The EMBO Journal vol.8 no.6 pp.1705 - 1710, 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

Dominique Weil[1,4], Marina D'Alessio[1], Francesco Ramirez[1], Wouter de Wet[2], William G.Cole[3], Danny Chan[3] and John F.Bateman[3]

[1]Department of Microbiology and Immunology, Morse Institute of Molecular Genetics, SUNY Health Science Center at Brooklyn, 450 Clarkson Avenue, Box 44, Brooklyn, NY 11203, USA, [2]Department of Biochemistry, Potchefstroom University, Potchefstroom 2520, South Africa and [3]Department of Pediatrics, University of Melbourne, Parkville, Victoria 3052, Australia
[4]Present address: Hôpital des Enfants Malades, Unité 12 INSERM, Paris 75015, France

Communicated by J.-L.Mandel

An unusual splicing mutation has been characterized in the pro α1(I) collagen gene of a sporadic case of Ehlers-Danlos Syndrome Type VII. Cloning of primer extended cDNA in conjunction with R-looping experiments established that nearly half of the pro α1(I) collagen gene transcripts are abnormally spliced, for they lack exon 6 sequences. Analysis of cloned genomic fragments revealed that one of the proband's alleles displays the substitution of an A for a G in the last nucleotide of exon 6. The change converts the normal Met (ATG) codon to Ile (ATA) and, in addition, obliterates a *Nco*I restriction site. The latter event was exploited to demonstrate the *de novo* nature of the mutation since DNA from the unaffected parents was fully digested with the enzyme, after *in vitro* amplification by the polymerase chain reaction. Further confirmation of the missplicing was obtained by transient expression into animal cells of allelic minigene constructs. Finally, Western blot analysis of cyanogen bromide cleaved collagen and nucleotide sequencing of appropriately selected cDNA clones demonstrated the production of relatively low amounts of correctly spliced molecules harboring the Ile substitution, as well.

Key words: alternative splicing/collagenopathies/RNA splicing/splicing signals specificity/type I collagen

Introduction

Type I collagen, a major extracellular component of several connective tissues, is composed of two α1(I) chains and one α2(I) chain that are initially synthesized as procollagen precursors. Structurally, the procollagen chains consist of a central domain made of several Gly-X-Y repeated tripeptides flanked at both ends by short, globular projections that are enzymatically removed after secretion of the assembled trimer (Miller and Gay, 1987). The genes encoding the subunits of type I collagen are relatively large and highly complex, for they range from 18 to 38 kb and contain >50 exons (Chu *et al.*, 1984; de Wet *et al.*, 1987).

Type I collagen mutations result in two clinically distinct entities: osteogenesis imperfecta, whose hallmark is bone fragility, and the Type VII form of Ehlers-Danlos syndrome (EDS VII), a rare, dominant condition characterized by joint dislocations and hypermobility (Tsipouras and Ramirez, 1987). Biochemical studies have established that cultured EDS VII fibroblasts accumulate partially cleaved molecules (pN-collagens) in the medium, as a result of structural mutations in either pro α1(I), EDS VIIA or pro α2(I) chains, EDS VIIB (Steinmann *et al.*, 1980; Eyre *et al.*, 1985; Cole *et al.*, 1986; Wirtz *et al.*, 1987). To be precise, in three EDS VII patients heterozygous deletions of 24 or 18 amino acids in the pro α1(I) or pro α2(I) chains respectively have been demonstrated (Eyre *et al.*, 1985; Cole *et al.*, 1986; Wirtz *et al.*, 1987). The deleted peptides, which include the N-proteinase cleavage sites of the procollagen chains, correspond precisely to the sequences encoded by the sixth exon of the cognate genes, COL1A1 and COL1A2 (Chu *et al.*, 1984; de Wet *et al.*, 1987). The deletion in one of the patients was recently shown to be caused by skipping of exon 6 sequences because of a T-to-C transversion at the obligatory GT dinucleotide of the 5′ splice site of intron 6 (Weil *et al.*, 1988).

We have now examined another of the aforementioned EDS VII patients, shown previously to produce shortened pro α1(I) chains (Cole *et al.*, 1986), and found that the deletion is similarly due to outsplicing of exon 6 sequences. In contrast to the previous case, however, two distinct features characterize the abnormality of this patient. Firstly, exon skipping is caused by a substitution of an A for a G in the last nucleotide of exon 6. Secondly, although the change inactivates splice-site recognition in the majority of the transcripts, low amounts of correctly spliced molecules are also produced. The identification of this unusual mutation, therefore, re-emphasizes the importance of the whole exon-intron junction sequence in conferring specificity to the selection of individual splicing signals.

