A 5' splice site mutation affecting the pre-mRNA splicing of two upstream exons in the collagen COL1A1 gene

Exon 8 skipping and altered definition of exon 7 generates truncated $\alpha 1(I)$ chains with a non-collagenous insertion destabilizing the triple helix

John F. BATEMAN,* Danny CHAN, Ingrid MOELLER, Marnie HANNAGAN and William G. COLE

Orthopaedic Molecular Biology Research Unit, Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Parkville, Victoria 3052, Australia

A heterozygous *de novo* G to A point mutation in intron 8 at the +5 position of the splice donor site of the gene for the $\alpha 1(I)$ chain of type I procollagen, COL1A1, was defined in a patient with type IV osteogenesis imperfecta. The splice donor site mutation resulted not only in the skipping of the upstream exon 8 but also unexpectedly had the secondary effect of activating a cryptic splice site in the next upstream intron, intron 7, leading to re-definition of the 3' limit of exon 7. These pre-mRNA splicing aberrations cause the deletion of exon 8 sequences from the mature mRNA and the inclusion of 96 bp of intron 7 sequence. Since the mis-splicing of the mutant allele product resulted in the maintenance of the correct codon reading frame, the resultant $\alpha 1(I)$ chain contained a short non-collagenous

32-amino-acid sequence insertion within the repetitive Gly-Xaa-Yaa collagen sequence motif. At the protein level, the mutant $\alpha 1(I)$ chain was revealed by digestion with pepsin, which cleaved the mutant procollagen within the protease-sensitive non-collagenous insertion, producing a truncated $\alpha 1(I)$. This protease sensitivity demonstrated the structural distortion to the helical structure caused by the insertion. In long-term culture with ascorbic acid, which stimulates the formation of a mature crosslinked collagen matrix, and in tissues, there was no evidence of the mutant chain, suggesting that during matrix formation the mutant chain was unable to be stably incorporated into the matrix and was degraded proteolytically.

INTRODUCTION

The molecular basis of the heritable connective tissue disorder osteogenesis imperfecta (OI) has been studied extensively and involves mutations in the genes for the $\alpha 1(I)$ chains and $\alpha 2(I)$ chains of type I collagen in at least 90% of cases (for reviews, see [1,2]). Mutations covering the range of possible gene defects including rearrangements, deletions, insertions and a spectrum of point mutations have been defined. While glycine substitution mutations within the collagen helix are probably the most common OI mutation, another apparently common mechanism for the production of a dysfunctional collagen helix involves exon-skipping mutations.

Exon-skipping mutations have been defined in several other diseases [3–11], and exon-skipping (or alternative splicing) is also a mechanism for generating protein structural diversity [12–20]. The fibrillar collagen genes such as those coding for type I collagen $\alpha 1(I)$ and $\alpha 2(I)$ chains (COL1A1 and COL1A2) are particularly susceptible to RNA splicing mutations because mRNA maturation involves the correct splicing-out of some 52 intronic sequences [21]. In most proteins exon-skipping leads to codon frameshifts and the introduction of premature stop codons, resulting in the failure of RNA nuclear transport and degradation, or to the production of drastically truncated and altered protein sequences which are degraded, thus commonly leading to a functionally null allele. In collagen, however, the codon reading frame is maintained among the 42 exons coding for the collagen triple helix, and thus exon-skipping in this region results in the production of a collagen protein which contains an internal deletion in the helix but retains the repeating Gly-Xaa-Yaa sequence of the helix [1,2].

Collagen mutations leading to exon-skipping have been defined in both the 5' splice donor site and 3' splice acceptor site consensus sequences [1,2]. In this paper we describe a COL1A1 G to A mutation at IVS8 nt 5 which disturbs the consensus sequence for the splice donor site. This mutation has a more complicated consequence than other similar mutations previously described, in that as well as preventing the definition and correct splicing of exon 8 it also leads to the re-definition of the upstream exon 7 by activation of a cryptic 5' splice site in intron 7. The re-defined exon 7 includes an intron sequence which results in the production of $\alpha 1(I)$ chains containing an insertion of non-collagenous protein sequence within the helix, along with the deletion of sequence corresponding to exon 8. This altered protein has a reduced thermal stability and is not incorporated into the mature collagenous matrix.

