Am. J. Hum. Genet. 61:1276–1286, 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

Ulrike Schwarze, Jayne A. Goldstein, and Peter H. Byers

Departments of Pathology and Medicine, University of Washington, Seattle

Summary

Ehlers-Danlos syndrome (EDS) type IV results from mutations in the *COL3A1* gene, which encodes the constituent chains of type III procollagen. We have identified, in 33 unrelated individuals or families with EDS type IV, mutations that affect splicing, of which 30 are point mutations at splice junctions and 3 are small deletions that remove splice-junction sequences and partial exon sequences. Except for one point mutation at a donor site, which leads to partial intron inclusion, and a single base-pair substitution at an acceptor site, which gives rise to inclusion of the complete upstream intron into the mature mRNA, all mutations result in deletion of a single exon as the only splice alteration. Of the exon-skipping mutations that are due to single base substitutions, which we have identified in 28 separate individuals, only two affect the splice-acceptor site. The underrepresentation of splice acceptor–site mutations suggests that the favored consequence of 3′ mutations is the use of an alternative acceptor site that creates a null allele with a premature-termination codon. The phenotypes of those mutations may differ, with respect to either their severity or their symptomatic range, from the usual presentation of EDS type IV and thus have been excluded from analysis.

Introduction

Ehlers-Danlos syndrome (EDS) type IV is an uncommon dominantly inherited disorder that is due to mutations in the *COL3A1* gene, which encodes type III procollagen, and is characterized by a predilection for bowel and arterial rupture and occasionally is complicated by rupture of the uterus during pregnancy. These complications generally lead to a shortened life span (Steinmann et al. 1993).

The *COL3A1* gene, located at 2q31-q32, contains 51 exons distributed over 44 kb (Chu and Prockop 1993). The gene encodes a protein of 1,467 amino acids, of which 1,029 are located within the core triple-helical domain characterized by the presence of glycine in every third position (Gly-X-Y, in which X and Y are any amino acid other than tryptophan and in which cysteine is confined to the last residue of the triple helix; Y is often hydroxyproline). The triple-helical domain is encoded by portions of 44 exons. Of these, 42 are cassettes that begin with a glycine codon and end with a Y-position codon, so that deletion of a single exon would result in an in-frame but shortened protein. Because type III procollagen is a homotrimer, the synthesis of an equal number of normal and abnormal chains results in production of a 7:1 ratio of abnormal:normal molecules. If the mutant *COL3A1* allele results in an exon-skipping event, the abnormal molecules contain one, two, or three shortened chains. The sites at which these alterations affect molecular function include molecular folding, secretion, and matrix formation (Smith et al. 1997).

EDS type IV has been proposed as a model for more-common forms of arterial aneurysms, which often cluster in families, but, to date, most studies have excluded the *COL3A1* gene as the locus for such mutations, in the absence of some findings of EDS type IV (Kuivaniemi et al. 1993; Tromp et al. 1993). In individuals with EDS type IV, approximately one-third of published mutations result in exon skipping, the majority of the remainder lead to substitution for single glycine residues within the triple-helical domain, and a small number are larger genomic deletions (Kuivaniemi et al. 1997). We have now identified in 33 individuals mutations that affect splicing, of which 30 are point mutations at splice junctions and 3 are small deletions that remove splice-junction sequences and partial exon sequences. Of the exon-skipping mutations that are due to single base substitutions and that we have identified in 28 separate individuals, only 2 affect the splice-acceptor site. This strong bias

Received June 20, 1997; accepted for publication October 2, 1997; electronically published November 14, 1997.

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

0002-9297/97/6106-0011$02.00

Table 1

Clinical Summary of 33 Studied Individuals Affected with EDS Type IV

					Skin		Blood Vessels			Complications				
Patient (Sex; Age [years][a])	Exon	Family History[b]	Face[c]	Joints[d]	Thin/ Translucent	Cigarette-Paper Scars	Bruising/ Bleeding	Venous Varicosities	Aneurysms	Arterial Rupture	Bowel Rupture	Spleen Rupture	Pneumothorax	Cause of Death (Age [Years])[e]
91-364 (F; 25)	7	−					+		+	+				A (25)
92-075 (M; 23)	8	−		+	−			−	+	+		+	+	
94-597 (F; 21)	8	+			+		+	+						
96-302 (F; 32)	9	−		+	+	+	+					+		
86-053 (M; 10)	9	+		+	+		−							
94-341 (F; 45)	9	+		+	+		+			+				
93-272 (M; 34)	14	−		+	+	+	+	+	+	+				
93-104 (F; 12)	14	+	+	+	+									
90-209 (M; 23)	17	+			+		+	+		+	+			A (24)
91-589 (F; 30)	18	−		+	+	+	+				+			
93-200 (F; 13)	18	−		+		+	+							
91-336 (M; 17)	22	−	+	+	+	+	+	+	+	+				
94-190 (F; 6)	24	−		+	+	+	+							
88-039 (F; 27)	24	−	+	+	+		+		+					
94-750 (M; 36)	24					+			+					
96-404 (F; 15)	24			−	+	+	+	+					+	
95-851 (M; 30)	27	−	+							+				A (30)
96-474 (F; 4)	28		+				+							
95-563 (F; 25)	30	−	+	+	+		+							
93-177 (F; 19)	34	−	+	+	+		+	+		+	+			A (20)
78-042 (M; 29)	34	+			+		+	+	+	+	+		+	A (32)
93-711 (F; 25)	37	+	+	+	+		+			+				A (25)
94-328 (F; 20)	38	+	+	+	+		+							
93-344 (F; 13)	38	+		+	+		+				+			
94-771 (M; 11)	41	−		+	+	+	+							
94-326 (F; 3)	41			+	+		+							
94-722 (M; 9)	42				+		+				+			
93-487 (M; 30)	43	−					+			+				
91-341 (M; 16)	43	−		+	+	−	+			+			+	
78-000 (F)	45	−								+				A (41)
91-062 (M; 24)	45	+		+	+	+	+		+				+	
92-011 (M; 20)	47	+			+	−					+			
95-229 (M; 13)	48	−				+	+			+				A (13)

Note.—A plus sign (+) denotes that the trait was present; and a minus sign (−) denotes that the trait was absent (a blank denotes that no information was available).

[a] Age shown is that when, on the basis of clinical grounds, diagnosis of EDS type IV was made.

[b] A plus sign (+) denotes that one parent (and other family members) were affected; and a minus sign (−) denotes that neither parent was affected.