Results

Cultured dermal fibroblasts from the EDS VIIA patient were previously shown to synthesize normal and shortened pro α1(I) collagen chains in the approximate ratio of 1.8-1.4 (Cole *et al.*, 1986). Peptide sequencing revealed that the deleted segment extended from Asn 136 in the C-terminal globular domain of the N-propeptide to Met 159 in the first Gly-X-Y triplet of the helical domain (Cole *et al.*, 1986), a region encoded by the sixth exon of COL1A1 (Chu *et al.*, 1984).

In order to confirm this observation at the mRNA level two alternative strategies were employed, namely R-looping analysis and cloning of primer extended cDNA. The former, aside from providing visual evidence of an aberrant exon-intron arrangement, enabled us to approximate the relative number of normal and abnormal heteroduplexes formed between the patient RNA and an appropriate subclone of COL1A1. Specifically, while in control DNA-RNA

hybrids the sequences of exons 6, 7, 8 and 9 were found to be interrupted by three small introns, nearly half of the proband's heteroduplexes displayed a distinctly abnormal R-looping pattern. Two rather than three small introns were visualized and, in addition, a larger intron, 907 ± 59 nucleotides in length, was seen between exons 5 and 7 (Table I). Within the limitations of this method, the size of the displaced DNA strand closely resembled that of the combined lengths of intron 5, exon 6 and intron 6 (Table I). Moreover, the relative abundance of abnormal molecules agreed with the biochemical estimate of shortened pro α1(I) chains (Cole *et al.*, 1986). This visual observation was independently supported by cloning of primer extended cDNA. Accordingly, two recombinants, each representative of either a positive or a negative pattern of hybridization to an exon 6 probe, were selected for sequence analysis. Comparison of the sequences disclosed the precise loss of the 72 bp of exon 6 in the negatively hybridizing cDNA (Figure 1). Collectively, these experiments suggested that the heterozygous mutation was the result of either a genomic deletion or an aberrant splicing.

To discriminate between these two possibilities, Southern blot hybridizations were performed on proband and control DNA cleaved with *Xba*I, *Pvu*II and *Nco*I. The three enzymes were chosen for their ability to detect structural changes in exon 6 (*Pvu*II) and at the junction sequences of intron 5/exon 6 (*Xba*I) and exon 6/intron 6 (*Nco*I) (D'Alessio *et al.*, 1988) (Figure 2). The results excluded a total deletion of exon 6 sequences, for they showed normal hybridization patterns with the first two enzymes (data not shown). Digestion with *Nco*I, on the other hand, did not provide conclusive information because this methylase-sensitive enzyme was found to cut fibroblast DNA very poorly, even when used in large excess. However, since cloned DNA was fully digested by *Nco*I, we analyzed in this manner several clones isolated from a size-selected *Hin*dIII library prepared from the proband DNA. This led to the identification of two sets of recombinants each displaying either the presence or the absence of the *Nco*I restriction site at the exon 6/intron 6 junction. Sequencing of two genomic clones disclosed that the heterozygous loss of the *Nco*I restriction site was caused by a single base substitution, A for G, in the last nucleotide of exon 6 (Figure 2). Incidentally, no additional changes were noted when the sequences comprised between introns 4 and 7 were compared in the two COL1A1 alleles (data not shown).

Table I. Size of exons and introns around the region of mismatch in DNA–RNA hybrids formed between the COL1A1 subclone and pro-α1(I) mRNA from control and EDS VIIA fibroblasts

Exon/intron	Size of exons/introns in R-loops formed with	
	Control RNA	EDS VIIA RNA
Exon 5	141 ± 35	128 ± 24
Intron 5	712 ± 83	907 ± 59
Exon 6	93 ± 14	–
Intron 6	189 ± 37	–
Exon 7	63 ± 22	60 ± 11
Intron 7	169 ± 45	163 ± 36

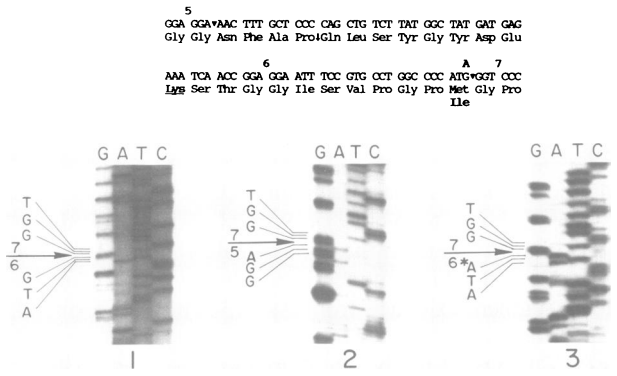

Fig. 1. Nucleotide sequences of the EDS VIIA cDNAs. The nucleotide sequence of exon 6 and part of exons 5 and 7 are shown at the top with the single base substitution (A for G) that results in the conversion of the Met codon of exon 6 to Ile indicated. The arrow signifies the N-proteinase cleavage site, while the crosslinking lysine residue is underlined. Exon boundaries are identified by the symbol ▼ in the sequence and by the arrows in the three panels. **Panel 1**, sequence of the normal transcript; **panel 2**, sequence of the outspliced transcript; **panel 3**, sequence of the correctly spliced product harboring the G to A transversion (asterisk). Only the nucleotides at the transitions between the various exons (numbers) are indicated.

