Am. J. Hum. Genet. 65:336–344, 1999

Redefinition of Exon 7 in the *COL1A1* Gene of Type I Collagen by an Intron 8 Splice-Donor–Site Mutation in a Form of Osteogenesis Imperfecta: Influence of Intron Splice Order on Outcome of Splice-Site Mutation

Ulrike Schwarze, Barbra J. Starman,* and Peter H. Byers

Departments of Pathology and Medicine, University of Washington, Seattle

Summary

Most splice-site mutations lead to a limited array of products, including exon skipping, use of cryptic splice-acceptor or -donor sites, and intron inclusion. At the intron 8 splice-donor site of the *COL1A1* gene, we identified a G+1→A transition that resulted in the production of several splice products from the mutant allele. These included one in which the upstream exon 7 was extended by 96 nt, others in which either intron 8 or introns 7 and 8 were retained, one in which exon 8 was skipped, and one that used a cryptic donor site in exon 8. To determine the mechanism by which exon-7 redefinition might occur, we examined the order of intron removal in the region of the mutation by using intron/exon primer pairs to amplify regions of the precursor nuclear mRNA between exon 5 and exon 10. Removal of introns 5, 6, and 9 was rapid. Removal of intron 8 usually preceded removal of intron 7 in the normal gene, although, in a small proportion of copies, the order was reversed. The proportion of abnormal products suggested that exon 7 redefinition, intron 7 plus intron 8 inclusion, and exon 8 skipping all represented products of the impaired rapid pathway, whereas the intron-8 inclusion product resulted from use of the slow intron 7–first pathway. The very low–abundance cryptic exon 8 donor site product could have arisen from either pathway. These results suggest that there is commitment of the pre-mRNA to the two pathways, independent of the presence of the mutation, and that the order and rate of intron removal are important determinants of the outcome of splice-site mutations and may explain some unusual alterations.

Received April 12, 1999; accepted for publication June 9, 1999; electronically published July 6, 1999.

Address for correspondence and reprints: Dr. Peter H. Byers, Department of Pathology, Box 357470, University of Washington, Seattle, WA 98195-7470. E-mail: pbyers@u.washington.edu

*Died on May 28, 1999.

0002-9297/99/6502-0009$02.00

Introduction

Most human genes have introns that must be removed precisely to create the mature mRNA that encodes the protein product. One common molecular mechanism of human genetic disease involves mutations at splice junctions (Cooper et al. 1995). In mammals, most introns have conserved 5′ (donor) and 3′ (acceptor) sequences that flank exons, a short polypyrimidine tract adjacent to the acceptor site, and a branch point sequence 18–40 nt from the acceptor dinucleotide (Hawkins 1988; Berget 1995). Mutations in these sites can lead to exon skipping, short deletions, or insertions in the mature mRNA. In most models of splicing, the exons are assumed to be independent, so that splicing of one intron occurs without regard to surrounding events. For most genes, however, the order of intron removal is unknown. In the few genes with a small number of introns in which the order of splicing has been examined, it is ordered but not processive (Lang and Spritz 1987; Gudas et al. 1990; Kessler et al. 1993). That is, splicing does not proceed unhindered from the 5′ end to the 3′ end of the initial transcript. Instead, there are rapid and slow splicing events that give rise to major and minor pathways.

Collagen genes are complex, and those that encode the fibrillar proteins have >50 exons (Byers 1995). Many mutations in these genes involve splice-consensus sequences (Kuivaniemi et al. 1997; Schwarze et al. 1997) and lead to a variety of outcomes that depend, in part, on the site of the change. We identified a point mutation in the 5′ donor sequence of intron 8 of the *COL1A1* gene. Unexpectedly, this mutation led to several splice outcomes, one of which was redefinition of the upstream exon 7. Current models would not ordinarily predict such an event (see Robberson et al. [1990] as an example), but, if the order of intron removal in the normal gene and the nature of intron sequences are considered, then the different products and their relative abundance appear to reflect the major and minor splicing pathways in this region. Thus, consideration of the order of intron

removal may clarify why certain unexpected splice outcomes occur.

Material and Methods

Clinical Summary

The proband was the product of an uncomplicated pregnancy and had a normal body length at birth. Between 10 mo and 9 years of age, she sustained several dozen spontaneous fractures to the bones of her legs, hands, and feet. After age 9 years, the fracture frequency decreased dramatically. She was growth-retarded at that point, with a height of 112 cm, which corresponds to her adult height. Her virtual cessation of growth was attributed, in part, to progressive scoliosis and moderate deformity of her lower extremities. Her mobility is reduced, and she spends most of her time in a wheelchair. Her sclerae have remained grayish-blue throughout her life. These clinical findings were thought to be consistent with the diagnosis of moderately severe osteogenesis imperfecta type IV.

Cell Culture and Analysis of Collagenous Proteins

Dermal fibroblasts were obtained from explants of a skin biopsy from the proband, with appropriate consent. Growth and maintenance of cells, radiolabeling of collagenous proteins, and analysis of pro-α chains and α chains by SDS-PAGE were performed as described elsewhere (Bonadio et al. 1985).

RNA Isolation

Total RNA or nuclear and cytoplasmic RNA were prepared from cultured fibroblasts of the proband and a control-cell strain by guanidinium thiocyanate-phenol-chloroform extraction (Chomczynski and Sacchi 1987). Nuclei and cytoplasm were separated by incubating the scraped cell pellet in Triton X-100 buffer (140 mM NaCl; 2 mM EDTA; 10 mM Tris, pH 7.4; 0.8% Triton X-100; 2 mM $MgCl_2$) for 5 min on ice, followed by centrifugation for 10 min at 1000 *g*, 4°C. The supernatant contained the cytoplasmic components. The nuclear pellet was gently vortexed for 5 s, washed in Triton X-100 buffer for 10 s, and centrifuged for 5 min under the same conditions as described above. The supernatant was discarded, and the pellet was saved to isolate nuclear RNA.