[c] Characteristic facial features include a thin pinched nose, thin lips, prominent-appearing eyes, hollow cheeks, and tightness of the skin over the face.

[d] Mild hypermobility limited to the small joints of the hands and feet.

[e] A = aortic dissection or arterial rupture.

toward the identification of 5′ splice-site mutations may reflect different consequences of 3′ mutations that result in less severe phenotypes, which may not be identified in the clinical range of EDS type IV.

Subjects, Material, and Methods

Clinical Summary

Clinical features of the 33 unrelated index individuals with EDS type IV who are reported in this study are represented in table 1. The clinical diagnosis was confirmed by biochemical demonstration of defects in the synthesis and secretion of type III procollagen. Of 28 index individuals for whom information on the parents was available, 11 had a family history compatible with the disease, and 17 were the first affected individuals in their families.

In 13 individuals, the diagnosis was made in childhood or adolescence; 2 of them (86-053 and 93-104) came to medical attention because of their positive family history rather than because of symptoms characteristic of EDS type IV. The other 11 individuals had a history of severe bruising or bleeding tendency, and 6 of them had presented to the emergency room with spontaneous bowel rupture (93-344 [see Kinnane et al. 1995] and 94-722), pneumothorax (96-404), or arterial rupture (91-336, 91-341, and 95-229). One (95-229) succumbed to an arterial rupture at age 13 years.

Fifteen adults required hospitalization for major clinical complications, and seven of them died from aortic dissection or arterial rupture at 24–41 years of age. Clinical findings in the affected daughter of patient 78 have been described by Weinbaum et al. (1987), and the family is included in the study of linkage of the *COL3A1* gene to the EDS type IV phenotype and the potential for prenatal diagnosis (Tsipouras et al. 1986).

Cell Culture and Analysis of Collagenous Proteins

Dermal fibroblasts were obtained from explants of skin biopsies from the 33 index individuals, with appropriate consent. Growth and maintenance of those and control-cell strains, radiolabeling of collagenous proteins, and analysis of proα chains and α chains by SDS-PAGE were performed as described elsewhere (Bonadio et al. 1985).

Preparation of cDNA and Genomic DNA, and Sequence Determination

Total cellular RNA was isolated from dermal fibroblasts of the probands (Chomczynski and Sacchi 1987), and cDNA was synthesized by use of random hexamers as primers for AMV-reverse transcriptase (Promega). PCR (Saiki et al. 1988) was used to amplify 15 overlapping regions of cDNA (average size 456 nt [range 400–563 nt]), which covered the complete coding sequence of the proα1(III) chain. The sequences of the oligonucleotides used as primers were derived from the previously published cDNA sequence for *COL3A1* (Ala-Kokko et al. 1989; Benson-Chanda et al. 1989) and are available on request from the authors. The amplified cDNA fragments were separated on 6% polyacrylamide gels. Prior to sequence determination by the dideoxy-chain termination method (Sanger et al. 1977) using T7 polymerase (Sequenase version 2.0; US Biochemicals), the PCR products were either purified on a 1% low-melting-temperature agarose gel (NuSieve GTG agarose; FMC) or were directly cloned into the PCR II vector, according to instructions provided in the TA cloning kit (Invitrogen). Genomic DNA was isolated from cultured fibroblasts by standard methods (Sambrook et al. 1989). The PCR was performed by use of 5′ and 3′ primers located within the adjacent exons upstream and downstream to an exon deleted from the cDNA. Sequence determination was accomplished as described above. The sequencing primers were either identical to the amplification primers, or, if long introns surrounded the deleted exon, nested primers derived from intron sequences were used.

Mutation Confirmation by Restriction Analysis

Digestion with restriction endonucleases (New England BioLabs/Boehringer) of PCR fragments derived from genomic DNA was performed to verify the presence of point mutations, which were predicted to create or disrupt endonuclease cleavage sites. In some cases an oligonucleotide that contained one mismatched nucleotide (mismatch primer) was used to introduce an endonuclease-restriction site during PCR.

Results

Characterization of mRNA Alterations Encoding Mutant Proα1 Chains of Type III Procollagen

To determine the structure of the mRNAs that encoded the proα1(III) chains of type III procollagen in cells from 33 individuals with EDS type IV, the *COL3A1* mRNA was reverse transcribed, and the product was amplified in overlapping fragments. After separation by PAGE, control cDNA yielded one normal-size fragment per amplification, whereas in 31 patients one cDNA region was detected from which one normal and a second, faster-migrating fragment were derived in equal amounts, consistent with heterozygosity for a small deletion (fig. 1). The boundaries of the deletions were identified by DNA sequence determination, and in each instance the faster-migrating fragment lacked a nucleotide sequence that corresponded precisely to the sequence of a single exon within the triple-helical domain. Exons 9

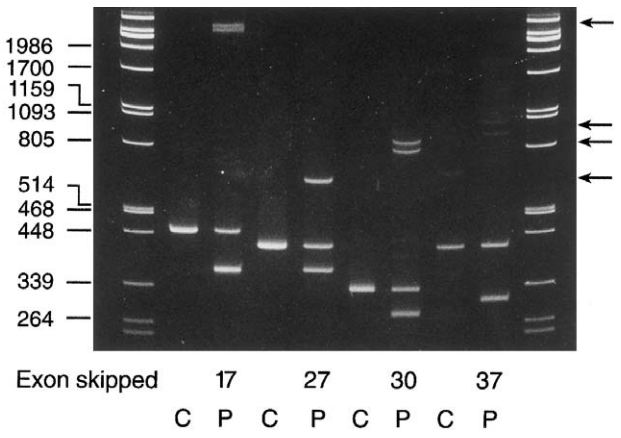

Figure 1 *COL3A1* reverse-transcriptase–PCR products from an unaffected individual (lanes C) and four patients (lanes P) with EDS type IV and exon-skipping mutations. Exon 17 = 99 bp (patient 90-209), exon 27 = 54 bp (patient 95-851), exon 30 = 45 bp (patient 95-563), and exon 37 = 108 bp (patient 93-711). The arrows on the right indicate the positions of heteroduplexes. Molecular-weight marker = −DNA × *Pst*I.

and 24 were found to be deleted in three and four unrelated patients or families, respectively. Sequences encoded by exons 8, 14, 18, 34, 38, 41, 43, and 45 were each deleted in two unrelated patients or families, whereas those of exons 7, 17, 22, 27, 28, 30, 37, and 47 were each identified in one patient or family (fig. 2*A*). The translational reading frame of the mRNA was maintained, and the carboxyl-terminal propeptide of the proα1(III) chain was left intact, to facilitate molecular assembly. Of these mutations, deletion of exons 8, 9, 18, 28, 30, 38, and 47 had not been identified previously (fig. 2*B*).