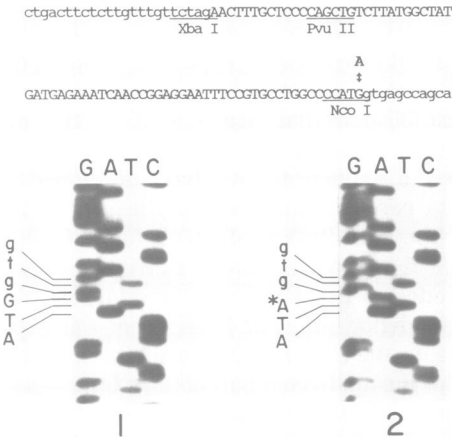

Fig. 2. Nucleotide sequences of the patient's *Nco*I (+) and *Nco*I (–) alleles. **Top**: nucleotide sequence of the 3′ end of intron 5 (lower-case letters), exon 6 (capital letters) and intron 6 (lower-case letters). The recognition sequences of the restriction enzymes *Xho*I, *Pvu*II and *Nco*I are indicated. The G-to-A substitution in the *Nco*I (–) allele of the EDS VIIA patient is shown above the sequence. **Bottom: panel 1**, *Nco*I (+) allele; **panel 2**, *Nco*I (–) allele with the asterisk highlighting the mutation.

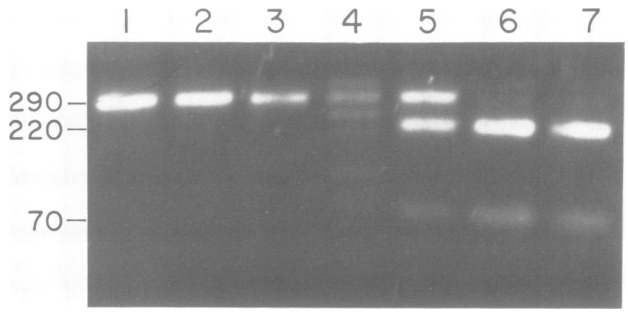

Fig. 3. Genotype analysis of the *Nco*I restriction pattern. Sizes of the amplified products are indicated on the left. **Lanes 1–3**, undigested PCR products; **lane 4,** φX174 DNA *Hae*III-digested markers; **lanes 5–7**, *Nco*I digestion of the PCR products from proband (5), mother (6) and father (7) genomic DNA.

Having established the *Nco*I heterozygosity in the proband, we tested the DNA of the unaffected parents in order to ascertain whether or not the G to A change was merely a neutral polymorphism. Hence, exon 6 sequences were specifically amplified from the DNA of each of the three family members by the polymerase chain reaction (PCR) (Saiki *et al.*, 1985) and digested to completion with *Nco*I. While both parental 290 bp PCR products were cleaved by the enzyme to give products of 220 and 70 bp, the *Nco*I heterozygosity of the proband DNA proved the *de novo* nature of the substitution (Figure 3).

To verify that the transcripts from the *Nco*I (−) allele were indeed misspliced, genomic fragments from both alleles were subcloned into an appropriate vector and transiently expressed in COS cells (Gluzman, 1981). Northern blot hybridization showed that RNA extracted from cells transfected with the *Nco*I (−) allele was appreciably shorter than that from cells harboring the *Nco*I (+) construct and, more importantly, that it did not hybridize to an exon 6 probe (Figure 4). Although the experiments demonstrated conclusively that in this patient disruption of normal splicing and consequent joining of exons 5 and 7 sequences was caused by the G-to-A transversion at the position immediately preceding the 5′ GT dinucleotide of intron 6, they did not rigorously exclude the presence of additional, minor transcripts.

To test this hypothesis, we exploited a fortuitous result of the mutation, notably the conversion of the normal Met (ATG) codon to Ile (ATA). We reasoned that because of this structural change, even if correct splicing of the affected allele seldom occurred, the resulting product should be identifiable by cyanogen bromide (CNBr) peptide mapping. Hence, procollagen from culture medium was purified and separated by DEAE−cellulose chromatography into early and late eluting fractions (Figure 5A). The samples were then digested with CNBr, run in a polyacrylamide gel along with a normal control, Western blotted and probed with antibody specific to the N-terminal propeptide of the pNα1(I) chain (Figure 5B). The reacting peptide common to the