EXPERIMENTAL

Clinical summary

The proband was the first member of her family to be affected with OI. She had severe bone fragility with marked retardation of growth. Her sclerae remained grey-blue throughout childhood and her teeth appeared normal. The severity of her clinical disease was consistent with OI type IV [22]. Skin biopsies were obtained from the patient, the parents and controls with the approval of the Human Ethics Committee of this hospital.

Fibroblast collagen biosynthetic labelling

Skin fibroblast cultures were established as previously described [23,24]. After culture for 3 days in growth medium containing

Abbreviation used: OI, osteogenesis imperfecta.

* To whom correspondence should be addressed.

0.25 mM sodium ascorbate (Sigma Chemical Co.), confluent fibroblast cultures were labelled with 10 $\mu\text{Ci}/\text{ml}$ L-[5- ^3H]proline (30 Ci/mmol, Amersham Corp.) for 18 h in medium containing 10% (v/v) dialysed foetal-calf serum (Flow Laboratories), 0.25 mM sodium ascorbate and 0.1 mM β -aminopropionitrile fumarate (Sigma Chemical Co.) [23,24]. In some experiments the 0.25 mM sodium ascorbate in the labelling medium was replaced with 0.1 mM $\alpha\alpha'$ -dipyridyl to prevent post-translational hydroxylation of the procollagen. The cell layer and medium fractions were separated for analysis and the procollagens, precipitated with 25% satd. $(\text{NH}_4)_2\text{SO}_4$, were subjected to limited proteolysis with pepsin [23,24].

Mammalian collagenase digestion

Aliquots of biosynthetically labelled fibroblast pepsin-treated collagen were digested with mammalian collagenase as described previously [25]. Samples were mixed with trypsin-activated collagenase at 25 °C for 24 h. The digestion products were mixed with sample loading buffer and analysed by SDS/10% (w/v) PAGE.

Extraction of matrix collagens

Collagens were extracted from either the *in vivo* skin matrix or the collagenous matrix produced *in vitro* by culture of skin fibroblasts in the presence of ascorbate [26]. Skin samples were freeze-milled, defatted and extracted as previously described [24]. Limited pepsin digestion was carried out and the solubilized collagens were analysed by gel electrophoresis. To produce *in vitro* collagenous matrices, cells were cultured for 28 days in the presence of sodium ascorbate. Under these culture conditions an extensive cross-linked collagenous matrix is deposited [26]. The cell layers were scraped into 50 mM Tris/HCl, pH 7.5, containing 0.15 M NaCl and proteinase inhibitors. The cell fractions were disrupted by sonication and extracted overnight at 4 °C. Insoluble material was removed by centrifugation and the procollagen and collagen in the supernatant (neutral salt-soluble fraction) was precipitated with ammonium sulphate (25%, v/v, saturated) and redissolved in 50 mM Tris/HCl (pH 7.5)/0.15 mM NaCl containing proteinase inhibitors. Insoluble collagen in the pellet was further extracted at 4 °C for 16 h with 0.5 M acetic acid and the collagen thus solubilized was lyophilized (acetic-acid-soluble fraction). Acetic-acid-insoluble residue was digested with pepsin (0.1 $\mu\text{g}/\text{ml}$ pepsin) in 0.5 M acetic acid for 16 h at 4 °C [25]. Digestion was terminated by lyophilization.

SDS/PAGE

Collagen α -chains were resolved on 5% (w/v) separating gels containing 2 M urea and detected by fluorography [23,24]. Coomassie Blue-stained gels were quantified by densitometry in comparison with standard collagen samples loaded on each gel.

Thermal stability of collagen

The freeze-dried [^3H]proline-labelled collagens were dissolved in 0.4 M NaCl/0.1 M Tris/HCl buffer, pH 7.4, at 4 °C. The samples were warmed stepwise (1 °C/10 min) from 34 to 43 °C and at 0.5 °C intervals samples were taken and digested with a mixture of trypsin and chymotrypsin [27]. The collagen that was resistant to proteolysis at each temperature interval was resolved by

electrophoresis and identified by fluorography. The thermal denaturation temperature, T_m , was defined as the temperature at which half of the collagen was degraded.

