III collagen is separated from type I based on the unique intrahelical cysteine disulfide bonds within type III collagen. The relative ratio of the α1(III)/α1(I) was obtained by quantitative densitometry of the pattern in lanes *3* and *4*. Lanes *1* and *3*, patient; lanes *2* and *4*, control.

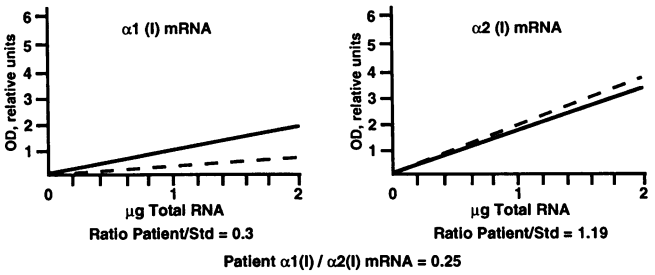

Figure 2. Plot of dot hybridization density and total RNA obtained from patient- and control-cultured fibroblasts. The slope for the $\alpha 1(I)$ mRNA of the patient is less than the control while that of $\alpha 2(I)$ is similar to control. This finding is consistent with a reduction of the $\alpha 1(I)$ mRNA that accumulates within the fibroblasts of the patient.

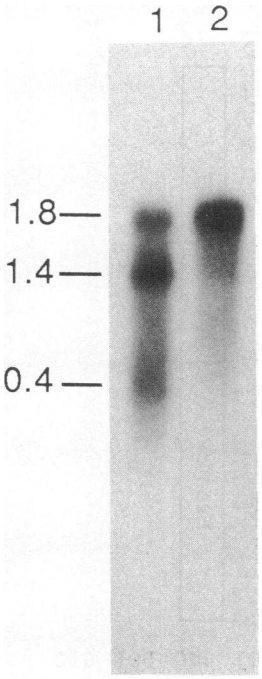

Figure 3. Indirect RNase protection of nuclear and cytoplasmic $\alpha 1(I)$ mRNA extracted from patient's cultured fibroblasts. The fully protected (1.8-nt) band present in both compartments is identical to the band found in normal fibroblasts. However, the lower bands (1.4 and 0.4 nt) that were present in the nuclear RNA (lane *1*) but absent from the cytoplasmic RNA (lane *2*) were unique to this patient.

6). The probe containing the included intron protected the predicted 442-nt fragment in the nuclear (Fig. 6 *A*, lanes *4* and *6*) but not the cytoplasmic compartment (Fig. 6 *A*, lanes *3* and *5*) of the patient RNA. No evidence of the 442-nt band was present in the control RNA (Fig. 6 *A*, lanes *7* and *8*). The RNA encoded by the normal COL1A1 allele from the patient as well as all the RNA from the control cell line did not protect the intron 26–containing probe. Instead the probe was cleaved producing two fragments (209 and 90 nt) collinear with the flanking exons. The cRNA probe derived from the normally spliced allele was protected equally well by total fibroblast RNA from the patient (Fig. 6 *B*, lane *6*) and control (Fig. 6 *B*, lane *5*), indicating that exon skipping was not occurring to a significant degree as a result of the donor site mutation (27). The intensity of the normally spliced transcript in the patient total RNA extract is significantly less than the control, confirming the low amount of normal $\alpha 1(I)$ mRNA that is present in this cell strain.

Discussion

The fibroblasts derived from this patient had a G → A transition at position + 1 of intron 26 in one of the two COL1A1 alleles so that the donor splice site was inactivated. As a consequence of this mutation, all of intron 26 was retained and the frame-shifted mRNA transcript contained several stop codons within and downstream of the intron insertion. Analysis of RNA transcripts from other multiexon genes such as dihydrofolate reductase (28) and triose phosphate isomerase (29) have shown that transport of the abnormal transcript into the cytoplasm was reduced when the stop codon was located in all but the terminal exon. The behavior of the mutant $\alpha 1(I)$ mRNA in the 053 cells is consistent with the model that a stop codon before the penultimate exon prohibits transport from the nucleus, since there was no evidence of abnormal transcript accumulation in the cytoplasm.