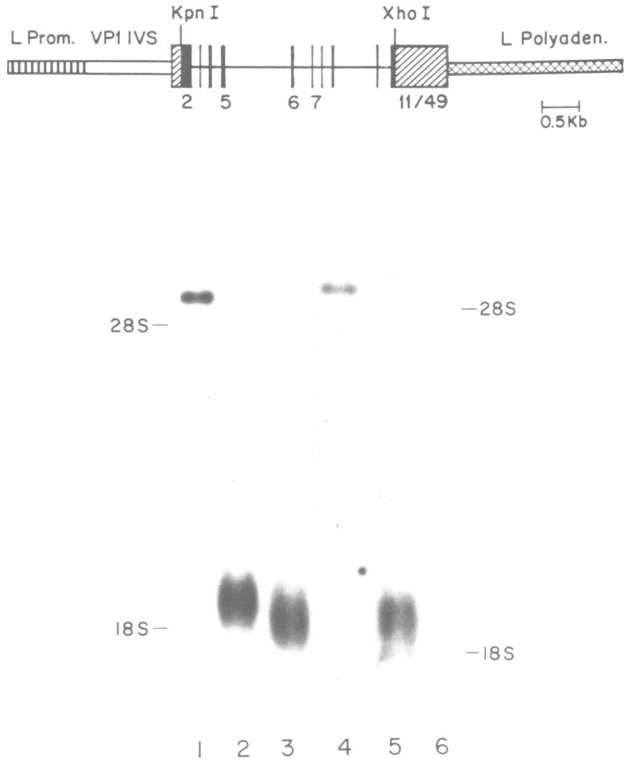

Fig. 4. Northern blot hybridization analysis of the transfected collagen minigene products. A schematic representation of the minigene construct is depicted at the top of the figure. As discussed in Materials and methods, the COL1A1 insert consists of cDNA (cross-hatched boxes) and genomic sequences in which the exons, represented by the black bars, are numbered underneath. The scale on the right side refers only to the eukaryotic insert and not to the SV40 elements depicted at both ends of the construct. VP1 IVS, intervening sequence of viral protein 1. Polyadenylated RNA was extracted from control fibroblasts (**lanes 1** and **4**), COS cells transfected with *Nco*I (+) minigene (**lanes 2** and **5**) and *Nco*I (−) minigene (**lanes 3** and **6**) constructs. Hybridization probes are: proα1(I) cDNA (**lanes 1−3**) and the exon 6 19mer (**lanes 4−6**). The relative migration of the rRNA markers is indicated on the right and left sides of the two Northern blots.

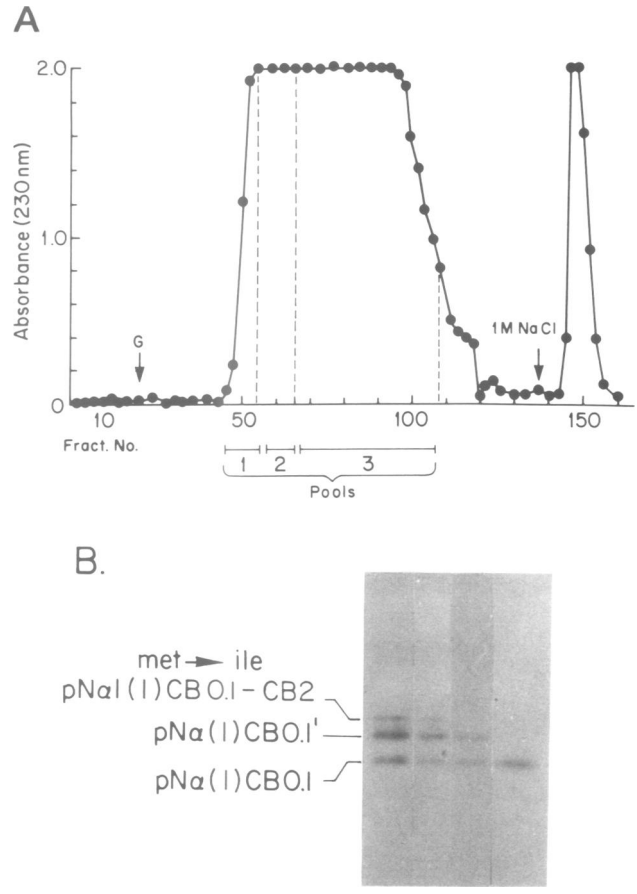

Fig. 5. Analysis of collagen cyanogen bromide peptides. Procollagen was precipitated from cultured EDS VIIA fibroblast culture medium and separated from proteoglycans by DEAE−cellulose chromatography (**panel A**). The letter G indicates the start of the gradient. The large peak containing the procollagen was divided into three pools (indicated by the numbers), which were subsequently digested with CNBr. The peptides (0.5−2 μg) were loaded onto 12.5% polyacrylamide gels containing SDS and subjected to electrophoresis. All lanes are Western blots using antiserum specific for the pNα1(I) Col1 fragment of the pNα1(I)CB O,1 peptide (**panel B**). From left to right: CNBr peptides of procollagens from pool 2, CNBr peptides of procollagens from pool 2, CNBr peptides of procollagens from pool 3, and CNBr peptides from normal procollagen. The normal CNBr peptide is designated pNα1(I)CB O.1; the mutant CNBr peptide containing the 24 amino acid peptide deletion is designated pNα1(I)CB O.1′; the mutant CNBr peptide containing the substitution of Met by Ile is designated pNα1(I)CB O.1-CB2.