Formation and chemical cleavage of mRNA cDNA heteroduplexes

Total RNA was isolated from confluent fibroblast cultures [28]. The mRNA·cDNA heteroduplexes were formed in a volume of 50 μl containing ~ 5 ng (50 000–100 000 d.p.m.) of ^{32}P -end-labelled control $\alpha 1(\text{I})$ cDNA probe, 5–10 μg of total RNA, 80% (v/v) formamide, 40 mM Pipes, pH 6.4, 1 mM EDTA and 0.4 M NaCl. The mixture was denatured at 80 °C for 5 min, incubated at 60 °C for 2 h and ethanol-precipitated. Chemical modification of mismatched nucleotides with hydroxylamine or osmium tetroxide, cleavage with piperidine and electrophoresis of the products on denaturing 7 M urea/5% (w/v) acrylamide gels have been described [29–32].

Amplification and sequencing

Total RNA was prepared from fibroblast cultures grown for 3 days with ascorbate at 31 or 37 °C [26]. First-strand cDNA was synthesized using a cDNA synthesis kit (Amersham Corp.) primed with a specific $\alpha 1(\text{I})$ oligonucleotide, 5' TCTGACCAGGAGCTCCATTT 3' (antisense), which was 3' to the mismatched bases, corresponding to nucleotides 1001–1020 of the normal $\alpha 1(\text{I})$ cDNA [33]. The base pairs of the COL1A1 are numbered according to D'Alessio et al. [21] and the cDNA sequences are numbered from the start of transcription of the pro $\alpha 1(\text{I})$ mRNA [33]. cDNA (~ 50 ng) was amplified by the PCR [34] through 30 cycles using Taq polymerase (Perkin Elmer-Cetus) [35]. Each cycle consisted of denaturation at 92 °C for 30 s, annealing of primers at 64 °C for 1 min and primer extension at 72 °C for 3 min [31,32]. PCR primers were 5' TCAACCGGAGGAATTTCCGT 3' (sense) and 5' ATCTCCATTCTTTCCAGGG 3' (antisense), which in controls amplified a 183 bp fragment corresponding to bases 630–812 of the $\alpha 1(\text{I})$ cDNA [33]. DNA was extracted from fibroblast cultures by digestion with 0.5 $\mu\text{g}/\text{ml}$ proteinase K in 3 ml of 10 mM Tris/HCl buffer, pH 8.2, containing 0.4 M NaCl, 2 mM EDTA and 1% (w/v) SDS at 50 °C for 16 h. The DNA was ethanol-precipitated from the supernatant and re-dissolved in 1 mM Tris/HCl, pH 7.5, containing 1 mM EDTA. COL1A1 genomic sequence encompassing the region from exon 6 to exon 9 was amplified for 30 cycles using the same set of primers. Each cycle consisted of denaturation at 93 °C for 90 s, annealing at 63 °C for 90 s and extension at 74 °C for 3 min.

cDNA and genomic amplification products were analysed by electrophoresis and bands of the predicted size were purified, treated with T4 polynucleotide kinase [36] and cloned into a *Sma*I-cut, dephosphorylated M13mp8 vector (Amersham Corp.). Multiple clones from independent amplification reactions were sequenced using a Sequenase kit (United States Biochemical Corp.) and the M13 universal primer. Direct sequencing of genomic PCR products was also performed as described by Bachmann [37] using an intron-7-specific primer (5' GTGAGTATCCAGGACGTCTT 3').

RESULTS

Collagen protein analysis

Preliminary electrophoretic analysis of the pepsin-digested collagen produced by O14 fibroblasts labelled in cell culture indicated the presence of an unique band migrating faster than the $\alpha 1(\text{I})$ chain of type I collagen [38]. The presence of this species in

Figure 1 SDS/5%-w/v-PAGE of fibroblast procollagen and collagen

(a) Fibroblast cultures were labelled for 18 h with [^3H]proline in the presence of ascorbate, and collagens in the cell layer (lanes 1–4) and secreted into the medium (lanes 5–8) were pepsin-digested and analysed without prior reduction (see the Experimental section for details). Lanes 1 and 5, OI4; lanes 2 and 6, control; lanes 3 and 7, the father; and lanes 4 and 8, the mother. (b) Fibroblast cultures were labelled at 32 °C with [^3H]proline, and collagens in the cell layer (lanes 1) and secreted into the medium (lanes 2) were pepsin-digested and analysed without prior reduction. (c) Fibroblast cultures were labelled for 18 h with [^3H]proline in the presence of $\alpha\alpha'$ -dipyridyl and the unhydroxylated procollagens in cell-layer samples were analysed after reduction with 10 mM dithiothreitol. Lane 1, OI4; lane 2, control. The migration positions of type I collagen pro α 1(I) and pro α 2(I) α 1(I) and α 2(I) chains, type V collagen α 1(V) and α 2(V) chains and type III collagen [α 1(III) $_3$] are shown. The abnormal faster-migrating α -chain is indicated by arrowheads.