With the two remaining patients, amplification of two different cDNA regions yielded one normal-size fragment and a second fragment of slower mobility, in equal amount, consistent with heterozygosity for a small insertion (fig. 3). Sequence determination of the abnormal fragment from the first patient (94-722) revealed an inclusion of the first 30 bp of intron 42. In the second patient (95-229), a 96-bp sequence was included in the mRNA between exons 47 and 48. In each case the translational reading frame was maintained, but the Gly-X-Y pattern of the triple helix was interrupted.

Identification of COL3A1 *Mutations Resulting in Exon Deletions*

To characterize the defect at the gene level and thereby to elucidate the cause of the exon-deletion events, *COL3A1* genomic DNA was amplified to include the exon missing from the cDNA. In all but three patients (93-104, 91-336, and 91-589), amplification yielded a single PCR product, which suggested that the mRNA deletion was caused by defective splicing of pre-mRNA. Direct sequence determination disclosed single base changes in all 28 patients whose DNA yielded a single band (fig. 2*A*). All but two point mutations identified in the present study (i.e., 26 of 28) caused disruption of the 5′ (donor) splice site. The G nucleotide at position +1 of the consensus sequence (Krawczak et al. 1992) was changed most frequently, with a predominance of transition mutations. The T nucleotide at position +2 and the G nucleotide at position +5 were changed at approximately equal frequency, with a slight preponderance of transition mutations over transversions. Mutations at positions −1 (the last position of the exon) and +3 were found in one patient each, and none was found at the fourth or sixth position. The two examples of mutations that disrupted the 3′ (acceptor) splice site were a G→C substitution at position −1 of intron 17 (93–200) and an A→C substitution at position −2 of intron 46 (92–011) (fig. 2*A*). The same IVS24G^{+1}→A mutation was identified in four unrelated families and had been identified previously in an additional family (Pope et al. 1996). The mutation affects a CpG site, as does the recurrent IVS20G^{+1}→A identified by others (Kontusaari et al. 1990; Anderson et al. 1993). Three other mutations in this collection have also been identified in other families, IVS37T^{+2}→C (Richards et al. 1994), the recurrent IVS41G^{+1}→A (Cole et al. 1990), and IVS45G^{+1}→A (Pope et al. 1996).

COL3A1 genomic-amplification products from the three remaining patients exhibited, in addition to the normal-size fragment, a second, shorter fragment in patients 93-104 and 91-336 and a longer fragment in patient 91-589. The nucleotide sequences from the fragments of altered size are depicted in figure 4. A 15-bp deletion in patient 93-104 (fig. 4*A*) removed the 5′ (donor) splice site of intron 14. The exact position of the genomic deletion could not be determined, because there was a 5-bp direct repeat that flanked the deleted sequence. There was a 57-bp genomic sequence deleted in patient 91-336 (fig. 4*B*), which obliterated the 3′ (acceptor) splice site of intron 21 and led to skipping of exon 22. A third patient (91-589; fig. 4*C*) had a complex deletion/insertion event that eliminated the 3′ (acceptor) splice site of intron 17 and resulted in skipping of exon 18. In this case, one 5-bp (cctgg) and one 6-bp (gtatac) direct repeat was found at multiple sites in the deleted/inserted sequence and in the flanking unaltered sequence.

Identification of COL3A1 *Mutations Leading to Inclusion of Intron Sequence in mRNA*

To identify, at the gene level, the defect that led to partial or complete intron retention, *COL3A1* genomic DNA was amplified to include the retained intron and