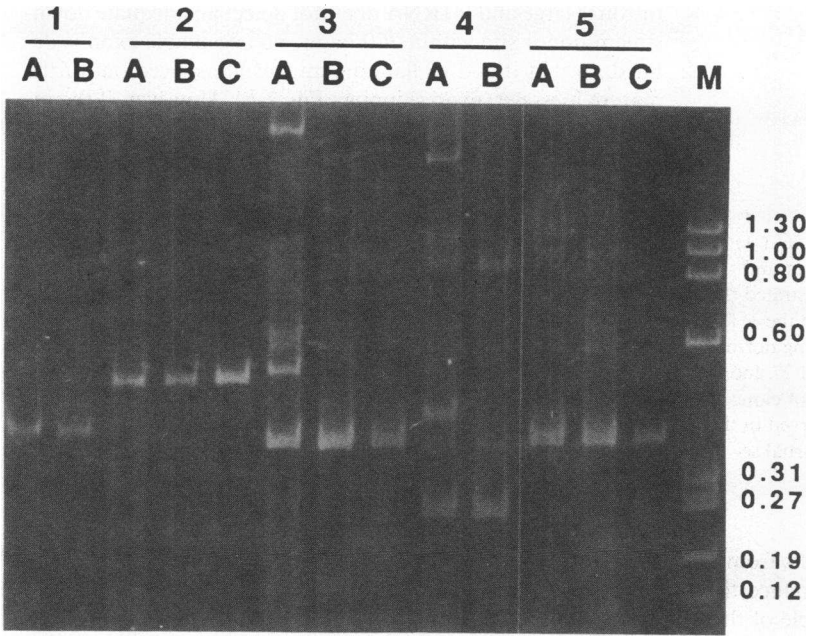

Figure 4. PCR of reverse-transcribed cDNA using nuclear RNA extracted from control- and patient-cultured fibroblasts. (Lanes *A*) Patient RNA; (lanes *B*) control RNA; (lanes *C*) cloned $\alpha 1(I)$ cDNA. The new band is present in lanes 3 *A* and 4 *A*, and is ~ 150 bp larger that the expected band. The heterodimer band is also present in these lanes. The primer pairs are given: (lanes *1*) exons 35–39; (lanes *2*) exons 30–36; (lanes *3*) exons 25–31; (lanes *4*) exons 24–28; (lanes *5*) exons 19–23. (Lane *M*) PhiX HaeIII DNA markers.

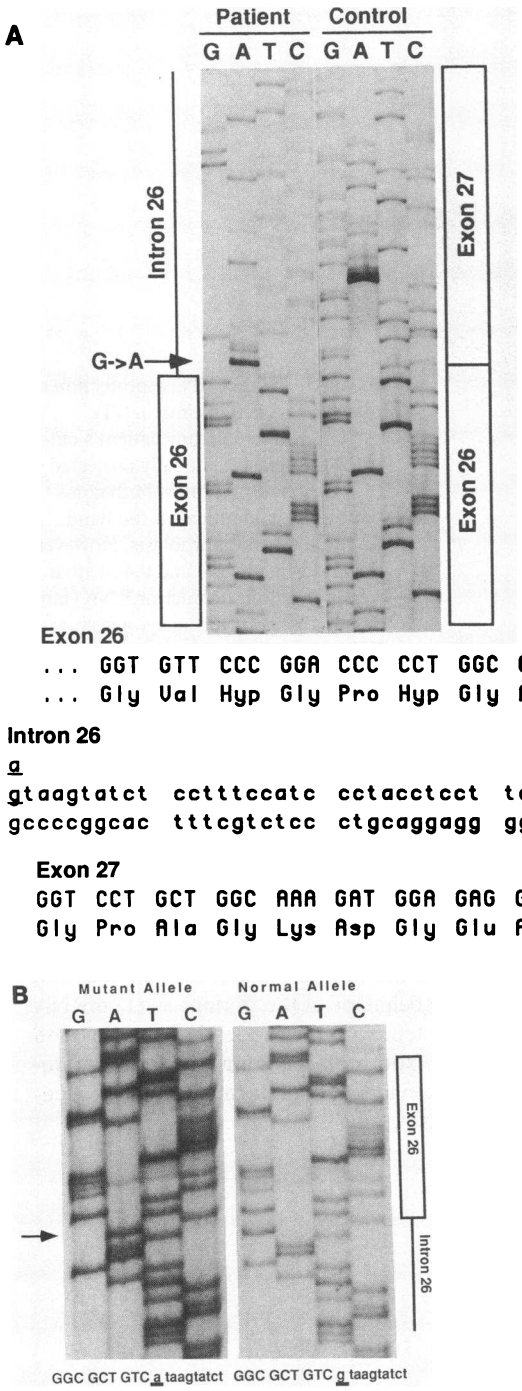

Figure 5. DNA sequencing of the mutation within intron 26. (*A*) DNA sequencing of cloned PCR fragments from the patient and control nuclear RNA. The abnormal patient allele demonstrated the presence of a complete intron 26 with a G → A transition in the + 1 donor position. The patient also had sequence demonstrating normal splicing of intron 26. The partial sequence of exons 26 and 27 and intron 26 is given below the figure. (*B*) DNA sequencing of cloned alleles of patient 053. The G → A transition that was observed in the nuclear RNA was confirmed in the genomic DNA. The partial sequence of exon and intron 26 is given below the figure.

A frame shift or premature stop codon that results from a mutation within an exon would be expected to produce the same effect as a retained intron, although an example of this type of mutation has not been reported. All of the defined mutations of COL1A1 or COL1A2 that induced a frame shift did not result in a premature stop codon before the terminal exon (30–33). Low levels of α1(I) mRNA (based on reversed ratio of α1/α2 mRNA ratio of 1:2) were observed with a frame shift in exon 51 that inactivated the authentic translation termination codon and extended the translated product to the first polyadenylation signal (34). A potential mechanism for the low α1(I) mRNA seen in this case is an indirect effect on splicing secondary to the altered transcriptional termination signal (35).

In fibrillar collagen genes all the exons that encode the triple-helical portion of the α chain are in-frame cassettes of the Gly X-Y triplet starting with the first G of the glycine codon and ending with the last nucleotide of the Y position amino acid. The collagen mutations that cause abnormalities of splicing were initially discovered because of synthesis of shortened α chains. The shortened chain resulted from exon skipping due to a mutation in either an intron donor or acceptor site (Table III). These mutations can result in more severe forms of OI because the shortened RNA transcripts are in frame, are transported to the cytoplasm, and are translated into a functional although truncated α chain. The shortened α chain is incorporated into a trimeric molecule that interferes with secretion, proteolytic processing, and fibrillogenesis, thus accounting for the strong dominant nature of mutations of structural macromolecules (36).