Preparation of cDNA and Genomic DNA, and Sequence Determination

All RNA preparations were treated with DNase I (Boehringer Mannheim) for 20 min at 37°C with the incubation buffer recommended by the manufacturer. After ethanol precipitation on dry ice for 15 min, cDNA was synthesized with random hexamers as primers for AMV-reverse transcriptase (Promega). *COL1A1* cDNA was amplified by PCR (Saiki et al. 1987) (denaturation at 94.5°C for 1.5 min, annealing at 67°C for 2 min, and extension at 72°C for 2 min, for 32 cycles) with primers located in exon 6 (E 6F) and exon 10 (E 10R) (table 1), and the cDNA fragments were separated on 6% polyacrylamide gels. Extra PCR bands, derived from the cDNA of the patient, were cut out from the gel and reamplified. Prior to sequence determination by the dideoxy-chain termination method (Sanger et al. 1977) with T7 polymerase (Sequenase® version 2.0, US Biochemicals), the PCR products were purified on a 1%–low-melting-temperature agarose gel (NuSieve GTG agarose; FMC). Genomic DNA was isolated from cultured fibroblasts by standard methods (Sambrook et al. 1989). The PCR was performed with primers located in exon 7 (E 7F) and exon 9 (E 9R) (table 1), under the same conditions used for the amplification of cDNA. The DNA sequence was determined, as above, with the primers used for amplification.

Isolation of Nuclear RNA and cDNA Synthesis

Fibroblasts from a control and from the patient were grown to near confluence in six 150 mm tissue-culture dishes in DMEM with 10% fetal bovine serum. The medium was then replaced by serum-free DMEM, and 5 μg of Actinomycin D (Boehringer Mannheim) per mil-

Table 1

Sequence of Oligonucleotide Primers

Primer	*COL1A1* Sequence 5′→3′
E 5F	AGATGGCATCCCTGGACAGCCTGGACTTC
E 6F	CAGCTGTCTTATGGCTATGATGAGAAATCA
I 6F	CCACATGCTTTCCTACAGAGGGATCACCA
E 7F	GTCCCTCTGGTCCTCGTGGTCTCCCTG
I 7F	TTTGGTCTTTTGGAGGGAAGACTGGGATGA
I 7F′	GTTGACAGGGACTTGTCTTCTAACCCATC
E 8F	CCAAGAATTCCAAGGTCCCCCTGGTGAGCC
I 8F	CACTCTCTATACAGATTCATACTCCTTCTA
E 9F	GGTCCCATGGGTCCCCGAGGTC
I 7R	GATGGGTTAGAAGACAAGTCCCTGTCAAC
E 8R	CTCCAGGGGGACCTTGGAAGCCTTGGGGAC
I 8R	GATAGAAGAGGTGGTTAGAATATGGATAAG
E 9R	TTCCAGGGGGACCTGGGGGACCTCGG
I 9R	GAAGGTCTCAGTCTTCTTTGGATTG
E 10R	CAGGACGACCAGGTTTTCCAGCTTCCCC
I 10R	ATGAAGCCCAAGTGCAGTGAAGCCCA

NOTE.—E indicates an exon sequence, I an intron sequence; F (forward) indicates a sense strand–primer, R (reverse) an antisense strand–primer. Underlined nucleotides indicate introduced mismatches. Sequences were derived from previously published genomic sequence of *COL1A1* (D'Alessio et al. 1988; Bateman et al. 1994) and from our own sequencing data (unpublished).

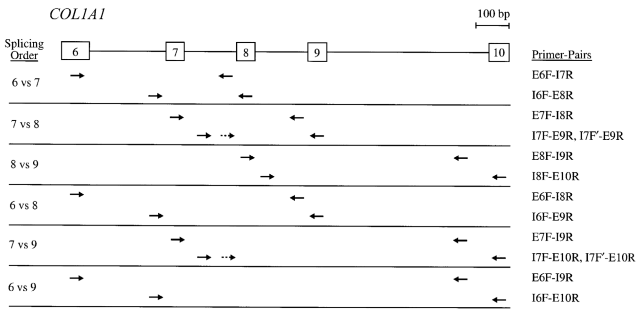

Figure 1 PCR primers used for the determination of the order of adjacent intron removal. (The schematic is drawn to scale.)

liliter was added to halt transcription. The incubation was terminated after 5 min, 10 min, 20 min, 40 min, or 65 min by repeated rinse with ice-cold 1 × PBS to inhibit nuclear mRNA export (Jarmolowski et al. 1994). One dish of each cell line was left untreated to serve as a baseline control. Isolation of nuclear and cytoplasmic RNA from the different timepoints, treatment with DNase I, and preparation of cDNA were performed as described above. The cDNA was synthesized in a 20-μl reaction volume, with 80% of the RNA preparation derived for each timepoint. To check for residual genomic DNA in the RNA preparation, one PCR reaction was done without prior reverse transcription. To be sure that there was no significant transfer of nuclear RNA into the cytoplasmic fraction during the separation procedure, we examined the presence of pre-mRNA in DNase I–treated RNA preparations from both cellular compartments by amplification of the respective cDNA with two intron primers derived from *COL3A1* genomic sequence (Zafarullah et al. 1990).