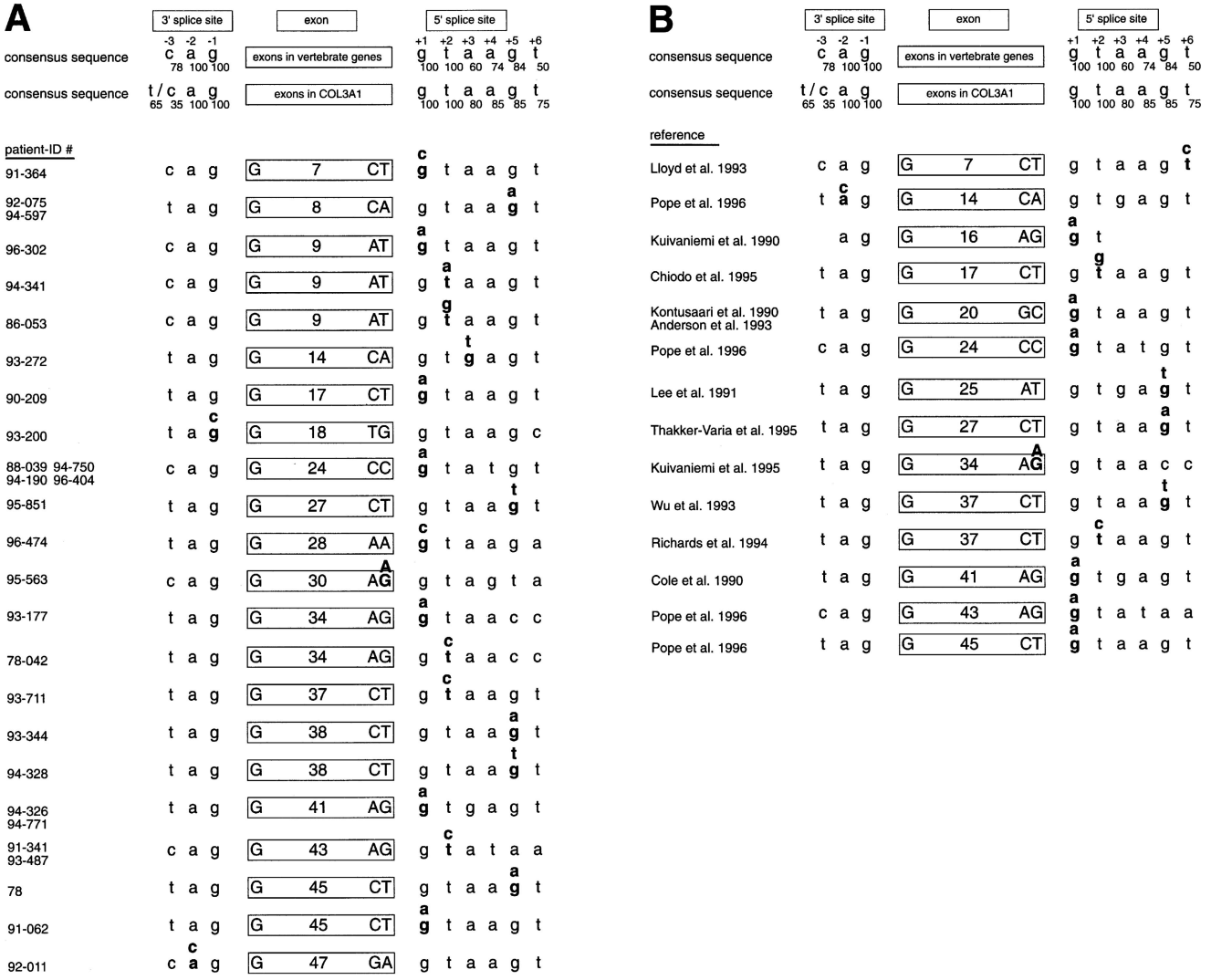

Figure 2 Point mutations at splice sites in *COL3A1*, resulting in exon deletions (*A*) identified in the present study and (*B*) reported in the literature. Subscript numerals in the splice-site consensus sequences refer to percentage frequency of occurrence. The consensus sequences in vertebrate genes have been reported by Padgett et al. (1986), and the *COL3A1*-specific consensus sequences are derived from characterized splice junctions of 20 exons. Exon sequences are signified by capital letters, and intron sequences are signified by lowercase letters; nucleotides at the site of mutation are signified by boldface lowercase letters.

the adjacent exons. In patient 94-722 sequence analysis revealed a G→A transition at the +1 position of the 5′ (donor) splice site of intron 42, which resulted in read-through to a cryptic donor site 30 nt downstream (AG/GTAGAA, where "/" denotes the site of cleavage). The same mutation, identified in an unrelated individual, had been reported previously (Kuivaniemi et al. 1990). In patient 95-229 the entire intron 47 (96 nt) was retained in the mRNA after an A→T transversion at the −2 position of the 3′ (acceptor) splice site (tag→ttg) of intron 47 (fig. 5). The mutation converted the termination codon (tag), which is part of the acceptor site, to a codon for leucine (ttg). The steady-state amount of the in-frame read-through transcript was not diminished, and the normal:intron-inclusion product ratio was ~1:1 in the PCR-amplified cDNA (fig. 3).

Analysis of Type III Collagen Production by Patients' Cultured Fibroblasts

Radiolabeled procollagens from the medium and the fibroblast cell layer were analyzed by SDS-PAGE—first, after addition of a reducing agent and, second, under nonreducing conditions after partial digestion with pepsin to remove the precursor-specific peptides at both ends of the triple-helical domain. The amount of type III procollagen secreted into the medium was diminished for all cell strains, but the electrophoretic mobility of the

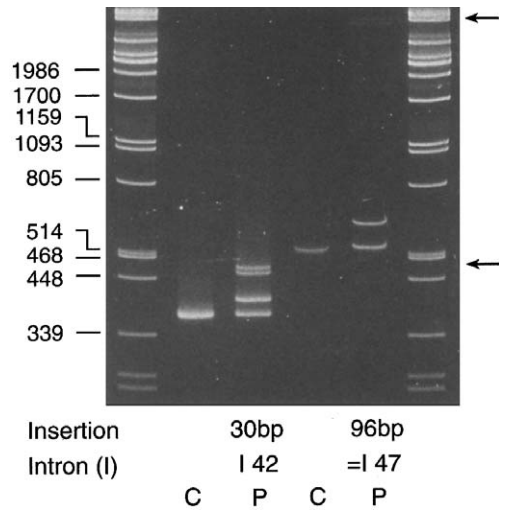

Figure 3 *COL3A1* reverse-transcriptase–PCR products from an unaffected individual (lanes C) and two patients (lanes P), with inclusion, in the mRNA, of either the first 30 bp of intron 42 (patient 94-722) or the entire 96-bp intron 47 (patient 95-229). The arrows on the right indicate the positions of heteroduplexes. Molecular-weight marker = −DNA × *Pst*I.

constituent chains was heterogeneous (fig. 6). In some cell strains, chains migrated slightly faster than normal, resulting in a less-sharp separation between the proα1(III) and proα1(I) chains (fig. 6). There was an increase in the amount of proα1(III) molecules retained intracellularly, with increasing delay of the electrophoretic mobility, the further C-terminal the exon deletion occurred (fig. 6).

Discussion

The accuracy of the cleavage and ligation reactions by which introns are removed from precursor mRNA depends on the interaction between consensus sequences located at each end of the introns, small nuclear ribonucleoproteins, and other protein-splicing factors (e.g., U2AF, PTB, ASF/SF2, and SC35 [Singh et al. 1995; Valcárcel et al. 1996; Wang et al. 1996; MacMillan et al. 1997]). When other aspects of the splicing machinery are normal, the efficiency of splicing is influenced not only by the conserved sequence elements at the 5′ (donor) and 3′ (acceptor) sites in the introns but by the length and sequence of introns (Sirand-Pugnet et al. 1995; Cogan et al. 1996), the length and sequence of adjacent exons (Sterner and Berget 1993; Peterson et al. 1994; van Oers et al. 1994; Del Gatto et al. 1996), and RNA secondary structure (Estes et al. 1992).

With the mutations that we and others have now reported, within the consensus acceptor and donor sites of the *COL3A1* gene, point mutations have been identified, that lead to exon skipping in 43 individuals or families (fig. 2*A* and *B*). Of these mutations, all but three are in the 5′ donor sequences. If a target size of 6 bp for the 5′ site (consensus sequence gtaagt) and 3 bp for the 3′ site (consensus sequence c/tag) is assumed (Krawczak et al. 1992), then the expected distribution of this number of mutations would be closer to a 2:1 ratio than is the 13:1 ratio that is observed. In the *COL1A1* and *COL1A2* genes, splice-site mutations that result in exon skipping in forms of osteogenesis imperfecta and EDS type VII are distributed in a pattern that more closely approximates the expected ratio (Kuivaniemi et al. 1997; Byers et al. 1997). The reason for the skewed pattern of observed mutations at splice sites in the *COL3A1* gene is not clear, but at least two explanations seem possible.