control and all of the mutant samples represents the normal pNα1(I) CB 0,1 peptide. The peptide found exclusively in all the fractions of the mutant cells was previously shown to include residues 1−135 of the pNα1(I) CB 0,1 peptides juxtaposed to residues 160−195 of the adjoining CB 2 peptide with deletion of the intervening 24 amino acids, including Met 159 (Cole *et al.*, 1986). A third peptide, seen only in the early eluting fraction of the DEAE−cellulose chromatography, was most likely, for the following reasons, to be the product of the correctly spliced transcripts of the *Nco*I (−) allele. Firstly, the migration pattern of this peptide was consistent with a molecule comprising normal sized pNα1(I) CB 0,1 and CB 2 peptides linked together because of a substitution of Met 159 by Ile (D'Alessio *et al.*, 1988). Secondly, the low representation of this variant chain, $<10\%$, was in agreement with the biochemical and electron-microscopic estimates of correctly and misspliced products.

Based on these data and in order to obtain additional evidence in support of alternative splicing, we proceeded to isolate the Ile coding cDNA. Since the mutation eliminates the *Nco*I restriction site in the affected gene and in the cDNA synthesized from its correctly spliced transcript as well, the relative amount of variant molecules was enriched by intercalating a *Nco*I digestion between two rounds of cDNA amplification by the PCR. The resulting products were then subcloned and screened with an exon 6 probe. Digestion of 20 positive clones with *Nco*I identified three negative cDNAs which, upon sequence inspection, were found to encode the variant Ile molecule (Figure 1).

Discussion

In this report we have described the characterization of the molecular defect in a sporadic case of EDS VIIA. This study has provided novel and important information pertaining to both collagen gene mutations and aberrant splicing in complex, multi-exon genes.

It is now well established that in higher eukaryotes accurate splicing of nuclear pre-mRNA is chiefly determined by the 5′ and 3′ splice sites in which the GT and AG dinucleotides respectively are absolutely conserved in nearly all introns (for reviews see Padgett *et al.*, 1986; Krainer and Maniatis, 1988). Although sequences around these invariable elements share considerable homologies, individual splice sites exhibit sufficient variability to leave unresolved the issue pertaining to the specificity of splice-site selection (Mount, 1982). Relevant to this, early studies supported the idea that exon sequences may not play a major role in RNA splicing (Padgett *et al.*, 1986; Krainer and Maniatis, 1988). However, such a notion has been recently confuted by more accurate investigations of sequence requirements on artificially generated RNA substrates that contain various combinations and manipulations of *cis*-competing splice sites (Reed and Maniatis, 1986). Two important parameters were found to regulate the specificity of the splicing reaction *in vitro*, namely the nature of the junction sequence context and the relative proximity of the splice sites. A model was therefore proposed whereby the splicing machinery scans the RNA substrate for adjacent splice sites and when two equally efficient signals are found to compete in *cis*, the internal one is selected, provided that its sequence context (including exon sequence) is not altered by structural changes (Reed and Maniatis, 1986). This model, in turn, explains the avoidance of exon-skipping during processing by complex transcripts, as well as the unobserved activation of weaker, cryptic splice sites in splicing mutations of multi-exon genes (Mitchell *et al.*, 1986; Hidaka *et al.*, 1987; Marvit *et al.*, 1987; Arpaia *et al.*, 1988; Ruffner and Dugaiczyk, 1988; Tromp and Prockop, 1988; Weil *et al.*, 1988).

The validity of this postulate is clearly substantiated by our findings in this naturally occurring collagen variant. Accordingly, alternative outsplicing of exon 6 sequences can be explained by a substantial but not complete reduction in the efficiency of the mutated exon 6/intron 6 splicing signal to compete in *cis* with the proximal exon 5/intron 5 splice site. Plausibly, the weakening of the signal results from the elimination by the G-to-A transversion of an important G:C base pairing between the collagen splice site and the 5′ end of the U1 snRNA (Rogers and Wall, 1980). This is the first description of alternative splicing caused by a single base substitution at the position immediately preceding the highly conserved GT dinucleotide. It is worth noting that an identical substitution in the sixth exon of COL1A2 also produces alternative splicing, albeit with a different relative ratio of the two transcripts (D.Weil *et al.*, manuscript in preparation). In contrast, in the previously described EDS VIIB variant with the mutated GT dinucleotide (Weil *et al.*, 1988), only exon-skipping occurs probably because this highly conserved element may have additional interacting functions during the splicing reaction (Aebi *et al.*, 1986). Taken together the data suggest that, although some of the intronic elements are absolutely indispensable for completing pre-mRNA processing, both coding and intervening sequences are necessary for determining the specificity of splice-site selection.