An explanation for why a similar mutation in a donor site results in exon skipping in the cases cited above and intron retention in our case comes from the model of exon definition developed by Robberson et al. (37) and Talerico and Berget (38). In this model, the first step is the binding of U2 RNA to a potential intron acceptor site followed by the second step, in which U1 RNA and specific proteins identify and bind to the strongest adjacent downstream donor site (39). This process defines the exons and initiates the splicing machinery to excise the intervening intron sequence. If U1RNA identifies an adequate donor within 300 bases (the upper limit in size for internal exons [40]), then an exon is defined (Fig. 7 *a*) and the intervening introns are spliced. With a mutant donor site, if the intron is large and U1RNA does not detect an adequate downstream donor site within 300 bases, no acceptable exon is defined, so that it and its flanking introns are spliced out of the mature message (exon skipping; Fig. 7 *b*). However, if the intron is small, then U1RNA will identify the authentic donor site of the subsequent downstream intron. This defines the two exons and the included intron as an acceptable "exon." Thus, the entire intron is retained in the mature RNA (intron retention; Fig. 7 *c*). This model accounts for why a mutant donor site can result in either exon skipping or intron retention and also predicts that all intron acceptor site mutations will result in exon skipping, because the defining process to identify an exon is initiated from the acceptor site.

Approximately half of the introns within the α1(I) gene are < 150 bp, making the risk for a donor site mutation leading to a retained intron a common possibility. There are 25 exon/intron/exon combinations that are < 300 bp, the size limit for an acceptable exon if the donor site of the included intron was inactivated. 17 of the candidate sites would induce a frame shift mutation, while 5 of the in-frame introns contain an in-frame stop codon. The phenotype of a patient harboring a mutation that introduces a frame shift–producing or stop codon–con-