Determination of the Order of Intron Removal

To determine the order of intron removal between exon 5 and exon 10 of *COL1A1* pre-mRNA, the splicing intermediates were analyzed by a PCR strategy described by Kessler et al. (1993). In brief, to determine the order of splicing of two adjacent introns, two sets of staggered primer pairs were used, each consisting of an exon and an intron primer—the latter being crucial to amplification of pre-mRNA in the presence of the predominant mature mRNA. By extending the sets of overlapping primer pairs over the entire area of interest (fig. 1), the order of intron removal could be inferred. The nucleotide sequences of the primers depicted schematically in figure 1 are listed in table 1. The order of removal of introns 5 and 6 versus intron 7 was determined by use of primers in exon 5 (E 5F) and intron 7 (I 7R). PCR was performed in a 50-μl volume containing 2 μl of the cDNA mixture, 5 μl of 10 × PCR buffer with 20 mM $MgCl_2$, 0.2 mM dNTPs, 1 U of *Taq* polymerase (Boehringer Mannheim), and 10 pmol of each primer. For all reactions, the sense primer was end-labeled with $\gamma[^{32}P]$ATP (3,000 Ci/mmol; Amersham) by T4 polynucleotide kinase (New England BioLabs). Amplifications were done through 27 cycles of denaturation at 94.5°C for 1.5 min, primer annealing at 67°C for 2 min, and primer extension at 72°C for 2 min. Primer combinations E 6F/I 8R, E 7F/I 9R, E 7F/I 8R, and I 8F/E 10R (table 1, fig. 1) required an annealing temperature of 60°C for 1.5 min. All PCR products (10-μl aliquots) were separated on 6% polyacrylamide gels. The gels were dried, and the radioactive bands were visualized by autoradiography. Prior to these experiments, *COL1A1* genomic DNA from control fibroblasts was amplified with the entire set of primer pairs, to characterize the sizes of the products.

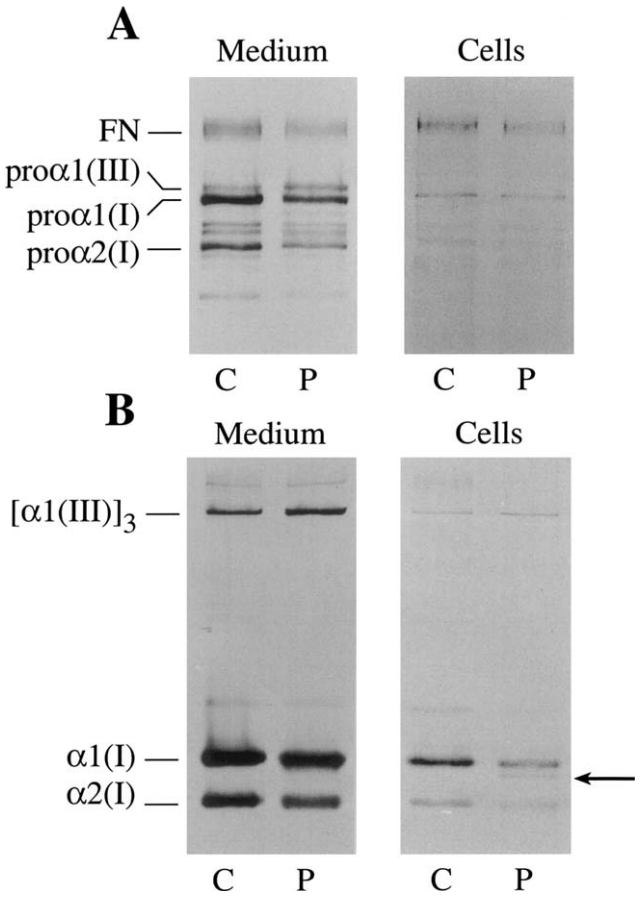

Figure 2 SDS-PAGE of radiolabeled procollagens (*A*) and collagens (*B*) from medium and fibroblast cell layer. C = control; P = proband. Procollagens were analyzed under reducing conditions; collagens were generated by partial proteolysis with pepsin and separated without reduction. The arrow points at a population of α1(I) chains of faster-than-normal electrophoretic mobility that is retained intracellularly.

Results

Analysis of Collagen Production by the Patient's Cultured Fibroblasts

Fibroblasts from the proband synthesized and secreted normal amounts of type III procollagen, but the amount of type I procollagen secreted into the medium was diminished (fig. 2*A*). The electrophoretic mobilities of the pro-α1(I) and pro-α2(I) chains were not altered, nor was there an appreciable amount retained in the cell layer. After partial proteolysis with pepsin of the secreted and intracellular collagens, the population of α1(I) chains was heterogeneous (fig. 2*B*). There was a very small amount of α1(I) chains, with slightly faster-than-normal mobility, secreted into the medium and a larger amount of α1(I) chains, with a more obvious increase in mobility, retained intracellularly.

Characterization of Abnormal Transcripts of COL1A1 *and Identification of the Gene Defect*

The observed protein alterations and the lack of overmodification of the α chains suggested that the molecular defect was located near the 5′ end of the triple-helix coding region of *COL1A1*. To identify the molecular defect, total RNA was reverse transcribed, *COL1A1* mRNA was amplified with primers E 6F and E 10R (table 1), and the products were separated by 6% PAGE. Amplification of the control sample yielded a single fragment of the expected size (241 bp). In the material from the proband, there were several bands in addition to the normal species. To identify the abnormal transcripts and to determine their cellular distribution, nuclear and cytoplasmic RNA from a control and the proband were reverse transcribed, and *COL1A1* mRNA was amplified with γ[^{32}P]-ATP–end-labeled exon primer E 6F and unlabeled primer E10R and was separated on a denaturing 8% polyacrylamide gel (fig. 3). The bands were excised from the gel and reamplified with the initial primers, and their DNA sequences were determined.