Figure 2 Pepsin sensitivity of OI4 procollagen

Labelled procollagens from OI4 and control fibroblast cultures were subjected to digestion with final concentrations of pepsin of 0.01–100 $\mu\text{g/ml}$ for 16 h at 4 °C. Samples were analysed without prior reduction on 5% (w/v) SDS/polyacrylamide gels. The migration positions of type I collagen α 1(I) and α 2(I) chains, partially processed procollagen pN α 1(I) and pN α 2(I) chains and type III collagen [α 1(III) $_3$] are shown. The mutant band was not detected in analyses of pepsin-digested collagens produced by fibroblasts from either parent or controls (Figure 1a, lanes 2–4 and 6–8). The mutant collagen species was also not apparent in analyses of unhydroxylated

cell layer and cell culture medium pepsin digests indicated that the species was secreted from the cell (Figure 1, lanes 1 and 5). When fibroblasts were biosynthetically labelled at 32 °C, rather than 37 °C, a greater proportion of the mutant species was evident (Figure 1b), demonstrating that the mutant protein was more stable at the lower temperature. The mutant protein band was confirmed as a collagenous species by digestion with bacterial collagenase, and Western gel analysis using specific antibodies identified the band as a type I collagen α 1(I) chain (results not shown). The mutant band was not detected in analyses of pepsin-digested collagens produced by fibroblasts from either parent or controls (Figure 1a, lanes 2–4 and 6–8). The mutant collagen species was also not apparent in analyses of unhydroxylated

Figure 3 Electrophoresis of mammalian collagenase digests of fibroblast pepsin-treated collagens

Labelled collagens were digested with mammalian collagenase and analysed on SDS/10% (w/v) polyacrylamide gels (see the Experimental section for details). Lane 1, control; lane 2 OI collagen. α 1(I)TC $_A$ and α 2(I)TC $_A$ designate the mammalian collagenase cleavage fragments from the N-terminal portion of the α 1 and α 2 chains. The abnormal faster migrating α 1(I)TC $_A$ is indicated by an arrowhead.

procollagens produced by the cells (Figure 1c, lane 1), suggesting that the mutant band may have arisen during the limited pepsin digestion of the samples before electrophoretic analysis.

This possibility was explored by subjecting the normally hydroxylated procollagens produced by the cells to increasing concentrations of pepsin (Figure 2). At a low pepsin concentration (0.01 $\mu\text{g/ml}$) only partial conversion of procollagen into collagen was achieved and the mutant collagen band was not detected. As proteolytic digestion conditions became harsher, the conversion of procollagen into collagen was more efficient. Under these more stringent digestion conditions (0.1–100 $\mu\text{g/ml}$ pepsin) the mutant-specific collagen band appeared. Digestion of OI4 procollagen with trypsin or chymotrypsin, as well as pepsin, at 20 °C also produced the smaller α 1(I) chain (results not shown). These results suggest that the mutant collagen species contained protease-sensitive sequences which, when cleaved, produced a protease-stable (at 20 °C) species which is smaller than a normal α 1(I) chain.

Digestion with mammalian collagenase (Figure 3) resulted in the production of two forms of the N-terminal fragment of α 1(I) [α 1(I)TC $_A$], further identifying the mutant pepsin-produced species as a smaller form of the α 1(I), which contained the protease-sensitive mutation within the N-terminal three quarters of the collagen α 1(I) helical domain. The normal migration of the C-terminal α 1(I) fragment [α 1(I)TC $_B$] and both fragments of the α 2(I) chain were consistent with this proposal (results not shown).