Another important set of considerations relates to the biosynthetic and phenotypic consequences of the mutation. The missplicing of exon 6 sequences does not in fact change the translation frame of the mRNA nor the ability of the resulting polypeptides to assemble into structurally abnormal trimers, ultimately responsible for the dominant phenotype. Moreover, the deletion, in addition to the N-proteinase cleavage site, eliminates one of the lysines involved in collagen crosslinking (Eyre *et al.*, 1984), a loss which may contribute to the pathogenesis of the syndrome. It is unknown whether the normal size, variant procollagen chains produced by the affected allele might also participate in the genesis of the phenotype, for example by disrupting the secondary structure of the N-propeptide.

In conclusion, this and published and unpublished data point to the existence of an interesting subgroup of EDS VII patients characterized by exon 6 outsplicing which in each of the cases is the result of a distinct mutation (Weil *et al.*, 1988; D.Weil *et al.*, manuscript submitted; F.Ramirez, unpublished data). It will therefore be of interest to examine additional EDS VII variants in order to appreciate the full spectrum of defects leading exon 6 to missplicing and/or assess whether or not this rare condition can also result from structural mutations.

Materials and methods

Cloning experiments

Collagen-specific cDNA cloning was accomplished using the synthetic primer 5′-CTTGGAAGCCTTGGGGACC-3′ which is complementary to the mRNA sequences encoded by the 5′ portion of exon 8 of COL1A1 (D'Alessio *et al.*,

1988). Conditions for first and second strand synthesis, as well as insertion of the resulting molecules into the *Eco*RI cloning site of the λgt10 vector were essentially as described (Davis *et al.*, 1986). Phage plaques were screened in parallel using the end-labeled exon 8 oligomer and an exon 6 19mer (5'-GGCACGGAAATTCCTCCGG-3') according to the previously detailed conditions (Weil *et al.*, 1988).

For genomic cloning, DNA purified form the proband's fibroblasts was cleaved to completion with *Hin*dIII and fractionated on a 10–40% sucrose gradient. DNA from fractions comprising 8- to 10-kb fragments were pooled and cloned into the *Hin*dIII-digested arms of the phage vector Charon 21 (Chu *et al.*, 1985). Recombinants were screened using a nick-translated genomic probe according to the standard protocol (Maniatis *et al.*, 1982).

R-loop hybridization and electron microscopy

Polyadenylated RNA was purified from proband and control fibroblasts and hybridized to RMS 8/7.6 E:H, a 7.6 kb *Eco*RI–*Hin*dIII pBR322 subclone of COL1A1 that encompasses the region containing exon 6 (Chu *et al.*, 1984). DNA and RNA, at concentrations of 6 and 20 μg/ml respectively, were hybridized in 50 μl volumes of buffer containing 70% formamide, 0.3 M NaCl and 0.01 M EDTA in 0.1 M Pipes, adjusted to pH 7.2 at 20°C. Hybridization was carried out first at 57°C for 12 h and then for an additional 8 h at 53°C. The hybrids were chilled to 0°C and treated with 1.0 M glyoxal for 2 h at 11°C. To remove free glyoxal, the R-loop mixtures were dialyzed at 4°C against buffer containing 0.5 M NaCl and 1.0 mM EDTA in 10 mM Tris–HCl, adjusted to pH 7.4 at 20°C (de Wet *et al.*, 1986). R-loop samples were spread from 50% formamide onto a water hypophase as described (Davis *et al.*, 1980). Films on collodion-covered grids were stained with uranyl acetate and then rotary shadowed with Pt/Pd. Nicked double-stranded replicative form DNA, as well as single-stranded virion DNA of phage φX174 were included in all spreads for length calibration. R-loop structures were visualized and photographed with a Philips 301 electron microscope and measured at a final magnification of 170 000 with a Numonics digitizer. The data is presented as mean ± SD on at least 20 molecules.

Minigene construction and transfection experiments

The two 9 kb *Hin*dIII genomic alleles and a full-size pro α1(I) cDNA were used as substrates for the construction of the collagen minigenes that were ultimately inserted into the expression vector pSVL (Pharmacia) (Templeton and Eckhart, 1984). Cleavage of the cDNA with *Xho*I and self-ligation of the clone eliminated nearly 3 kb of coding sequences and juxtaposed the *Xho*I restriction sites of exons 11 and 49 (Chu *et al.*, 1984). The deleted cDNA was then digested with *Kpn*I, whose recognition site resides within exon 2, and *Xho*I. Subsequently, within these two sites the *Kpn*I, *Xho*I double-digested subfragment of each of the two 9 kb *Hin*dIII genomic alleles was subcloned. This resulted in the generation of a collagen minigene construct that consists of genomic sequences from exon 2 to 11 inserted between the cDNA sequences that span from the start site of translation to exon 2 and from 49 to the end of the message. Further restriction of the minigene with *Eco*RI excised the collagen sequences from the pUC9 plasmid and, in addition, removed the 3' untranslated region, as well as the last 30 bp of coding sequences (Chu *et al.*, 1984). After elution, the fragment was blunt-ended and subcloned into the *Sma*I restriction site of the expression vector pSVL, which includes the following SV40 elements: late promoter, VP1 intron and late polyadenylation signals (Templeton and Eckhart, 1984).