The most abundant abnormal transcript in the cells of the proband retained the 5′ 96 nt of the 158-nt intron 7 and skipped the 54-nt exon 8. Export into the cytoplasm appeared to be unimpeded. There were three less abundant abnormal transcripts. In the first, exon 8 was skipped, and the product was exported to the cytoplasm. In the second, intron 8 (163 nt) was retained, and, in the third, both intron 7 and intron 8 were retained. The latter two transcripts were seen exclusively in the nucleus, and in both of these transcripts there was a single G→A nucleotide change at the +1 position of intron 8. In the nucleus there was an additional transcript, with very low abundance, from which the last 26 nt of exon 8 were deleted. Both exported transcripts would produce shortened proteins and no doubt are responsible for the abnormal proteins identified.

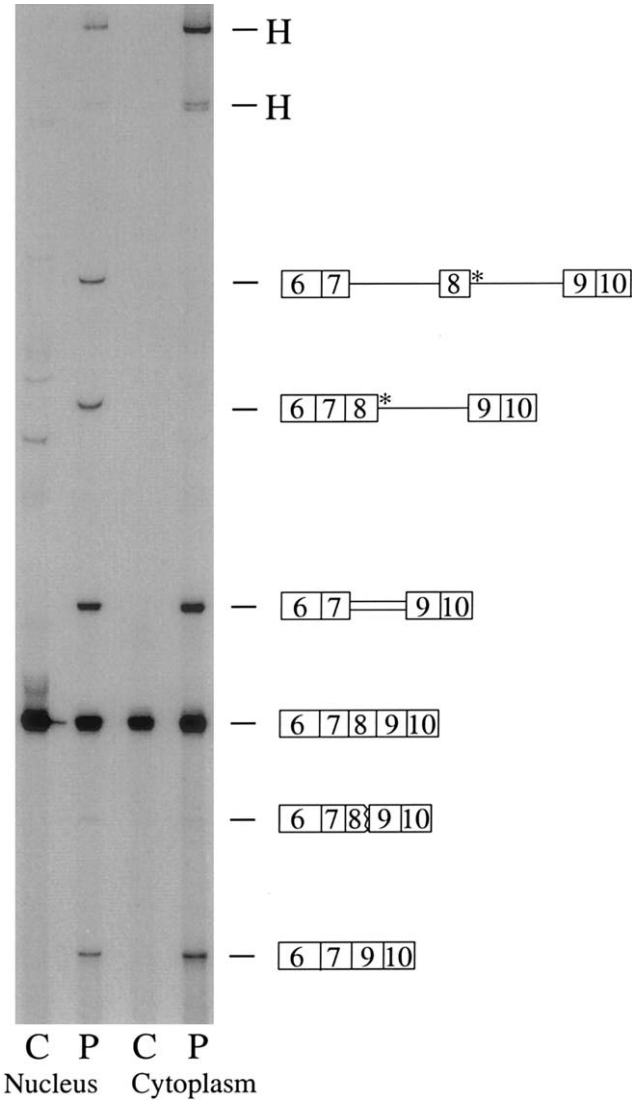

Figure 3 *COL1A1* amplification products of cDNA derived from DNase-treated nuclear and cytoplasmic RNA. C = control; P = proband. The sense primer that annealed to exon 6 sequence (E 6F) was endlabeled using γ[^{32}P]-ATP; the antisense primer (E 10R) remained unlabeled. To improve resolution, separation was performed in a denaturing 8% polyacrylamide gel. A sequencing reaction was run on the same gel and served as a size standard. The abnormal transcripts schematically shown on the right were verified by sequence determination. The schematics are drawn to scale. Single horizontal lines represent full-length introns; the double line denotes the retention of 96 nt of intron 7, resulting in exon 7 redefinition; and the undulated vertical line shows the truncation of exon 8. The asterisk indicates a G→A transition at the +1 position of intron 8. H = heteroduplex.

Genomic DNAs of a control and of the proband were amplified with primers in exon 7 (E 7F) and exon 9 (E 9R) (table 1) and were sequenced directly. The only sequence alteration found was the G→A transition at +1

of the 5′ (donor) splice site in one allele of intron 8 (IVS 8 G+1→A) (fig. 4). There was a cryptic donor splice site within intron 7 (tg/g[+97]taaga, where "/" denotes the site of cleavage) of both the normal allele and the allele that contained the intron 8 mutation. In the mutant allele, this cryptic donor site was used in a portion of the transcripts and resulted in the retention of 96 nt of intron 7 (redefinition of exon 7). The 96-nt insertion encoded an in-frame polypeptide of 32 residues that did not have the canonical Gly-X-Y repeat characteristic of the collagen triple helix (fig. 4*A*). An additional cryptic donor site within exon 8 (TG/GTGAGC) gave rise to truncation of exon 8 by 26 nt (fig. 4*B*).