Collagen melting temperature

Estimation of the susceptibility to thermal denaturation of mutant and normal type I collagen synthesized by OI fibroblasts was made by measuring its resistance to protease digestion after heating to different temperatures (Figure 4). Pepsin-treated collagen helices containing the shortened mutant α 1(I) chains were much more readily thermally denatured than control collagen. Because of the relatively small proportion of the mutant present, it was not possible to measure accurately the T_m , but since all the mutant became unstable between 34 and 36 °C the estimated T_m of the mutant collagen was ~ 35 °C. The T_m of control collagen was the same as that determined in numerous other studies of the thermal denaturation of type I collagen molecules, i.e. 40.9 °C [27]. The T_m of OI type III collagen was the same as control values (results not shown).

Figure 4 Thermal denaturation of type I collagen

[³H]Proline-labelled collagens prepared from the medium of OI4 were treated with pepsin and after inactivation of the pepsin the collagens were warmed slowly from 20 to 38 °C. Serial samples were taken and digested with trypsin/chymotrypsin and the resistant chains were resolved by electrophoresis (see the Experimental section for details). The migration positions of type I collagen $\alpha 1(I)$ and $\alpha 2(I)$ chains and type III collagen [$\alpha 1(III)_3$] are shown. The abnormal faster-migrating $\alpha 1(I)$ -chain is indicated by an arrowhead.

Extraction of collagens from the dermis and collagenous matrices produced *in vitro*

To determine whether the mutant collagen was deposited into a crosslinked collagenous matrix, tissue-equivalent fibroblast collagen matrices and dermal samples from the patient and controls were extracted and the collagen was analysed after pepsin digestion. Fibroblasts were cultured in the presence of ascorbic acid for 28 days, to allow the deposition of a cross-linked collagenous matrix [26]. Collagen deposition was monitored by Coomassie Blue staining of serial extracts and the species synthesized determined by pulse-labelling and analysis of collagens secreted into the medium (results not shown). The OI fibroblasts showed a substantial reduction in the extent of collagen matrix deposition *in vitro* (66% of that in control cells). Deficiencies of collagen matrix formation by OI fibroblasts *in vitro* have been demonstrated previously [39]. The OI4 fibroblast collagen matrix required pepsin digestion for extraction, indicating the presence of extensive collagen crosslinking. In this mature matrix there was no evidence for the presence of the mutant type I collagen, whereas during pulse-labelling the synthesis of the mutant was clearly demonstrated although the amount of the stable mutant represented < 10% of the total type I collagen synthesized. These data demonstrated that while the mutant $\alpha 1(I)$ is synthesized, it is not incorporated into the matrix *in vitro*. Pepsin extraction of dermal samples from the patient confirmed that the mutant chains were absent from the mature collagenous matrix *in vivo*.

Detection and sequencing of abnormal mRNA sequences

Heteroduplexes were formed using a control $\alpha 1(I)$ cDNA 427 bp *NarI-NarI* fragment which spanned the N-terminal half of the procollagen molecule, and mRNA samples were extracted from cultured OI4 and control fibroblasts. Both hydroxylamine/piperidine and osmium tetroxide/piperidine treatment resulted in multiple cDNA cleavages, demonstrating the presence of mismatched Cs and Ts (Figure 5). The upper and lower size limits of the cleavage products defined the boundaries of the mis-

Figure 5 Detection of the $\alpha 1(I)$ mRNA mutations by chemical modification and cleavage

Total RNA from OI and parental fibroblasts was hybridized to a 427 bp *NarI-NarI* ³²P-labelled fragment of the $\alpha 1(I)$ cDNA and allowed to react with hydroxylamine or osmium tetroxide for 0–30 min as indicated, treated with piperidine and the resultant fragments resolved on a 5% (w/v) polyacrylamide gel (see the Experimental section for details). In OI4 multiple cleavages at mismatched and thus reactive Cs by hydroxylamine and Ts by osmium tetroxide produced cleavage products of ~250–300 bp. ³²P-labelled *HaeIII* fragments of ϕ X174 DNA molecular mass standards (right) were used to calculate the sizes of the cleavage fragments.