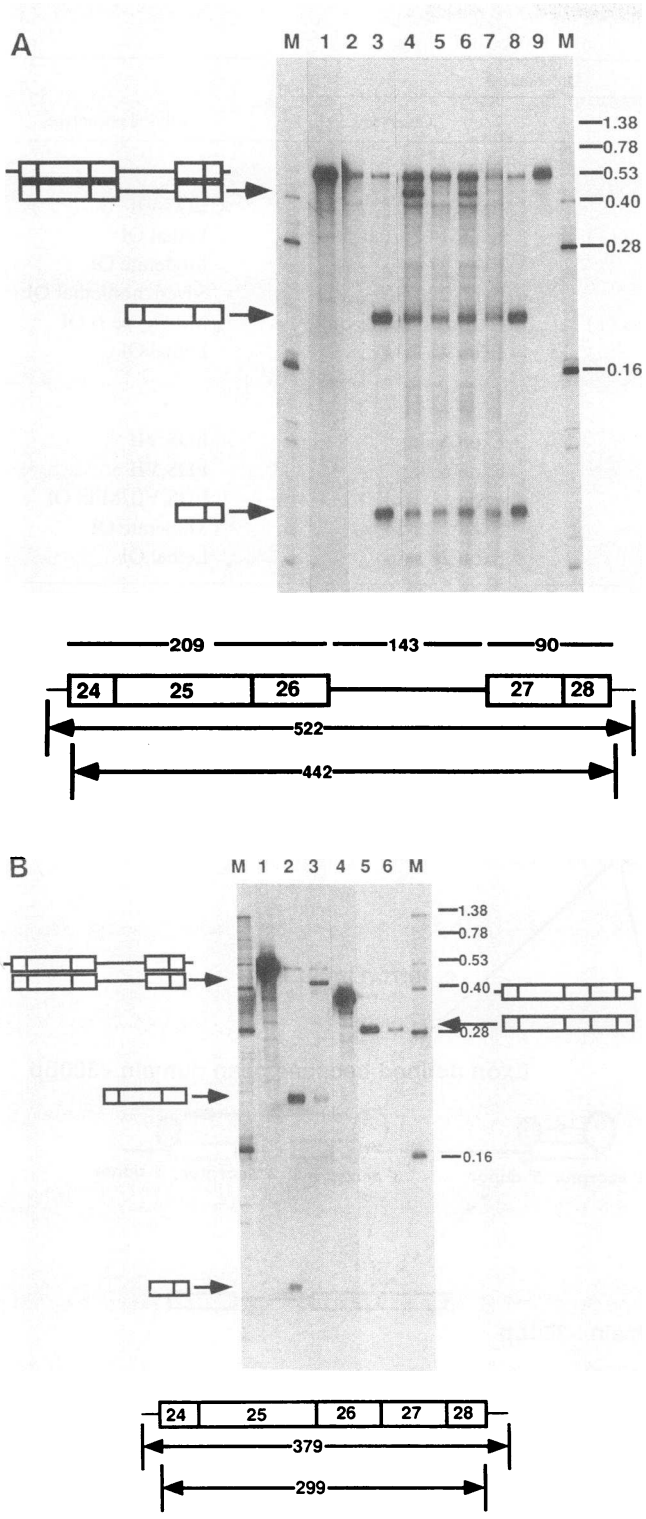

Figure 6. Direct RNase protection of RNA from patient and control fibroblast strains. (*A*) RNA was hybridized with the intron 26–containing cRNA probe derived from cloning and sequencing experiment discussed in Fig. 5 *b*. Two separate preparations of nuclear and cytoplasmic RNA from the patient's fibroblast (lanes *3–6*) were analyzed. The probe is 522 nt (lanes *1* and *9*) and is incompletely digested by the RNase treatment (lane *2*). It protects a 442-nt fragment of RNA from the nuclear compartment from the patient fibroblasts (lanes *4* and *6*), which is not found in the cytoplasmic compartment from the same cell preparations (lanes *3* and *5*). RNA from a control cell strain did not show this band (lane *7*, cytoplasmic; lane *8*, nuclear). The smaller size results from RNase cleavage of the polylinker portion of the cloning vector. The probe is not fully protected by normally spliced RNA and is cleaved into two exon-containing fragments containing exons 24–26 (209 nt) and exons 27–28 (90 nt). The two fragments were present in all lanes containing α1(I) collagen mRNA. Kinased RNA markers are in the lanes *M*. The probe and the expected band sizes are diagrammed below the figure. (*B*) RNase protection with the intron containing (lanes *1–3*) and fully spliced (lanes *4–6*) cRNA probes. Lane *1*, intron containing probe alone; lane *2*, patient cytoplasmic RNA; lane *3*, patient nuclear RNA; lane *4*, spliced probe alone; lane *5*, total RNA from control cells; lane *6* total RNA from patient cells. Kinased RNA markers are in lanes *M*. A diagram of the fully spliced cRNA is given beneath the figure.

taining retained intron would be expected to be similar regardless of the site of the intron. However, variability might be observed if the exon/intron/exon size approaches or exceeds 300 bases. At that point exon definition would not be determined invariably and the entire domain would be either retained or skipped. Disease severity would reflect the proportion of mRNA that contained the skipped exon.

Review of published data of collagen splicing mutations is consistent with the exon definition model, with one possible exception (Table III). All acceptor mutations resulted in exon skipping (49, 50), as did donor mutations where the exon/intron/exon size (scanning domain) was > 300 nt (5, 45–48). In a number of these cases (49) the mRNA product of the mutant allele was lower than predicted by the model. It seems likely that cryptic splice sites may have been used and produced some transcripts containing frame shift mutations that were not transported to the cytoplasm. Since nuclear RNA was not examined in these cases, this possibility needs further examination.

The exon definition model may require modification to accommodate an exon 14 donor site mutation in COL1A1 (50). Here intron retention is predicted that would insert three in-frame stop codons and produce an OI type I phenotype. Instead, exon 14 was skipped and resulted in a lethal phenotype (Table III). A potential explanation for this exception is the nature of the mutation. This donor site mutation was located at position + 5, which is a less highly conserved consensus base and produced either normal splicing or exon skipping. Normal splicing was decreased at elevated temperatures. This suggests that the exon-defining mechanism can be disrupted if there is weak binding of U1RNA at the mutant donor site. In those transcripts where U1 binds the mutant site but does not result in cleavage, there is interference with splicing of the upstream intron such that the upstream intron forms a lariat with the downstream branch point resulting in exon skipping. Cooperation between branch points and donor sites in determining the extent of exon skipping has been observed in a different context (39, 51). Other factors such as strength of competing donor and acceptor splice sites and exon size also influence which splicing alternative will predominate, and may account for the heterogeneity of disease severity observed in mild to moderately deforming OI that result from mutations that affect splicing of collagen mRNA. The exception observed in this case suggests that an included intron will only be observed when the donor site is completely inactivated and the scanning mechanism progresses to the next downstream strong donor site.

Analysis of patients with a null collagen allele has been difficult because protein data pointing to the site of a potential

Table III. Outcome of Reported Splicing Mutations

Intron No.	Intron mutation	Scan domain	Outcome		Clinical phenotype
			Predicted	Observed	
			α1(I) mRNA		
6 (45)	Donor (−1)	345	Exon 6 skip	Exon 6 skip	EDS VII
14 (50)	Donor (+5)	213	Intron 14 inclusion (+)	Exon 14 partial skip*	Lethal OI
16 (41)	Acceptor		Exon 17 skip	Exon 17 skip	Moderate OI
21 (42)	Acceptor		Exon 22 skip	Exon 22 skip	Severe nonlethal OI
26 (This case)	Donor (+1)	251	Intron 26 inclusion (+)	Intron 26 inclusion (+)	Mild (type I) OI
47 (46)	Donor (+1)	482	Exon 47 skip	Exon 47 skip	Lethal OI
			α2(I) mRNA		
6 (47)	Donor (+1)	749	Exon 6 skip	Exon 6 skip	EDS VII
6 (48)	Donor (−1)	749	Exon 6 skip	Exon 6 skip	EDS VII
10 (43)	Acceptor		Exon 11 skip	Exon 11 skip	EDS VII/Mild OI
12 (5)	Donor (+2)	1399	Exon 12 skip	Exon 12 skip	Moderate OI
27 (44)	Acceptor		Exon 28 skip	Exon 28 skip	Lethal OI

The intron containing an acceptor or donor site mutation is designated in columns 1 and 2. The scanning domain (exon/intron/exon size) is only given for donor site mutations because it is this type of mutation in which the scanning domain appears to be important. The predicted and actual outcomes of the acceptor or donor mutation are indicated. Whether a retained intron would contain a stop codon or induce one by a frame shift within the mature RNA is give as + or −. The disease severity observed as a consequence of the mutation is given. EDS VII is Ehlers-Danlos syndrome, type VII. * This exception is discussed in the text.