One explanation is that the phenotypes associated with 3′ splice-site mutations may differ from the usual presentation of EDS type IV, which has been the means by which the individuals studied by us and by others have been ascertained. A second possibility is that the favored result of 3′ splice-site mutations is creation of a null allele but that the ability to discern a heterozygous null by protein assays is limited, so that we might not have studied the cells.

The molecular consequences of 5′ splice-site mutations depend largely on the position of the mutation in the consensus sequence and on the length of the downstream intron. The three major consequences of 5′ splice-site mutations are expected to be (1) skipping of the entire preceding exon, (2) read-through of the intron, with intron inclusion in the mature mRNA, and (3) use of cryptic donor sites either in the preceding exon or downstream in the intron. In the *COL3A1* gene, exon skipping within the triple-helical domain results in an in-frame mRNA, because of the cassette structure of the exons (each starts with an intact glycine codon and ends with an intact Y-position codon). These mRNA species would be expected to have normal stability.

Intron inclusion may be favored if the size of the intron does not yield a redefined exon >300 bp (Roberson et al. 1990). For example, with an intron 20 $G^{+1}\rightarrow A$ transition, intron 20 is retained in a third of the mutant transcripts that are detectable in total RNA at steady state (Kuivaniemi et al. 1990). Most introns in the *COL3A1* gene that have been characterized are >300 bp. Thus, of the remaining mutations that we and others have identified, only the disruption of the 5′ (donor) splice site of intron 34 might have led to inclusion of a the complete downstream 85-nt intron. The product of this alternative would contain a stop codon, and thus the mRNA would most likely be unstable or not transported efficiently from the nucleus (Stover et al. 1993; Maquat 1995, 1996; Redford-Badwal et al. 1996). This alternative does not appear to be used, since the normal:

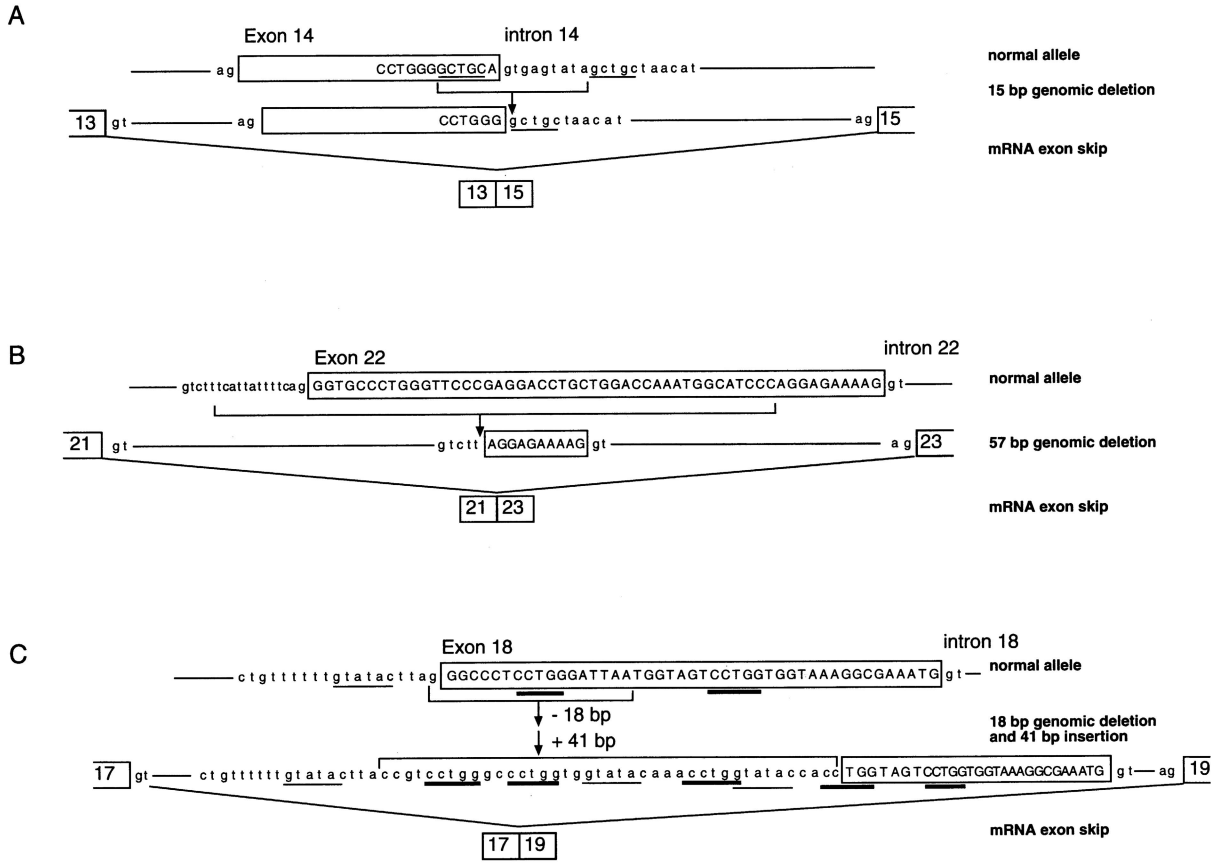

Figure 4 Schematic representation of two small genomic deletions (*A* and *B* [patients 93-104 and 91-336, respectively]) and one complex deletion/insertion event (*C* [patient 91-589]) in *COL3A1,* leading to exon skips. Exons are represented as boxes (not drawn to scale), and intron sequences are signified by lowercase letters; one short direct repeat in *A* and two different short repeats in *C* are underlined.

exon-skipped product ratio is ~1:1 in the PCR-amplified cDNA. All other introns in the *COL3A1* gene in which we and others have identified 5′ splice-site mutations would lead to redefined exons >300 nt, so that skipping would probably be favored.

Of the reported mutations, only three appear to use cryptic splice sites either within the intron or within the preceding exon (Kuivaniemi et al. 1990). The $G^{+1}\rightarrow A$ mutations in these introns lead to alternative splicing in which exon skipping is one of the possible outcomes. With the mutation in the 421-bp intron 16, skipping is the major product, and insertion of 24 nt is a minor product. With the mutation in the smaller, 132-bp intron 20, both a 24-nt insertion and insertion of the whole intron constitute significant products, and exon skipping is a minor outcome. With mutations in the large intron 42 (753 bp), the only splicing product results in read-through to a cryptic donor site 30 nt downstream, an observation that we also made in one unrelated patient (94-722) in our study.