Figure 6 Electrophoresis of collagen cDNA-PCR products from control and OI fibroblasts

cDNA was amplified by PCR using oligonucleotides in exons 6 and 9 (see Experimental section for details). Aliquots of each reaction mixture were resolved by electrophoresis on a 2.5% (w/v) agarose gel and stained with ethidium bromide. The sizes of the amplified products are indicated. Lane 1, control fibroblasts; lane 2, OI fibroblasts grown at 37 °C; lane 3, OI fibroblasts grown at 31 °C.

matched region and suggested the presence of a deletion that resulted in a mismatched sequence of ~50–60 bp in the control $\alpha 1(I)$ cDNA, the location of which suggested that the mismatched sequences were encoded by exon 8. Chemical cleavage of heteroduplexes formed with mRNA from fibroblast cultures of the parents of OI4 indicated that the exon 8 mismatch was not present.

To define the mutation, first-strand $\alpha 1(I)$ cDNA was prepared from mRNA extracted from cells grown at 37 or 31 °C, and amplified by the PCR using unique oligonucleotides in exon 6 and 9, chosen to span the region containing the mismatches (Figure 6). Amplification of control mRNA resulted in the production of a single major species (183 bp) corresponding to

Figure 7 Normal and mutant genomic COL1A1 DNA sequences

(a) The exon sequences are shown in upper-case letters and intron sequence in lower-case letters. The nucleotides in the normal sequence which are altered in the OI4 allele are underlined with the substituted nucleotide in the mutant sequence shown above. The nucleotide numbering scheme is based on that of D'Alessio et al. [21] re-numbered (see Results section for details) to account for differences to the published sequence. The second line gives the deduced amino acid sequence of the mutant allele product and the third line the normal amino acid sequence. Amino acids are numbered from the N-terminus of the collagen triple helix. (b) Diagrammatic representation of the RNA splicing pattern of the normal and mutant alleles.

the expected size of the exon 6 to exon 9 PCR product (Figure 6, lane 1). In the OI4 fibroblasts, grown at either 37 or 31 °C (Figure 6, lanes 2 and 3), two amplified products of similar abundance were apparent. One of these corresponded to the expected 183 bp fragment seen in the controls, but the other was a larger product of 225 bp which must therefore contain an additional sequence inserted into the mature mRNA. Sequencing of M13 subclones of the amplified cDNA products identified the mutation as an insertion of 96 bp of intron 7 sequence and the deletion of exon 8 sequences from the mRNA. The abnormal sequence was confirmed by sequencing of multiple clones from two separate PCR reactions. Clones containing the normal sequence were also identified, confirming that the patient was heterozygous for the $\alpha 1(I)$ mutation.

Definition of the gene defect

To define the genomic mutation that resulted in the exon-skipping and inclusion in the mature mRNA of intron sequences, genomic sequencing of OI4 DNA was performed by PCR amplification, cloning and sequencing. These data demonstrate a G to A mutation within the 5' splice donor site at position IVS8 nt 5 (Figure 7a). In the same clones exhibiting the G to A

mutation a second alteration of C to G within the coding sequence of exon 8 was observed. These changes were confirmed by direct sequencing of genomic PCR products (results not shown). The C to G alteration would result in a Pro to Ala substitution in the exon-8-encoded sequence, but since exon 8 is spliced out in the mutant allele product, this sequence variant is not expressed at the protein level.

There were numerous other inconsistencies with the published genomic sequence [21], but since all these changes were found in clones of both alleles of OI4 and in normal controls, they most likely represent errors in the original published data or differences between the individual sequenced previously [21] and our population. The majority of the differences are single-base insertions, deletions or reassignments in intron sequences. For the record, deletions of T occurred at positions 3306, 3312, 3317, 3337 and 3342 and of A at 3309 and 3354, and a CA deletion occurred at 3541. Base reassignments of T to A at 3266, C to T at 3589 and single-base insertions of G at 3728 and C at 3797 also occurred (sequence numbers correspond to those used by D'Allesio et al. [21]). As well as these minor sequence differences, a strikingly different sequence stretch was defined. The 22 bp sequence stretch from position 3549 to position 3570 in the published sequence [21] was not detected in either allele of OI4 or in analyses of two

control DNA samples; instead, a unique sequence of 34 bp was defined in this region. These sequence alterations are shown in Figure 7(a) and the altered sequence is re-numbered accordingly. According to the re-numbered sequence the G to A 5' splice mutation occurred at nt 3658.