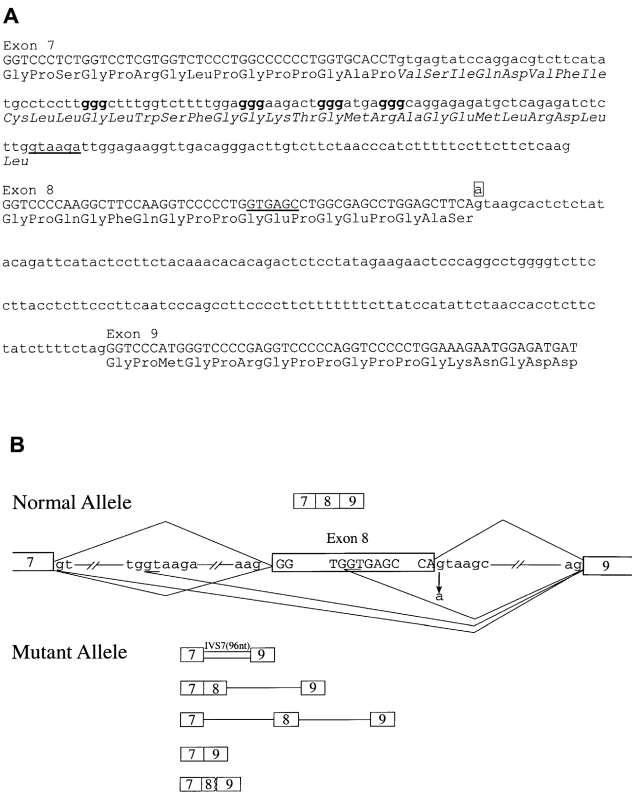

Figure 4 Normal and mutant DNA sequence of *COL1A1*. *A*, The sequence of exons 7, 8, and 9 is shown in uppercase letters and that of introns 7 and 8 in lowercase letters. The splice-donor–site mutation in intron 8 (G+1→A) is pointed out above the normal sequence. The underlined nucleotides represent cryptic splice-donor sites. G-triplets between the constitutive and cryptic intron 7 splice-donor sites are in bold. The deduced noncollagenous amino acid sequence due to exon 7 redefinition is shown in italics. *B*, Schematic representation of part of the *COL1A1* gene illustrating the G→A transition mutation at the donor site of intron 8 in the mutant allele and the resulting alternative splice pathways (diagonal lines). The highly conserved GT dinucleotide is underlined in the sequence of cryptic donor sites present in intron 7 and in exon 8. The aberrant mRNA transcripts derived from the mutant allele are listed so that the most abundant is at the top and the least abundant at the bottom. The three transcripts in between were expressed at similar levels.

The Order of Intron Removal

Although the abnormal mRNA transcripts were identified, it remained unclear why a donor-site mutation induced the activation of a cryptic donor site in the *upstream* intron as the preferred alternative splice pathway, and why this mutation caused "splice paralysis" with nuclear accumulation of splice intermediates. One possibility was that the order of intron removal in this region of the gene determined the selection of alternate donor sites. To characterize the order of intron removal, several intron/exon primer pairs were used to amplify cDNA synthesized from nuclear RNA from cells treated with Actinomycin D for different periods.

When we placed primers in intron 6 and exon 8, using nuclear RNA from the control cells, the vast majority of the amplified product at 0 min contained intron 6 and intron 7, and the minor product contained only intron 6 (fig. 5*A*). There was virtually no transfer of the label from the large to the small fragment over time, which is consistent with removal of intron 6 prior to the excision of intron 7 in most transcripts. If intron 7 were removed first, the majority of radiolabel should be found in the smaller fragment during the chase period. When the primers were placed in exon 6 and intron 8, there was transfer of the label from the largest fragment at 0 min to an intermediate fragment, from which intron 6 had been excised (fig. 5*B*). Although some label was transferred to a fragment from which intron 7 had been excised, the majority of the product was no longer detectable, because the sequence of intron 8 was removed prior to the removal of intron 7. Thus, in most transcripts, the removal of intron 6 preceded the removal of intron 8, which, in turn, preceded the excision of intron 7. When primers were placed in exon 7 and intron 8 (fig. 5*C*), there was some transfer of label from the large to the small fragment, which is consistent with a minor pathway in which intron 7 was removed first. The majority of label simply disappeared between 10 and 20 minutes, which is consistent with rapid excision of intron 8. The observation that intron 8 was removed prior to the excision of intron 7 in the majority of transcripts was further corroborated by placement of primers in intron 7 and exon 9 (fig. 5*D*). As the amount of the full-size fragment decreased over time, there was a concomitant increase in the amount of product from which intron 8 had been excised. Removal of introns 5, 6, and 9 was rapid and preceded excision of intron 8 (data for introns 5 and 9 not shown). These findings were best explained by ordered intron removal in this region of the *COL1A1* allele products (fig. 6). In the major pathway, introns 5, 6, and 9 were excised very rapidly, and we could not determine the precise order of their removal. Subsequently, intron 8 was removed, followed