With 3′ splice-site mutations, the several alternatives include exon skipping and the use of alternative acceptor sites within the intron or the succeeding exon (most common). Exon skipping would lead to a stable product, as exemplified by the mutations that we identified, in introns 17 and 46, that led to skipping of exons 18 and 47, respectively. We examined the sequences of exons 6–49, to identify potential splice-acceptor sites (c/tag sequences). When we assumed 12 bp as a minimal exon size (Hawkins 1988; Berget 1995), we found 84 potential sites, of which 76 would result in out-of-frame mRNA sequences that we would expect to be unstable or to be retained within the nucleus (Maquat 1995, 1996; Stover et al. 1993). This suggests that many 3′ splice-site mutations could give rise to "null" alleles of *COL3A1,* because the mRNA product contains a premature-termination codon that is succeeded by at least one intron and is thus subject to nonsense-mediated decay (Carter et al. 1996).

It is clear that cells from all individuals so far identified with the EDS type IV phenotype produce an abnormal type III procollagen protein. We are uncertain of the

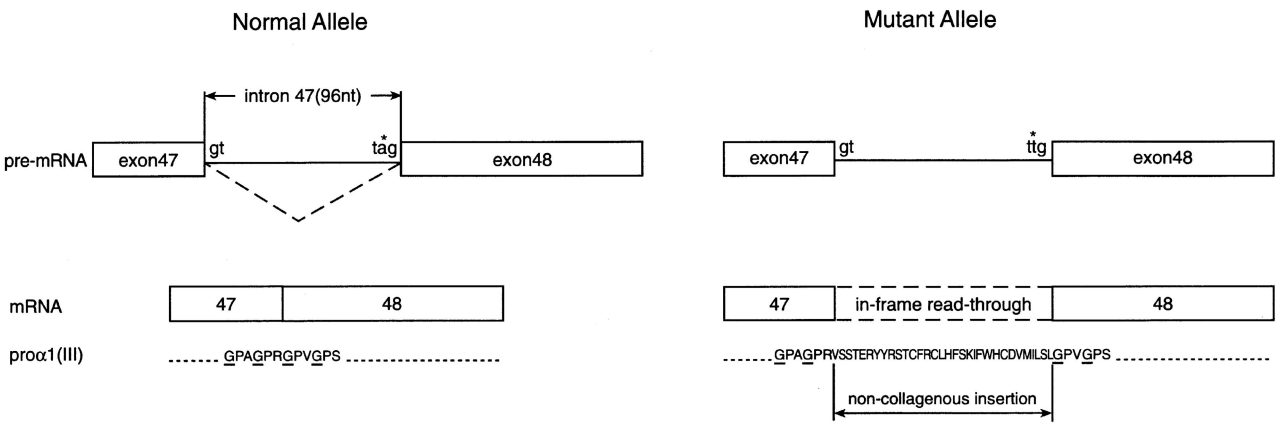

Figure 5 Schematic depiction of a point mutation (IVS47 $A^{-2}\rightarrow T$), at the 3′ (acceptor) splice site of intron 47, that changes a termination codon (TAG) to a codon for leucine (TTG) and results in retention of intron 47 (96 nt) in the mature transcript (patient 95-229).

expected phenotype for *COL3A1* null alleles. Two large studies of individuals with cerebral aneurysms (Kuivaniemi et al. 1993) and abdominal aortic aneurysms (Tromp et al. 1993) make it unlikely that *COL3A1* null alleles are the usual causes of those conditions, inasmuch as the majority of the people studied had evidence of expression of both *COL3A1* alleles. Heterozygous mice with a *COL3A1* null allele generated by targeted gene inactivation were phenotypically normal (Liu et al. 1997). However, late onset of signs might be missed by a short follow-up period. Null alleles of the *COL1A1* gene result in the mildest form of osteogenesis imperfecta, and in these individuals the mutations have been shown to be (i) point mutations that alter 5′ (donor) splice sites and lead to intron inclusion or the use of cryptic sites that are out of frame and encode downstream termination codons; (ii) small deletions or insertions within exons, which cause translational frameshifts and new termination codons downstream; and (iii) point mutations that create a premature-termination codon (Stover et al. 1993; Willing et al. 1994, 1996). Given this, milder or late-onset vascular involvement may yet be a candidate phenotype for the *COL3A1* null-allele genotype.

In the present study we have shown that a point mutation in the 3′ splice site can result in the inclusion of the complete upstream intron. This appears to be a rare event with human mutations. An A→T transversion at the −2 position of the 3′ splice site of intron 47 (in patient 95-229) results in retention, in the mature transcript, of the 96-nt intron 47, in which an in-frame termination codon embedded in the site is changed to a leucine codon. It is the only splice outcome (fig. 5). Since the length of the retained intron is a multiple of three,

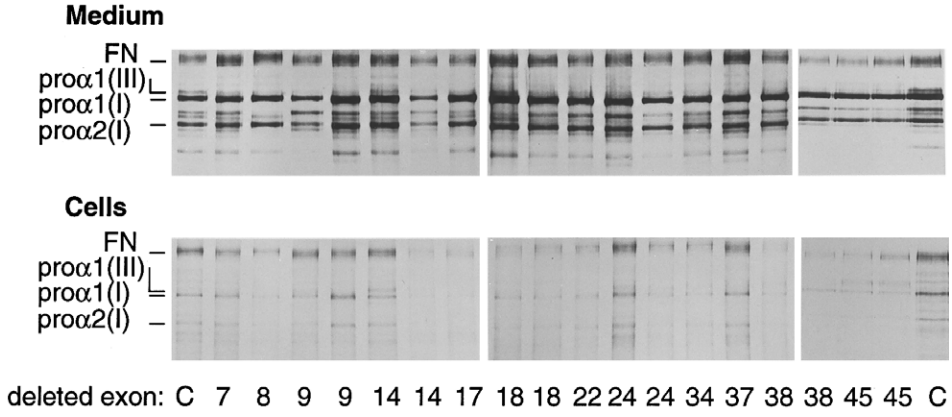

Figure 6 SDS-PAGE of radiolabeled procollagens from medium (*top*) and fibroblast cell layer (*bottom*), analyzed under reducing conditions. The first and the last lane show a control cell line (C), and the lanes in between them show cell lines of 18 patients with EDS type IV and are arranged according to the relative position of the exon deletion along the *COL3A1* gene, with the N-terminus leftmost. FN = fibronectin.

and since the mutation converts the only termination codon (also part of the 3′ splice site) in that sequence to a codon for leucine, the reading frame remains uninterrupted and thus is probably normally translated. The repeating amino acid sequence $(Gly\text{-}X\text{-}Y)_n$ of the collagen triple helix is abolished in the area of the insertion, so this chain would severely disrupt the helical structure and result in a nonfunctional protein. Furthermore, there would be none of the fully helical shorter molecules that are seen in exon-skipping mutations.