For each of the transfection experiments five 175-cm^2 plates of COS cells (Gluzman, 1981) were transfected with 40 μg of minigene DNA by the DEAE–dextran method (Sompayrac and Danna, 1981). Polyadenylated RNA was purified ~48 h after transfection and analyzed by Northern blot hybridizations (Maniatis *et al.*, 1982).

PCR amplification and DNA sequencing

The oligonucleotides used for the genomic amplification were 5'-GCTGTCTTATGGCTATGATGA-3' and 5'-CATGGTGATCCCTCT-GTAGGA-3'. The two oligomers are complementary to the sequence 5' to the *Nco*I restriction site of exon 6 and to the 3' most segment of intron 6 respectively (D'Alessio *et al.*, 1988). Two micrograms of genomic DNA were amplified for 25 cycles with 2.5 U of Taq polymerase (Perkin-Elmer) and 200 ng of each primer in a total volume of 100 μl (Saiki *et al.*, 1985). Denaturation was for 1.5 min at 94°C. Annealing was at 52°C for 2.5 min. Extension was at 72°C for 3.5 min. Amplified DNA was detected by staining with ethidium bromide after electrophoresis in a 2% NuSieve (FMC) agarose gel.

The oligonucleotides used for the cDNA amplification were 5'-CGGAATTCCCGAAACCACCGGCGTGGAAGGA-3' and 5'-GGGATC-CCGCCTTGGAAGCTTGGGGACC-3'. The former corresponds to the coding sequence encompassing exons 3 and 4, while the latter is complementary to the coding strand of exon 8 (D'Alessio *et al.*, 1988). The two oligomers contain additional sequences specific for *Eco*RI and *Bam*HI restriction sites respectively. Amplification conditions were as described above except that the annealing temperature was 45°C. Moreover, and as discussed in the text, the cDNA was subjected to two rounds of 25-cycle amplification intercalated by a single *Nco*I digestion. Subcloning into pUC18 was achieved by double digestion with *Bam*HI and *Eco*RI of the PCR products purified by two phenol extractions and ethanol precipitations.

Nucleotide sequencing was performed according to the dideoxy chain termination procedure (Sanger *et al.*, 1977) on double-stranded DNA (Hattori and Sakaki, 1986).

Protein analysis

Procollagen was precipitated from cultured EDS VIIA fibroblasts, grown in Dulbecco's modified Eagle's minimal essential medium containing 10% fetal calf serum, by the addition of ammonium sulfate to a final concentration of 25% saturation. The precipitate was dissolved in 0.05 M Tris–HCl, pH 7.5, 0.15 M NaCl and proteinase inhibitors, precipitated with ethanol (final concentration 18%) redissolved in 0.05 M Tris–HCl, pH 7.5, 2 M urea, 1 mM EDTA and dialyzed against the same buffer for 24 h at 4°C. Procollagen was separated from proteoglycans by DEAE–celluose and appropriate fractions were pooled, dialyzed against 0.1 M acetic acid, lyophylized and digested with CNBr (Smith *et al.*, 1972). Between 0.5 and 2 μg of the CNBr-derived peptides were electrophoresed in 12.5% polyacrylamide gels containing SDS, Western blotted and reacted with an antiserum specific for the pNα(I) Col1 fragment of the pNα1(I) CB O, 1 peptide (Cole *et al.*, 1986). Relative proportions of the three pNα1(I) CB O, 1 peptides in the total DEAE–cellulose procollagen peak were calculated from the concentrations of the protein (absorbance areas) of pools 1, 2 and 3 and from densitometry determinations of the relative amounts of the peptides on the Western blots obtained from these pools.

Acknowledgements

We are greatly indebted to Dr Rudolf Jaenisch for providing us with the full-size proα1(I) cDNA and Dr Jeff Bonadio for many helpful suggestions on the amplification experiments. We thank Ms Sharon Boast and Mr James Prince for their excellent technical assistance and Ms Roseann Lingeza for the preparation of the manuscript. M.D.A. is on leave of absence from International Institute of Genetics and Biophysics, CNR, Naples, Italy. This work was supported by grants from the National Institutes of Health (AR 38648), the March of Dimes, Birth Defects Foundation (1-1042), INSERM, the South African Medical Research Council and the National Health and Medical Research Council of Australia.