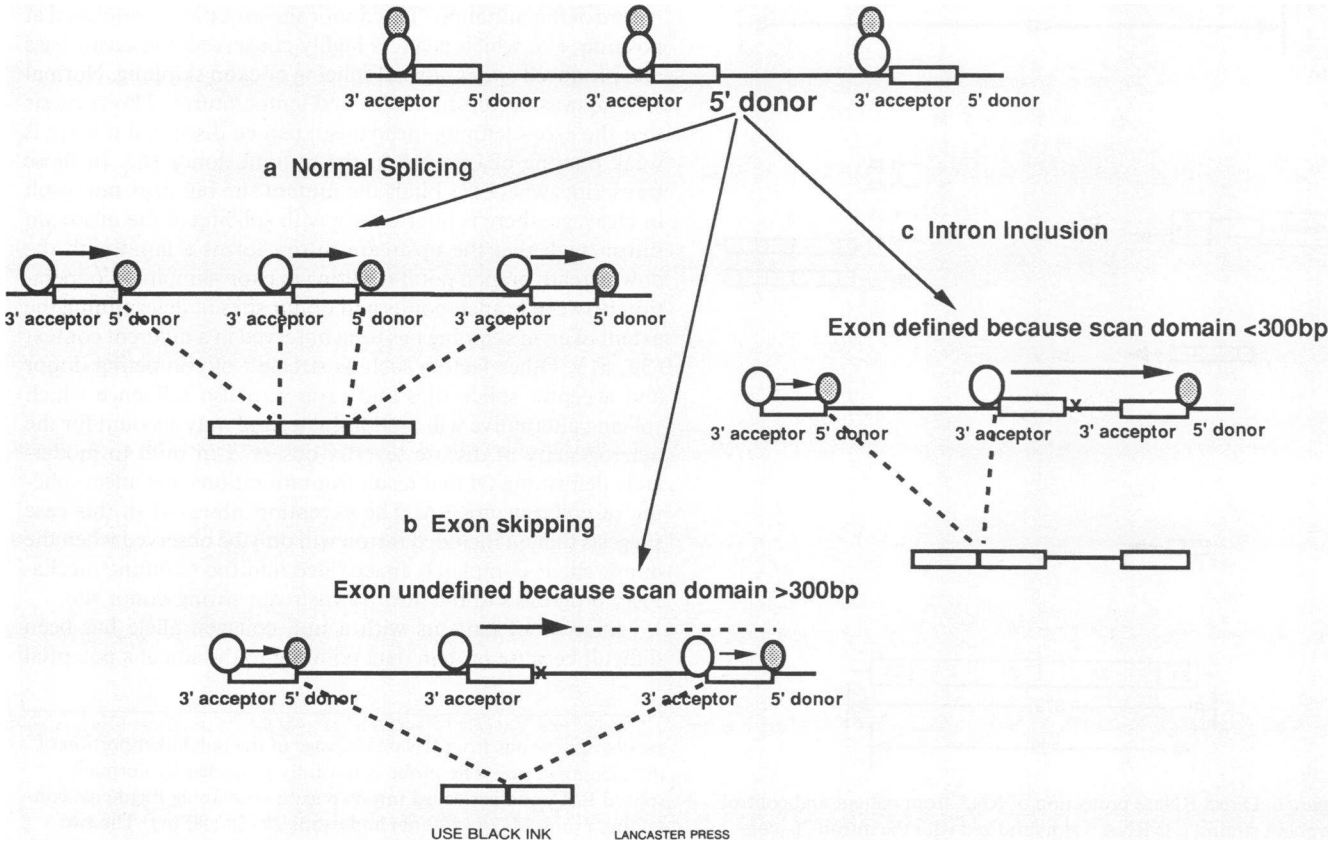

Figure 7. Outcome of splicing with a mutation that inactivates a donor site. U2 RNA (*open circle*) identifies the 3′ acceptor site and subsequently U1 RNA (*filled circle*), and associated proteins identify the downstream donor site. (*a*) Normal splicing occurs when exons are identified and the intervening introns are removed. (*b*) Exon skipping occurs with a damaged donor site. If the scanning mechanism does not identify an adequate donor within 300 bp of the acceptor site, the region is not considered an exon and the region is spliced. The outcome is that the exon adjacent to the mutant donor site is skipped. (*c*) Intron inclusion also can occur with a damaged donor site. If the scanning mechanism does identify a strong donor site within 300 bp of the acceptor site, then the exon is defined by the position of the donor site. If the new donor site is the next downstream exon, then the entire intron is included as part of the enlarged exon. If an adequate donor is identified within the intron, then cryptic splicing and inclusion of a partial intron are observed.