The other 3′ splice-site mutation that we identified, IVS17G^{-1}→C, and the previously described IVS13A^{-2}→C (Pope et al. 1996) both convert the nonsense "codon" TAG of the acceptor sequence to sense "codons." In both instances the intron of which they are part contains stop codons, and intron 17 is large, so that the opportunity for read-through is absent.

In three individuals, elimination of a splice junction, by small genomic deletions or a complex deletion/insertion event (fig. 4), explained a single exon deletion. The 15-bp deletion that removed the 5′ (donor) splice site of intron 14 involved a 5-bp flanking direct repeat, consistent with deletion by "slipped mispairing" during DNA replication (Efstratiadis et al. 1980). The mechanism for a 57-bp deletion by which the 3′ (acceptor) splice site of intron 21 was removed is not clear. There is no flanking direct repeat or inverted repeat that could allow the formation of a stem-loop structure during replication and lead to excision by DNA repair enzymes. In a third individual with an exon 18 skip, the genomic defect was an 18-bp deletion/41-bp insertion that obliterated the 3′ (acceptor) splice site of intron 17. The iterative occurrence of short sequence stretches in the inserted sequence that were originally located within and adjacent to the deleted 18 bp is consistent with "stutter" of the DNA polymerase.

Type III collagen is a component of the extracellular matrix of many tissues and is particularly abundant in skin, blood vessels, and hollow organs, which bear the brunt of mutations in the *COL3A1* gene. Ultrastructural analyses of various tissues derived from *Col3a1*$^{-/-}$ mutant mice, which exhibit features closely resembling the phenotype of EDS type IV in humans, showed both a marked reduction of the number of collagen fibrils and a significant increase in the mean diameter of highly variable and disorganized type I collagen fibrils, which implies a critical role for type III collagen in determining the size and structure of these fibrils (Liu et al. 1997). In the presence of mutations, the size of fibrils changes to reflect the location of the mutation along the chain (Smith et al. 1997). Because collagen is a linear, ropelike molecule, it might be expected that mutations along the gene (with consequences along the protein) might have different phenotypic outcomes. On the basis of the data that we have collected on these families, we were unable to determine a precise phenotype/genotype correlation. That is, bowel rupture and arterial rupture appeared to be as common among families in which the mutations were in the 3′ end of the coding region as they were among those families with mutations in the 5′ end of the region encoding the triple helix. The most likely explanation for our failure to find correlation of mutation position and phenotype is that the criteria for inclusion in the study group require that most of the phenotypic consequences be met in one or more affected individuals. Thus the range of phenotypic variation is likely to be small and may be limited to minor features, such as acrogeria. An alternative explanation is that only the normal homotrimers and abnormal homotrimers (in which all three chains have deleted an exon) are efficiently secreted, so that the major matrix molecule population with all exon-skipping mutations is similar.

Acknowledgments

We thank the many families and clinicians who provided the cell strains and clinical data, Laura Suesserman for maintaining the cells, and Barbara Kovacich for typing the manuscript.

References

Ala-Kokko L, Kontusaari S, Baldwin CT, Kuivaniemi H, Prockop DJ (1989) Structure of cDNA clones coding for the entire preproα1(III) chain of human type III procollagen. Biochem J 260:509–516

Anderson DW, Thakker-Varia S, Stolle CA (1993) A G^{+1}→A change in IVS20 of the type III procollagen gene leads to cryptic splicing of the mRNA in an EDS IV patient. Matrix 13:33

Benson-Chanda V, Su M-W, Weil D, Chu M-L, Ramirez F (1989) Cloning and analysis of the 5′ portion of the human type-III procollagen gene (*COL3A1*). Gene 78:255–265

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, Duvic M, Atkinson M, Robinow M, Smith LT, Krane SM, Greally MT, et al (1997) Ehlers-Danlos syndrome type VIIA and VIIB result from splice junction mutations or genomic deletions that involve exon 6 in the *COL1A1* and *COL1A2* genes of type I collagen. Am J Med Genet 72: 94–105

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

Chiodo AA, Sillence DO, Cole WG, Bateman JF (1995) Abnormal type III collagen produced by an exon-17-skipping mutation of the *COL3A1* gene in Ehlers-Danlos syndrome type IV is not incorporated into the extracellular matrix. Biochem J 311:939–943

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

Chu M-L, Prockop DJ (1993) Collagen: gene structure. In: Royce PM, Steinmann B (eds) Connective tissue and its heritable disorders. Wiley-Liss, New York, pp 149–165

Cogan JD, McCarthy EMS, Prince MA, Lekhakula S, Futrakul A, Bundey S, Phillips JA III (1996) A novel mechanism of aberrant splicing. Am J Hum Genet Suppl 59:A10

Cole WG, Chiodo AA, Lamande SR, Janeczko R, Ramirez F, Dahl H-HM, Chan D, et al (1990) A base substitution at a splice site in the *COL3A1* gene causes exon skipping and generates abnormal type III procollagen in a patient with Ehlers-Danlos syndrome type IV. J Biol Chem 265: 17070–17077

Del Gatto F, Gesnel M-C, Breathnach R (1996) The exon sequence TAGG can inhibit splicing. Nucleic Acids Res 24: 2017–2021

Efstratiadis A, Posakony JW, Maniatis T, Lawn RM, O'Connell C, Spritz RA, DeRiel JK, et al (1980) The structure and evolution of the human β-globin gene family. Cell 21:653–668

Estes PA, Cooke NE, Liebhaber SA (1992) A native RNA secondary structure controls alternative splice-site selection and generates two human growth hormone isoforms. J Biol Chem 267:14902–14908

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

Kinnane J, Priebe C, Caty M, Kuppermann N (1995) Perforation of the colon in an adolescent girl. Pediatr Emerg Care 11:230–232

Kontusaari S, Tromp G, Kuivaniemi H, Ladda RL, Prockop DJ (1990) Inheritance of an RNA splicing mutation (G^{+1} ^{IVS20}) in the type III procollagen gene (COL3A1) in a family having aortic aneurysms and easy bruisability: phenotypic overlap between familial arterial aneurysms and Ehlers-Danlos syndrome type IV. Am J Hum Genet 47:112–120