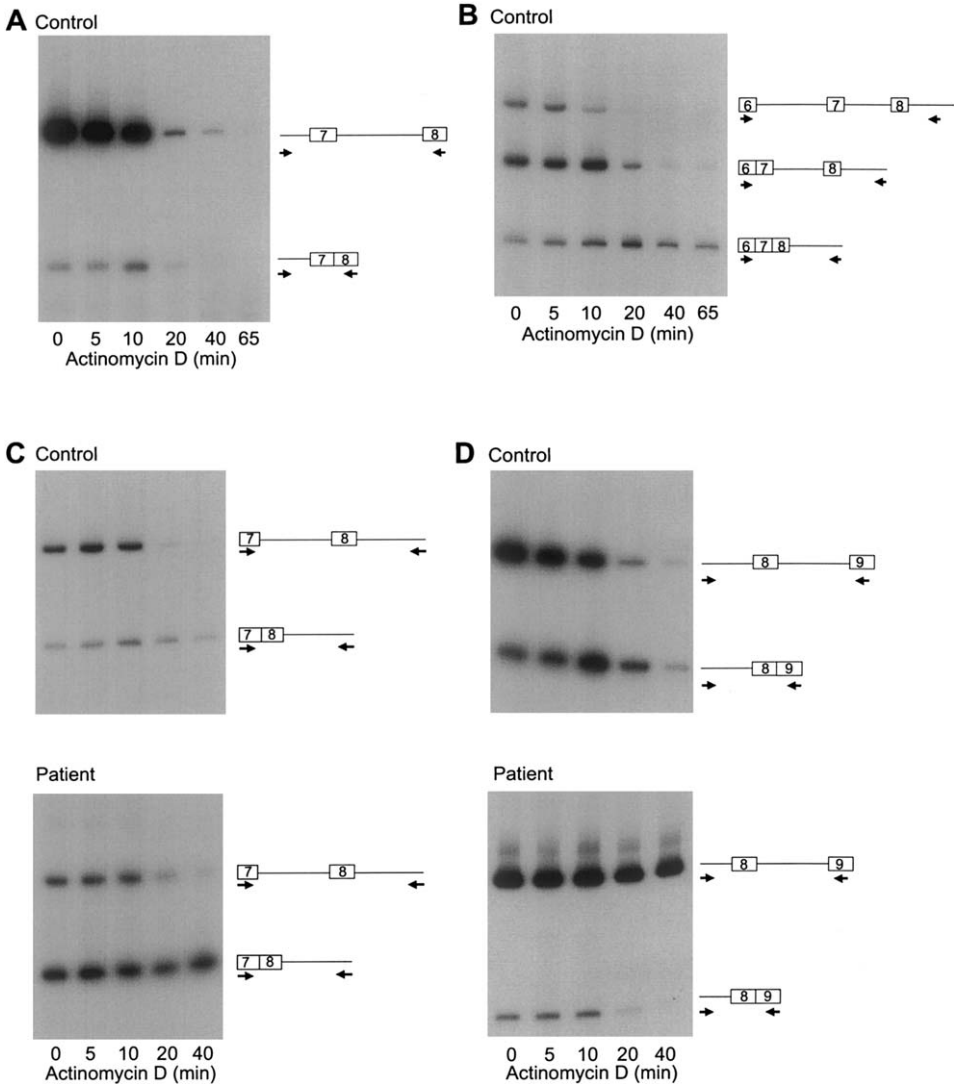

Figure 5 Relative order of removal of introns 6, 7, and 8 of *COL1A1*. Nuclear RNA was prepared from cultured fibroblasts after defined time intervals of exposure to Actinomycin D, treated with DNase, reverse transcribed, amplified with intron-exon primer pairs (the sense primer of which was radioactively endlabeled), and separated by PAGE. The splicing intermediates, visualized by autoradiography, are schematically depicted to scale to the right. Arrows indicate the annealing sites of the amplification primers (table 1, fig. 1). *A*, Control: Intron 6 was excised prior to the excision of intron 7 in most transcripts. *B*, Control: After the removal of intron 6, in the larger proportion of transcripts the excision of intron 8 preceded the excision of intron 7. *C*, Control: The majority of label disappeared between 10 and 20 minutes, consistent with removal of intron 8 prior to the removal of intron 7. Patient: Long-term retention of intron 8 following removal of intron 7. *D*, Control: Removal of intron 8 prior to removal of intron 7 was the predominant pathway of intron excision. Patient: The majority of label remained unchanged due to both intron 7 and intron 8 retention. (The intron 7 primer was placed within the 3′ 62 nucleotides and corresponds to primer I 7F′ in table 1 and fig. 1.)

by excision of intron 7. In the minor pathway, intron 7 removal preceded excision of intron 8.

When we studied cells from the patient, using primers in exon 7 and intron 8 (fig. 5*C*), there was accumulation of product with prolonged retention of intron 8 after removal of intron 7, which corresponded to one of the products retained in the nucleus. When an intron 7 primer placed within the 3′ 62 nt of the intron (to assure that it would be removed in the course of normal splicing) and an exon 9 primer were used (fig. 5*D*), a small amount of label was lost, which is consistent with normal splicing, presumably of the normal allele. The majority of label, however, remained in the initial product for the full duration of the chase, which indicated that both intron 7 and intron 8 were retained, as was confirmed by sequence determination of the amplified product. This corresponded to the other of the two retained nuclear products.

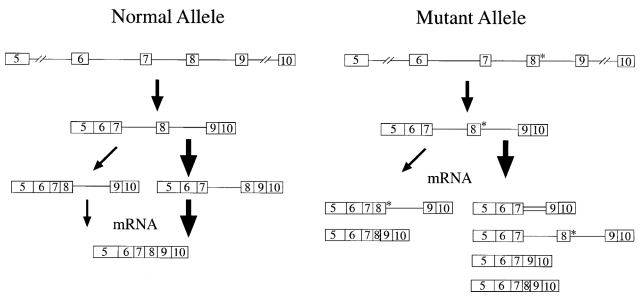

Figure 6 Minor and major pathway of intron removal between exon 5 and exon 10 of the normal and mutant allele of *COL1A1*. Excision of introns 5, 6, and 9 was rapid, and the precise order of their removal could not be determined. In the major pathway (wide arrow) intron 8 was removed prior to intron 7; in the minor pathway (thin arrow) intron 7 removal preceded excision of intron 8. The different splice products derived from the mutant allele (the site of the mutation is indicated by an asterisk) and their relative proportion reflected splicing via the major and minor pathway and suggested commitment of the pre-mRNA to the two pathways independent of the presence of the mutation.

Discussion

Splice-donor–site mutations most commonly lead to exon skipping, to use of cryptic sites within the downstream intron or an adjacent exon, or to intron inclusion if the intron is small (Talerico and Berget 1990). Splice-acceptor–site mutations often lead to the use of cryptic acceptor sites within the downstream exon or, less frequently, within the upstream intron (Cooper et al. 1995; Robberson et al. 1990). In some instances, splice-acceptor–site mutations lead to exon skipping or (rarely) give rise to inclusion of the upstream intron (e.g. Schwarze et al. 1997). The abundance and stability of mRNA generated by the use of alternative sites depends on whether the products are in-frame, whether they lead to premature termination codons within the coding sequences, and whether the termination codon is succeeded by an intron (Maquat 1996; Carter et al. 1996).