mutation are lacking. Initial characterization of nuclear RNA not only suggests potential mechanisms to account for the underproduction of mature mRNA from the mutant allele, it provides localization of an unprocessed intron that might not be obtained from analysis of total RNA because of the enrichment of the abnormal transcript in the nuclear compartment. The difficulty of detecting null-producing mutations in cDNA derived from total RNA is illustrated by the presence of 2 abnormal of 19 sequenced cDNA transcripts of the 2.5-kb ferrochelastase mRNA (52), < 2% of the expected cystic fibrosis transmembrane conductance regulator transcript in bronchial epithelial cells detected by mutation specific oligonucleotide hybridization (53), and < 5% of a frame shift–producing exon skip mutation of the 5.3-kb low density lipoprotein receptor mRNA (54). It is likely that frame shift mutations that result from insertions or deletions within exons of COL1A1 will also account for type I OI but these will be hard to localize by the RNase protection techniques used in this case. PCR of fragments of nuclear mRNA enriched for the presence of the mutant transcript followed by an analysis using techniques capable of detecting a 1-bp change will be required to identify such cases. Discovery of mutations within the regulatory domains of the COL1A1 gene that lead to reduced transcriptional activity will be an even greater challenge because the regions essential for the activity of this gene in collagen-producing tissues are not yet clearly defined (55). These will be important abnormalities to identify because it is likely that subtle mutations that result in a partially null allele of the COL1A1 gene may be a common cause of type I OI and heritable forms of osteoporosis.

Acknowledgements

This work was supported by National Institutes of Health grant AR-30426. This work was done in part as fulfillment of D. Primorac's Ph.D. requirement at the University of Zagreb.

References

1. Rowe, D. W. 1991. Osteogenesis imperfecta. *In* Bone and Mineral Research. Vol 7. Ed. Johan, N. M. Heersche, and John A. Kanis, editors. Elsevier Science Publishers B. V., Amsterdam. 209–241.

2. Byers, P. H., and R. D. Steiner. 1992. Osteogenesis imperfecta. *Annu. Rev. Med.* 43:269–282.

3. Steinmann, B., V. H. Rao, A. Vogel, P. Bruckner, R. Gitzelmann, and P. H. Byers. 1984. Cysteine in the triple-helical domain of one allelic product of the α1(I) gene of type I collagen produces a lethal form of osteogenesis imperfecta. *J. Biol. Chem.* 259:11129–11138.

4. Chu, M. L., C. J. Willams, G. Pepe, J. L. Hirsch, D. J. Prockop, and F. Ramirez. 1984. Internal deletion in a collagen gene in a perinatal lethal form of osteogenesis imperfecta. *Nature (Lond.).* 304:78–80.

5. Chipman, S. D., J. R. Shapiro, M. B. McKinstry, M. L. Stover, P. Branson, and D. W. Rowe. 1992. Expression of mutant type I collagen in osteoblast and fibroblast cultures from a proband with osteogenesis imperfecta type IV *J. Bone Miner. Res.* 7:793–805.

6. Kuivaniemi, H., G. Tromp, and D. J. Prockop. 1991. Mutations in collagen genes: Causes of rare and some common diseases in humans. *FASEB (Fed. Am. Soc. Exp. Biol.) J.* 5:2052–2060.

7. Bateman, J. F., I. Moeller, M. Hannagan, D. Chan, and W. G. Cole. 1992. Characterization of three osteogenesis imperfecta collagen alpha 1(I) glycine to serine mutations demonstrating a position-dependent gradient of phenotypic severity. *Biochem. J.* 288:131–135.

8. Shapiro, J., M. L. Stover, V. Burn, M. McKinstry, A. Burshell, S. Chipman, and D. Rowe. 1992. An osteopenic nonfracture syndrome with features of mild osteogenesis imperfecta associated with the substitution of a cysteine for glycine at triple helix position 43 in the proα1(I) chain of type I collagen. *J. Clin. Invest.* 89:567–573.

9. Spotila, L. D., C. D. Constantinou, L. Sereda, A. Ganguly, B. L. Riggs, and D. J. Prockop. 1991. Mutation in a gene for type I procollagen (COL1A2) in a woman with postmenopausal osteoporosis: evidence for phenotypic and genotypic overlap with mild osteogenesis imperfecta. *Proc. Natl. Acad. Sci. USA.* 88:5423–5427.

10. Wenstrup, R. J., A. W. Shrago-Howe, L. W. Lever, C. L. Phillips, P. H. Byers, and D. H. Cohn. 1991. The effects of different cysteine for glycine substitutions within alpha 2(I) chains. Evidence of distinct structural domains within the type I collagen triple helix. *J. Biol. Chem.* 266:2590–2594.

11. Barsh, G. S., K. E. David, and P. H. Byers. 1982. Type I osteogenesis imperfecta: a nonfunctional allele for proα1(I) chains of type I procollagen. *Proc. Natl. Acad. Sci. USA.* 79:3838–3842.

12. Rowe, D. W., J. R. Shapiro, M. Poirier, and S. Schlesinger. 1985. Diminished type I collagen synthesis and reduced α1(I) collagen mRNA in cultured fibroblasts from patients with dominantly inherited (type I) osteogenesis imperfecta. *J. Clin. Invest.* 71:689–697.

13. Willing, M. C., C. J. Pruchno, M. Atkinson, and P. H. Byers. 1992. Osteogenesis imperfecta type I is commonly due to a COL1A1 null allele of type I collagen. *Am. J. Hum. Genet.* 51:508–515.

14. Genovese, C., and D. W. Rowe. 1987. Analysis of cytoplasmic and nuclear mRNA in fibroblasts from patients with type I osteogenesis imperfecta. *Methods Enzymol.* 145:223–235.

15. Genovese, C., A. Brufsky, J. R. Shapiro, and D. W. Rowe. 1989. Detection of mutations in human type I collagen RNA by indirect RNase protection. *J. Biol. Chem.* 264:9632–9637.