Krawczak M, Reiss J, Cooper DN (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet 90:41–54

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, Prockop DJ, Wu Y, Madhatheri SL, Kleinert C, Earley JJ, Jokinen A, et al (1993) Exclusion of mutations in the gene for type III collagen (*COL3A1*) as a common cause of intracranial aneurysms or cervical artery dissections: results form sequence analysis of the coding sequences of type III collagen from 55 unrelated patients. Neurology 43:2652–2658

Kuivaniemi H, Tromp G, Bergfeld WF, Kay M, Helm TN (1995) Ehlers–Danlos syndrome type IV: a single base substitution of the last nucleotide of exon 34 in *COL3A1* leads to exon skipping. J Invest Dermatol 105:352–356

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

Lee B, Vitale E, Superti-Furga A, Steinmann B, Ramirez (1991) G to T transversion at position +5 of a splice donor site causes skipping of the preceding exon in the type III procollagen transcripts of a patient with Ehlers-Danlos syndrome type IV. J Biol Chem 266:5256–5259

Liu X, Wu H, Byrne M, Krane S, Jaenisch R (1997) Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development. Proc Natl Acad Sci USA 94:1852–1856

Lloyd J, Narcisi P, Richards A, Pope FM (1993) A T^{+6} to C^{+6} mutation in the donor splice site of *COL3A1* IVS 7 causes exon skipping and results in Ehlers-Danlos syndrome type IV. J Med Genet 30:376–380

MacMillan AM, McCaw PS, Crispino JD, Sharp PA (1997) SC35-mediated reconstitution of splicing in U2AF-depleted nuclear extract. Proc Natl Acad Sci USA 94:133–136

Maquat LE (1995) When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. RNA 1:453–465

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

Padgett RA, Grabowski PJ, Konarska MM, Seiler S, Sharp PA (1986) Splicing of messenger RNA precursors. Annu Rev Biochem 55:1119–1150

Peterson ML, Bryman MB, Peiter M, Cowan C (1994) Exon size affects competition between splicing and cleavage-polyadenylation in the immunoglobulin gene. Mol Cell Biol 14: 77–86

Pope FM, Narcisi PN, Nicholls AC, Germaine D, Pals G, Richards AJ (1996) *COL3A1* mutations cause variable clinical phenotypes including acrogeria and vascular rupture. Br J Dermatol 135:163–181

Redford-Badwal DA, Stover ML, Valli M, McKinstry MB, Rowe DW (1996) Nuclear retention of *COL1A1* messenger RNA identifies null alleles causing mild osteogenesis imperfecta. J Clin Invest 97:1035–1040

Richards AJ, Narcisi P, Ferguson C, Cobben JM, Pope FM (1994) Two new mutations affecting the donor splice site of *COL3A1* IVS37 and causing skipping of exon 37 in patients with Ehlers-Danlos syndrome type IV. Hum Mol Genet 3:1901–1902

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

Singh R, Valcárcel J, Green MR (1995) Distinct binding spec-

ificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. Science 268:1173–1176

Sirand-Pugnet P, Durosay P, Brody E, Marie J (1995) An intronic (A/U)GGG repeat enhances the splicing of an alternative intron of the chicken β-tropomyosin pre-mRNA. Nucleic Acids Res 23:3501–3507

Smith LT, Schwarze U, Goldstein J, Byers PH (1997) Mutations in the *COL3A1* gene result in the Ehlers-Danlos syndrome type IV and alterations in the size and distribution of the major collagen fibrils of the dermis. J Invest Dermatol 108: 241–247

Steinmann B, Royce PM, Superti-Furga A (1993) The Ehlers-Danlos syndrome. In: Royce PM, Steinmann B (eds) Connective tissue and its heritable disorders. Wiley-Liss, New York, pp 351–407

Sterner DA, Berget SM (1993) In vivo recognition of a vertebrate mini-exon as an exon-intron-exon unit. Mol Cell Biol 13:2677–2687

Stover ML, Primorac D, Liu SC, McKinstry MB, Rowe DW (1993) Defective splicing of mRNA from one *COL1A1* allele of type I collagen in non-deforming (type I) osteogenesis imperfecta. J Clin Invest 92:1994–2002

Thakker-Varia S, Anderson DW, Kuivaniemi H, Tromp G, Shin H-G, van der Rest M, Glorieux FH, et al (1995) Aberrant splicing of the type III procollagen mRNA leads to intracellular degradation of the protein in a patient with Ehlers-Danlos type IV. Hum Mutat 6:116–125

Tromp G, Wu Y, Prockop DJ, Madhatheri SL, Kleinert C, Earley JJ, Zhuang J, et al (1993) Sequencing of cDNA from 50 unrelated patients reveals that mutations in the triple-helical domain of type III procollagen are an infrequent cause of aortic aneurysms. J Clin Invest 91:2539–2545

Tsipouras P, Byers PH, Schwartz RC, Chu ML, Weil D, Pepe G, Cassidy SB, et al (1986) Ehlers-Danlos syndrome type IV: cosegregation of the phenotype to a *COL3A1* allele of type III procollagen. Hum Genet 74:41–46

Valcárcel J, Gaur RK, Singh R, Green MR (1996) Interaction of U2AF^{65} RS region with pre-mRNA of branch point and promotion base pairing with U2 snRNA. Science 273: 1706–1709

van Oers CCM, Adema GJ, Zandberg H, Moen TC, Baas PD (1994) Two different sequence elements within exon 4 are necessary for calcitonin-specific splicing of the human calcitonin/calcitonin gene-related peptide I pre-mRNA. Mol Cell Biol 14:951–960

Wang J, Takagaki Y, Manley JL (1996) Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. Genes Dev 10:2588–2599

Weinbaum PJ, Cassidy SB, Campbell WA, Rickles FR, Vintzileos AM, Nochimson DJ, Tsipouras P (1987) Pregnancy management and successful outcome of Ehlers-Danlos syndrome type IV. Am J Perinatol 4:134–137

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

Willing MC, Deschenes SP, Slayton RL, Roberts EJ (1996) Premature chain termination is a unifying mechanism for COL1A1 null alleles in osteogenesis imperfecta type I cell strains. Am J Hum Genet 59:799–809

Wu Y, Kuivaniemi H, Tromp G, Strobel D, Romanic AM, Prockop DJ (1993) Temperature sensitivity of aberrant RNA splicing with a mutation in the G^{+5} position of intron 37 of the gene for type III procollagen from a patient with Ehlers-Danlos syndrome type IV. Hum Mutat 2:28–36