References

Aebi,M., Horning,H., Padgett,R.A., Reiser,J. and Weissmann,C. (1986) *Cell*, **47**, 555–565.

Arpaia,E., Dumbrille-Ross,A., Maler,T., Neote,K., Tropak,M., Troxel,C., Stirling,J,L., Pitts,J.S., Bapat,B., Lamhonwah,A.M., Mahuran,D.J., Schuster,S.M., Clarke,J.T.R., Lowden,J.A. and Gravel,R.A. (1988) *Nature*, **333**, 85–86.

Chu,M.L., de Wet,W., Bernard,M., Ding,J.F., Morabito,M., Myers,J., Williams,C. and Ramirez,F. (1984) *Nature*, **310**, 337–340.

Chu,M.L., Gargiulo,V., Williams,C.J. and Ramirez,F. (1985) *J. Biol. Chem.*, **260**, 691–694.

Cole,W.G., Chan,D., Chambers,G.W., Walker,I.D. and Bateman,J.F. (1986) *J. Biol. Chem.*, **261**, 5496–5503.

D'Alessio,M., Bernard,M., Pretorius,P.J., de Wet,W. and Ramirez,F. (1988) *Gene*, **67**, 105–115.

Davis,L.G., Dibner,M.D. and Battey,J.F. (1986) *Basic Methods in Molecular Biology*. Elsevier, New York.

Davis,R.W., Simon,M. and Davidson,N. (1980) *Methods Enzymol.*, **65**, 494–560.

de Wet,W., Sippola,M., Tromp,G., Prockop,D.J., Chu,M.L. and Ramirez,F. (1986) *J. Biol. Chem.*, **261**, 3857–3862.

de Wet,W., Bernard,M., Benson-Chanda,V., Chu,M.L., Dickson,L., Weil,D. and Ramirez,F. (1987) *J. Biol. Chem.*, **262**, 16032–16036.

Eyre,D.R., Paz,M.A. and Gallop,P.M. (1984) *Annu. Rev. Biochem.*, **53**, 717–748.

Eyre,D.R., Shapiro,F.D. and Aldridge,J.F. (1985) *J. Biol. Chem.*, **260**, 11322–11329.

Gluzman,Y. (1981) *Cell*, **23**, 175–182.

Hattori,M. and Sakaki,Y. (1986) *Anal. Biochem.*, **152**, 232–238.

Hidaka,Y., Palella,T.D., O'Toole,T.E., Tarle,S.A. and Kelley,W.N. (1987) *J. Clin. Invest.*, **80**, 1409–1415.
Krainer,A. and Maniatis,T. (1988) In Hames,P.D. and Glover,D.M. (eds), *Transcription and Splicing; Frontiers in Molecular Biology*. IRL Press, Oxford, pp. 131–206.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Marvit,J., Di Lella,A.G., Brayton,K., Ledley,F.D., Robson,K.J.H. and Woo,S.L.C. (1987) *Nucleic Acids Res.*, **15**, 5613–5628.
Miller,E.J. and Gay,S. (1987) *Methods Enzymol.*, **144**, 3–41.
Mitchell,P.J., Urlaub,G. and Chasin,L. (1986) *Mol. Cell. Biol.*, **6**, 1926–1935.
Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.
Padgett,R.A., Grabowski,P.J., Konarska,M.M., Seiler,S. and Sharp,P.A. (1986) *Annu. Rev. Biochem.*, **55**, 1119–1150.
Reed,R. and Maniatis,T. (1986) *Cell*, **46**, 681–690.
Rogers,J. and Wall,R. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 1877–1879.
Ruffner,D.E. and Dugaiczyk,A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2125–2129.
Saiki,R.K., Scharf,S., Faloona,F., Mullis,K.B., Horn,G.T., Erlich,H.A. and Arnheim,N. (1985) *Science*, **230**, 1350–1354.
Sanger,R., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Smith,B.D., Byers,P.H. and Martin,G.R. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 3260–3262.
Sompayrac,L.M. and Danna,K.J. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 7575–7678.
Steinmann,B., Tuderman,L., Peltonen,L., Martin,G.R., McKusick,V.A. and Prockop,D.J. (1980) *J. Biol. Chem.*, **255**, 8887–8893.
Templeton,D. and Eckhart,W. (1984) *Mol. Cell. Biol.*, **4**, 817–821.
Tromp,G. and Prockop,D.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 5254–5258.
Tsipouras,P. and Ramirez,F. (1987) *J. Med. Genet.*, **24**, 2–8.
Weil,D., Bernard,M., Combates,N., Wirtz,M.K., Hollister,D., Steinmann,B. and Ramirez,F. (1988) *J. Biol. Chem.*, **263**, 8561–8564.
Wirtz,M.K., Glanville,R.W., Steinmann,B., Rao,V.H. and Hollister,D. (1987) *J. Biol. Chem.*, **262**, 16376–16385.

Received on February 9, 1989; revised on March 3, 1989