DISCUSSION

The proband was shown to have OI type IV due to a heterozygous *de novo* mutation in the gene for the $\alpha 1(I)$ chain of type I procollagen, COL1A1 (Figure 7b). The G-to-A point mutation in intron 8 at the +5 position of the 5' splice donor site resulted in skipping of the upstream exon 8. The novel feature of this mutation was that as well as exon-skipping the mutation activated a cryptic splice site in the upstream intron 7, leading to re-definition of the 3' limit of exon 7. The result of this pre-mRNA splicing aberration was the deletion of exon 8 sequences from the mature mRNA and the inclusion of 96 bp of intron 7 5' sequence. Since the mis-splicing of the mutant allele product resulted in the maintenance of the correct codon reading frame, the resultant $\alpha 1(I)$ chain contained a short non-collagenous amino acid sequence insertion within the repetitive Gly-Xaa-Yaa collagen sequence motif.

This mutation has important consequences for the further development of our mechanistic understanding of *in vivo* pre-mRNA splicing. Since experimental approaches have predominantly used *in vitro* model systems, such as minigene constructs, it is of importance to correlate these *in vitro* mechanisms with the effect of mutations on pre-mRNA splicing of complex multi-exon genes, like the collagen α -chains, *in vivo*. A widely accepted model for pre-mRNA splicing proposes that exon definition occurs at the initial stages of spliceosome assembly [40]. In this model, after spliceosome binding to the 3' acceptor site consensus sequence, the spliceosome complex scans for a suitable 5' donor site, (C/A)AGGT(A/G)AGT [41,42]. Once such a consensus sequence is recognized, the exon limits are defined and subsequent spliceosome interactions define the resulting intron which is removed via a set of transesterification steps in which intron loops are removed by interactions with specific intronic branch point sequences.

According to this model, 5' donor site mutations can result in two general phenotypes. In the first, spliceosome scanning within the vicinity of the mutated donor site locates an appropriate cryptic 5' site, and the exon adjacent to the mutation is re-defined as either larger or smaller than its original size [42,43]. If no suitable new 5' donor site can be identified by the spliceosome complex within ~ 300 bp downstream of the 3' site, the maximum exon size is exceeded [44], and the sequence is not recognized as an exon, and the entire potential exon is spliced out along with the adjacent intron sequences. Such exon-skipping has resulted from a number of 5' splice site mutations [1,2] and may even be the most common phenotype of donor site mutations [44]. In the case of the fibrillar collagens, the codon reading frames are not split between the exons coding for the helical domains and thus exon-skipping results in the excision of exon-encoded sequences and a resultant protein truncation, but with maintenance of the Gly-Xaa-Yaa sequence motif.

In COL1A1, intron 8 contains no recognizable cryptic 5' splice site sequences within the exon size window, explaining the failure of the spliceosome complex to define exon 8, and out-splicing of exon 8 sequences from the mature mRNA occurs in OI4. In this patient, however, the second consequence of the mutation in the intron 8 donor site was the re-definition of exon 7 by activation of a cryptic 5' splice donor site in intron 7, 96 bp downstream from the normally used 5' site. A scoring system based on the

frequency of each nucleotide in published consensus sequences has been devised [45] to give a theoretical measure of the potential strength of putative 5' and 3' splicing signals. These calculations on the genomic sequence spanning exons 7–9 indicated that the G to A mutation at IVS8 nt 5 reduced the score of this 5' donor site to below that of actively utilized splice sites. Downstream to the 5' site, no strong 5' signals were identified. These predictions were consistent with the failure of exon 8 definition and out-splicing. The activated intron 7 cryptic 5' site in the mutant allele product was a strong donor site and, in fact, scored better than the normally used upstream intron 7 5' site.

Given that the cryptic site, which is present in both normal and mutant alleles, is a strong 5' donor consensus sequence, why is it not used during pre-mRNA splicing in the normal allele? The answer may be the proximity of the cryptic site to the intron 7/exon 8 junction. The cryptic site is only 62 bp from the intron/exon boundary and this is less than the observed minimum mammalian intron size of 70 bp [41], which is presumed to be necessary for intron rearrangements during looping and removal [40]. However, in the mutant allele product, exon 8 sequences are not defined due to the donor site mutation, and the proximity of the cryptic site to exon 8 is no longer a restriction, since intron 7 is effectively lengthened by the inclusion of exon 8 and intron 8 sequences (see Figure 7b). While this remains speculative, the data support the model proposed for an initial exon definition step in pre-mRNA splicing and they also provide evidence of the minimum size of introns required for normal splicing. Furthermore, these *in vivo* findings support the proposal that mechanisms are concerted rather than sequential, and sequence contexts and co-operation between spliceosomes may influence splice sequence selection [40].