To initiate splicing, the 5′ splice-donor site, the branch point, and the adjacent polypyrimidine tract of the 3′ splice-acceptor region are recognized by elements of the spliceosome in a process in which commitment to splicing occurs (Reed and Palandjian 1997). After the assembly of the spliceosome, a complex set of activation steps and rearrangements leads to scission, lariat formation, 3′-end cleavage, and ligation to juxtapose the coding elements in the mature mRNA (Staley and Guthrie 1998). In most models of splicing (see Berget 1995, for example), only the removal of single introns is considered, and the implicit assumption is that splicing occurs in a generally 5′ to 3′ direction in the precursor mRNA (Lang and Spritz 1983). If splicing is not processive, then one or more regions of the precursor mRNA may contain introns surrounded by large domains of already spliced sequences. Under these circumstances, the available choices for the splicing machinery may be more restricted or may require a more imaginative cellular process to create an in-frame mRNA in the face of donor- or acceptor-site mutations. In those small genes that have been studied, such as β-globin (Lang et al. 1985), rat α-lactalbumin and murine interleukin-3 (Lang and Spritz 1987), thymidine kinase (Gudas et al 1990), and adenine phosphoribosyltransferase and dihydrofolate reductase (Kessler et al 1993), intron removal appears to be orderly, but not strictly processive, from the 5′ end to the 3′ end of the precursor mRNA. Furthermore, these studies indicate that there are major and minor pathways with a preferred, but not absolute, order of intron removal. There are no published data on the order of intron removal in larger, more complex genes, like *COL1A1*.

In the region of the normal *COL1A1* gene encompassed by exons 5–10, we found that introns 5, 6, and 9 were removed quickly. Intron 8 was then usually spliced before intron 7, although there was a minor pathway in which intron 7 was removed first (fig. 6). In the cells from the patient we describe here (and one previously described by Bateman et al. 1994), a splice-donor mutation in intron 8 of the *COL1A1* gene of type I collagen had several effects: (1) redefinition of the upstream exon 7, in the absence of changes in the preceding intron or in the sequence of either exon 7 or intron 7 itself; (2) skipping of exon 8; (3) use of a cryptic donor site in exon 8; (4) inclusion of intron 8; and (5) retention of both intron 7 and intron 8. This mutation might be expected to skip exon 8 because of failure of exon definition (Robberson et al. 1990). With a small intron, inclusion might occur, and, on occasion, an alternative donor site could be used within either the intron or the preceding exon. Redefinition of the upstream exon does not fit with most splicing models (Berget 1995). It does, however, correspond to products that could be predicted if the order of intron removal is considered.

Because of the speed with which the surrounding introns are removed, in the precursor mRNA there is a "large" exon upstream of exon 7, followed by intron 7, exon 8, intron 8, and then spliced product of (at least) exons 9 and 10. With a mutation in the constitutive donor site of intron 8, the weak donor site within the exon, the strong cryptic site within intron 7 and the constitutive intron 7 donor site could attract U1 snRNP, the first component of the 5′ complex (Michaud and Reed 1993). If intron 7 is spliced first, then the only splicing alternatives that involve intron 8 are the use of the exon 8 cryptic donor site or intron inclusion (fig. 6). In the molecules "committed" to the intron 8–first pathway, several alternatives can be envisaged. Splice "paralysis" could lead to retention of both intron 7 and

intron 8. Competition between the two strong splice-donor sites in intron 7 would be expected to lead to exon 7 redefinition (use of the cryptic site) or skipping of exon 8 (use of the constitutive site). Use of the exon 8 cryptic donor site would give rise to exon truncation (fig. 6). The relative abundance of all the splice products in the cell strongly suggests that there is commitment of the pre-mRNA to the two pathways, independent of the presence of the mutation, and that each product is derived primarily from one of the pathways.

Although the cryptic intron 7 donor site is used frequently, the constitutive site remains active. One possible explanation for its activity is the presence of several G-triplets in the immediate downstream region (fig. 4*A*). McCullough and Berget (1997) demonstrated the role of these motifs in directing the splicing machinery to the adjacent upstream sites, and Cogan et al. (1997) and McCarthy and Phillips (1998) showed that deletion of a single G from a G-triplet in intron 3 of the human-growth-hormone gene led to skipping of the upstream exon. An alternative possibility is that the relative length of the binding tracts of U1 snRNA and U6 snRNA, which can influence the efficiency of 5′ donor site splicing (Murray and Jarrell 1999; Staley and Guthrie 1999), permits competition between the two sites.

The factors that control the order of intron removal are not clear. The rate of intron removal could be influenced by the "strength" of the donor and acceptor sites, whether there are alternative or cryptic sites within the introns (or exons), the presence of splice enhancers or suppressors, the presence of SR protein–binding sites, or the size of adjacent intron/exon units and whether they permit simultaneous occupancy by spliceosomes. The extent to which these *cis*-acting elements contribute to the order of intron removal will likely depend on how accessible they are to *trans*-acting factors. Hence, RNA conformation also needs to be considered.

Most splice-site mutations produce only a single outcome, although few are subjected to the scrutiny employed here. There are a number of instances in which the outcome of splice mutations can be complex, with multiple products, and other instances in which the same splice-site alteration (e.g., Kuivaniemi et al. 1990) in different introns may have different effects. From these indicators, it seems likely that in those regions the involved genes may, like the region of the *COL1A1* gene studied here, be processed through major and minor pathways and that use of these pathways will explain the multiple outcomes.