16. Peterkofsky, B., and B. Dieglemann. 1971. Use of a mixture of proteinase-free collagneases for the specific assay of radioactive collagen in the presence of other proteins. *Biochemistry.* 6:988–993.

17. Sykes, B., B. Puddle, M. Francis, and R. Smith. 1976. The estimation of two collagens from human dermis interrupted gel electrophoresis. *Biochem. Biophys. Res. Commun.* 72:1472–1480.

18. Genovese, C., D. W. Rowe, and B. Kream. 1984. Construction of DNA sequences complementary to rat α_1 and α_2 collagen mRNA and their use in studying the regulation of type I collagen synthesis by 1,25 dihydroxy vitamin D. *Biochemistry.* 23:6210–6216.

19. Chu, M-L., J. C. Myers, M. P. Bernard, J. F. Ding, and F. Ramierz. 1982. Cloning and characterization of five overlapping cDNAs specific for the human proα1(I) collagen chain. *Nucleic Acids Res.* 10:5925–5933.

20. Myers, J. C., M. L. Chu, S. H. Faro, W. J. Clark, D. J. Prockop, and F. Ramirez. 1981. Cloning a cDNA for the pro-α2 chain of human type I collagen. *Proc. Natl. Acad. Sci. USA.* 78:3516–3520.

21. Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 13:6–13.

22. Myers, R. M., Z. Larin, and T. Maniatis. 1985. Detection of single base substitutions by ribonuclease cleavage of mismatches in RNA:DNA duplexes. *Science (Wash. DC).* 230:1242–1246.

23. Lichtler, A., N. L. Barrett, and G. G. Carmichael. 1992. Simple, inexpensive preparation of T1/T2 ribonuclease suitable for use in RNase protection experiments. *Biotechniques.* 12:231–232.

24. Chow, Q., M. Russell, D. E. Buck, J. Raymond, and W. Block. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy number amplification. *Nucleic Acids Res.* 20:1717–1723.

25. Maquat, L. E., and A. J. Kinniburgh. 1985. A β^0 thalassemic β-globin RNA that is labile in bone marrow cells is relatively stable in HeLa cells. *Nucleic Acids Res.* 13:285–2867.

26. Baserga, S. J., and E. J. Benz, Jr. 1988. Nonsense mutations in the human β-globin gene affect mRNA metabolism. *Proc. Natl. Acad. Sci. USA.* 85:2056–2060.

27. Kuivaniemi, H., S. Kontusaari, G. Tromp, M. Zhao, C. Sabol, and D. J. Prockop. 1990. Identical G + 1 to A mutations in three different introns of the type III procollagen gene (COL3A1) produce different patterns of RNA splicing in three variants of Ehlers-Danlos syndrome IV. An explanation for exon skipping with some mutations and not others. *J. Biol. Chem.* 265:12067–12074.

28. Urlaub, G., P. J. Mitchell, C. J. Ciudad, and L. A. Chasin. 1989. Nonsense mutations in the dihydrofolate reductase gene affect RNA processing. *Mol. Cell. Biol.* 9:2868–2880.

29. Cheng, J., M. Fogel-Petrovic, and L. E. Maquat. 1990. Translation to near the distal end of the penultimate exon is required for normal levels of spliced triosephosphate isomerase mRNA. *Mol. Cell. Biol.* 10:5215–5225.

30. Bateman, J. F., S. R. Lamande, H. M. Dahl, D. Chan, T. Mascara, and W. G. Cole. 1989. A frameshift mutation results in a truncated nonfunctional carboxyl-terminal proalpha 1(I) propeptide of type I collagen in osteogenesis imperfecta. *J. Biol. Chem.* 264:10960–10964.

31. Pihlajaniem, T., L. A. Dickson, F. M. Pope, V. V. R. Korhonen, A. Nicholls, D. J. Prockop, and J. C. Myers. 1984. Osteogenesis Imperfecta: cloning of a pro-α2(I) collagen gene with a frameshift mutation. *J. Biol. Chem.* 259:12941–12944.

32. Chu, M-L., D. Rowe, A. C. Nicholls, F. M. Pope, and D. J. Prockop. 1984. Presence of translatable mRNA for pro α2(I) chains in fibroblasts from a patient

with osteogenesis imperfecta whose type I collagen does not contain α2(I) chains. *Collagen Relat. Res.* 4:389–394.

33. Chipman, S. D., H. O. Sweet, D. J. McBride, M. Davisson, S. C. Marks, A. R. Shuldiner, R. J. Wenstrup, D. W. Rowe, and J. R. Shapiro. 1993. Defective proα2(I) collagen synthesis in a recessive mutation in mice: a potential model of human osteogenesis imperfecta. *Proc. Natl. Acad. Sci. USA.* 90:1701–1705.

34. Willing, M. C., D. H. Cohn, and P. H. Byers. 1990. Frameshift mutation near the 3′ end of the COL1A1 gene of type I collagen predicts an elongated proalpha 1(I) chain and results in osteogenesis imperfecta type I. *J. Clin. Invest.* 85:282–290.

35. Niwa, M., and S. M. Berget. 1991. Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes & Dev* 5:2086–2095.

36. Kadler, K. E., A. Torre-Blanco, E. Adachi, B. E. Vogel, Y. Hojima, and D. J. Prockop. 1991. A type I collagen with substitution of a cysteine for glycine-748 in the α1(I) chain copolymerizes with normal type I collagen and can generate fractallike structures. *Biochemistry.* 30:5081–5088.