Although the possible outcomes of splice mutations can be predicted, it is difficult to determine, on the basis of sequence changes alone, which of one or more possible choices will be made. If the phenotype is known, such prediction may be of little consequence. However, as mutation analysis moves to genomic characterization alone and predictive information is sought, it would be helpful to have models that can predict splice outcome. For example, in the *COL1A1* gene, both the IVS47G+1→A and the IVS48G+1→A would be expected to lead to exon skipping and produce the lethal OI type II phenotype. Although the IVS47 mutation does just that (unpublished observations), the IVS48 mutation results in the OI type I phenotype, apparently because the use of an alternative donor site leads to a frame shift and an unstable mRNA (Willing et al. 1994). While it is a tall order, a detailed analysis of splice order in complex genes like *COL1A1* may produce sufficient insight to provide a predictive model that could be applied to many genes in which knowledge of the effect of a splice-site mutation would have implications for therapy or counseling.

Acknowledgments

This work was supported in part by National Institutes of Health grant AR41223. The paper is dedicated to the memory of Barbra Starman.

References

Bateman JF, Chan D, Moeller I, Hannagan M, Cole WG (1994) A 5′ splice site mutation affecting the pre-mRNA splicing of two upstream exons in the collagen *COL1A1* gene. Biochem J 302:729–735

Berget SM (1995) Exon recognition in vertebrate splicing. J Biol Chem 270:2411–2414

Bonadio J, Holbrook KA, Gelinas RE, Jacob J, Byers PH (1985) Altered triple helical structure of type I procollagen in lethal perinatal osteogenesis imperfecta. J Biol Chem 260: 1734–1742

Byers PH (1995) Disorders of collagen structure and synthesis. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) The metabolic and molecular bases of inherited disease, 7th edition. McGraw-Hill, Inc., New York, pp 4029–4078

Carter MS, Li S, Wilkinson MF (1996) A splicing-dependent regulatory mechanism that detects translation signals. EMBO J 15:5965–5975

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159

Cogan JD, Prince MA, Lekhakula S, Bundey S, Futrakul A, McCarthy EMS, Phillips JA III (1997) A novel mechanism of aberrant pre-mRNA splicing in humans. Hum Mol Genet 6:909–912

Cooper DN, Krawczak M, Antonarakis SE (1995) The nature and mechanisms of human gene mutation. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) The metabolic and molecular bases of inherited disease, 7th edition. McGraw-Hill, Inc., New York, pp 259–292

D'Alessio M, Bernard M, Pretorius PJ, De Wet W, Ramirez F (1988) Complete nucleotide sequence of the region encompassing the first twenty-five exons of the human proa1(I) collagen gene (*COL1A1*). Gene 67:105–115

Gudas JM, Knight GB, Pardee AB (1990) Ordered splicing of thymidine kinase pre-mRNA during the S phase of the cell cycle. Mol Cell Biol 10:5591–5595

Hawkins JD (1988) A survey on intron and exon lengths. Nucleic Acids Res 16:9893–9908

Jarmolowski A, Boelens WC, Izaurralde E, Mattaj IW (1994) Nuclear export of different classes of RNA is mediated by specific factors. J Cell Biol 124:627–635

Kessler O, Jiang Y, Chasin LA (1993) Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. Mol Cell Biol 13:6211–6222

Kuivaniemi H, Kontusaari S, Tromp G, Zhao M, Sabol C, Prockop DJ (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. J Biol Chem 265: 12067–12074

Kuivaniemi H, Tromp G, Prockop DJ (1997) Mutations in fibrillar collagens (types I, II, III, and XI), fibril-associated collagen (type IX), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels. Hum Mutat 9:300–315

Lang KM, Spritz RA (1983) RNA splice site selection: evidence for a 5′→3′ scanning model. Science 220:1351-1355

——— (1987) In vitro splicing pathways of pre-mRNAs containing multiple intervening sequences. Mol Cell Biol 7: 3428–3437

Lang KM, van Santen VL, Spritz RA (1985) The two intervening sequences of human β- and γ-globin pre-mRNAs are excised in a preferred temporal order *in vitro*. EMBO J 4: 1991–1996

Maquat LE (1996) Defects in RNA splicing and the consequence of shortened translational reading frames. Am J Hum Genet 59:279–286

McCarthy EMS, Phillips JA III (1998) Characterization of an intron splice enhancer that regulates alternative splicing of human GH pre-mRNA. Hum Mol Genet 7:1491–1496

McCullough AJ, Berget SM (1997) G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. Mol Cell Biol 17: 4562–4571

Michaud S, Reed R (1993) A functional association between the 5′ and 3′ splice sites is established in the earliest prespliceosome complex (E) in mammals. Genes Dev 7: 1008–1020

Murray HL, Jarrell KA (1999) Flipping the switch to an active spliceosome. Cell 96:599–602

Reed R, Palandjian L (1997) Spliceosome assembly. In: Krainer AR (ed) Eukaryotic mRNA processing. IRL Press, Oxford, pp 103–129

Robberson BL, Cote GJ, Berget SM (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol Cell Biol 10:84–94

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

Schwarze U, Goldstein JA, Byers PH (1997) Splicing defects in the *COL3A1* gene: marked preference for 5′ (donor) splice-site mutations in patients with exon-skipping mutations and Ehlers-Danlos syndrome type IV. Am J Hum Genet 61:1276–1286

Staley JP, Guthrie C (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell 92:315–326

——— (1999) An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p. Mol Cell 3:55–64

Talerico M, Berget SM (1990) Effect of 5′ splice site mutations on splicing of the preceding intron. Mol Cell Biol 10: 6299–6305

Willing MC, Deschenes SP, Scott DA, Byers PH, Slayton RL, Pitts SH, Arikat H, et al (1994) Osteogenesis imperfecta type I: molecular heterogeneity for COL1A1 null alleles of type I collagen. Am J Hum Genet 55:638–647

Zafarullah K, Kleinert C, Tromp G, Kuivaniemi H, Kontusaari S, Wu Y, Ganguly A, et al (1990) G to A polymorphism in exon 31 of the *COL3A1* gene. Nucleic Acids Res 18:6180