37. Robberson, B. L., G. J. Cote, and S. M. Berget. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* 10:84–94.

38. Talerico, M., and S. M. Berget. 1990. Effect of 5′ splice site mutations on splicing of the preceding intron. *Mol. Cell. Biol.* 10:6299–6305.

39. Stolow, D. T., and S. M. Berget. 1991. Identification of nuclear proteins that specifically being to RNAs containing 5′ splice sites. *Proc. Natl. Acad. Sci. USA.* 88:320–324.

40. Hawkins, J. D. 1988. A survey of intron and exon lengths. *Nucleic Acids Res.* 16:9893–9908.

41. Pruchno, C. J., G. A. Wallis, B. J. Starman, A. Ayleswort, and P. H. Byers. 1989. Inefficient splicing and production of both a normal and shortened mRNA from the same COL1A1 allele of type I collagen in a father and son with osteogenesis imperfecta (OI). *Am. J. Med. Genet.* 45:213a. (Abstr.)

42. Wallis, G. A., B. J. Starman and P. H. Byers. 1989. Clinical heterogeneity of OI explained by molecular heterogeneity and somatic mosaicism. *Am. J. Hum. Gen.* 45:228*a*. (Abstr.)

43. Kuivaniemi, H., C. Sabol, G. Tromp, M. Sippola-Thiele, and D. J. Prockop. 1988. A19-base pair deletion in the pro-alpha 2(I) gene of type I procollagen that causes in-frame RNA splicing from exon 10 to exon 12 in a proband with atypical osteogenesis imperfecta and in his asymptomatic mother. *J. Biol. Chem.* 263:11407–11413.

44. Tromp, G., and D. J. Prockop. 1988. Single base mutation in the proalpha 2(I) collagen gene that causes efficient splicing of the RNA from exon 27 to exon 29 and synthesis of a shortened but in-frame proalpha 2(I) chain. *Proc. Natl. Acad. Sci. USA.* 85:5254–5258.

45. Weil, D., M. D'Alessio, F. Ramirez, W. DeWet, W. G. Cole, D. Chan, and J. F. Bateman. 1989. A base substitution in the exon of a collagen gene causes alternative splicing and generates a structurally abnormal polypeptide in a patient with Ehlers-Danlos syndrome type VII. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:1705–1710.

46. Byers, P. H. 1990. Brittle bones: fragile molecules; Disorders of collagen gene structure and expression. *Trends Genet.* 6:293–300.

47. Weil, D., M. D'Alessio, F. Ramirez, B. Steinmann, M. K. Wirtz, R. Glanville, and D. W. Hollister. 1989. Temperature-dependent expression of a collagen splicing defect in the fibroblasts of a patient with Ehlers-Danlos syndrome Type VII. *J. Biol. Chem.* 264:16804–16809.

48. Weil, D., M. D'Alessio, F. Ramirez, and D. R. Eyre. 1990. Structural and functional characterization of a splicing mutation in the pro-alpha 2(I) collagen gene of an Ehlers-Danlos type VII patient. *J. Biol. Chem.* 265:16007–16011.

49. Byers, P. H., G. A. Wallis, and M. C. Willig. 1991. Osteogenesis imperfecta: translation of mutation to phenotype. *J. Med. Genet.* 28:433–442.

50. Bonadio, J., F. Ramirez, and M. Barr. 1990. An intron mutation in the human alpha 1(I) collagen gene alters the efficiency of pre-mRNA splicing and is associated with osteogenesis imperfecta type II. *J. Biol. Chem.* 265:2262–2268.

51. Dominski, Z., and R. E. Kole. 1992. Cooperation of pre-mRNA sequence elements in splice site selection. *Mol. Cell. Biol.* 12:2108–2114.

52. Nakahashi, Y., H. Fujita, S. Taketani, N. Ishida, A. Kappas, and S. Sassa. 1992. The molecular defect of ferrochelatase in a patient with erythropoietic protoporphyria. *Proc. Natl. Acad. Sci. USA.* 89:281–285.

53. Hamosh, A., B. C. Trapnell, P. L. Zeitlin, C. Montronse-Rafizadeh, B. J. Rosenstein, R. C. Crystal, and G. R. Cutting. 1991. Severe deficiency of cystic fibrosis transmembrane conductance regulator messenger RNA carrying nonsense mutation R553X and W1316X in respiratory epithelial cells of patients with cystic fibrosis. *J. Clin. Invest.* 88:1880–1885.

54. Koivisto, L-M., H. Turtola, K. Aaltoj-Setala, B. Top, R. R. Frants, P. T. Kovanen, A. C. Syvanen, and K. Kontula. 1992. The familial hypercholesterolemia (FH)-North Karelia mutation of the low density lipoprotein receptor gene deletes seven nucleotides of exon 6 and is a common cause of FH in Finland. *J. Clin. Invest.* 90:219–228.

55. Pavlin, D., A. C. Lichtler, A. Bedalov, B. E. Kream, J. R. Harrison, H. F. Thomas, G. A. Gronowicz, S. H. Clark, C. O. Woody, and D. W. Rowe. 1992. Differential utilization of regulatory domains within the α1(I) collagen promoter in osseous and fibroblastic cells. *J. Cell Biol.* 116:227–236.