If splicing of the mutant allele were variable, the possible products would be the normal sequence including exon 8 sequences, a simple exon 8 skip, a product including exon 8 but with mis-splicing at the intron 7 cryptic site, or the observed product consisting of the exon 8 skip and intron 7 cryptic site activation. When the patient's mRNA was PCR-amplified only two products were detected (Figure 6): the expected 183 bp band, containing the normal exon 8 sequence, and the mutant 225 bp band. Direct sequencing of the 183 bp band demonstrated that it did not contain the mutant-allele specific variation of the exon 8 sequence (C to G; see Figure 7), confirming that the exon-8-encoded region was derived solely from the normal allele. Sequencing of the 225 bp band demonstrated that it corresponded to the exon 8 skip and intron 7 sequence insertion. Thus the mutation results in the production of only one form of mis-spliced product expressed at approximately equal efficiency to the normal COL1A1 allele product. Several cases of mis-splicing have shown temperature sensitivity, indicating that these mutations altered splicing preference but normal splicing patterns could be restored by culture of cells at less stringent temperatures [46]. However, in this case, the mis-splicing was not temperature-sensitive, suggesting that the G to A point mutation in intron 8 at the +5 position of the splice donor site altered the splicing pattern irreversibly. This is consistent with the model proposed where the mutation leads to a large reduction in the predicted splice site strength and the activation of a theoretically strong cryptic site in intron 7.

The activation of the cryptic splice site in intron 7 led to a predicted insertion of 32 amino acids coded by intron sequence, and the deletion of the 18-amino-acid exon 8 sequence from the predicted protein sequence (Figure 7a). Since the codon reading frame was maintained, protein sequences C-terminal to exon 9 were unchanged. While the exon 8 deletion maintained the Gly-Xaa-Yaa repeating sequence of the collagen helix, and thus could

most probably be accommodated without severe disruption to the helical structure, the 32-amino-acid non-collagen sequence would be expected to create a major disturbance to the helix at the point of insertion, 18 amino acids from the N-terminal end of the triple helix and the junction with the N-telopeptide. The production of a truncated mutant $\alpha 1(I)$ by digestion with pepsin directly demonstrated the disruption to the helical structure at the point of the insertion, since normal triple-helical collagen is pepsin-resistant. Further evidence of the disturbance to the helix was provided by thermal stability studies showing a marked reduction in the T_m of the mutant collagen.

The data suggest that the mutation disrupts the co-operative stability of the collagen helix [47], and it is unlikely that the downstream N-terminal amino acids will re-nucleate to form a stable triple helix, since the 18-amino-acid collagenous sequence stretch is too short for formation of a triple helix in aqueous solution (H.-P. Bachinger, personal communication). The consequent dramatic reduction in the thermal stability of the collagen molecule indicates a critical role for this N-terminal helical domain in stabilizing the triple-helix structure. In support of this, the absence of this stable region in the N-termini has been shown to have a destabilizing effect in type V and XI collagens [48].

The synthesis of normal and mutant $\alpha 1(I)$ mRNA in fibroblast cultures were comparable, while at the protein level only small amounts of the mutant was detected after pepsin digestion. This discrepancy most probably results from the increased degradation of the mutant collagen, which is destabilized by the non-collagenous insertion. Once a cut has been made within the proteolytically exposed region of the insertion, the mutant collagen helix should readily denature since the reduced T_m is below physiological temperatures, rendering the mutant molecules vulnerable to further proteolytic attack. The data demonstrating a higher proportion of mutant at 32 °C provide further evidence of the instability of the mutant protein. The mechanism of this degradation is not known and both intracellular degradation of the mutant [2,23] and extracellular turnover may be involved. In long-term culture with ascorbic acid, which stimulates the formation of a mature collagen matrix, and in tissues, no mutant chain was present, providing further evidence that *in vivo* the mutant collagen is degraded and unable to be stably incorporated into the collagen matrix. The reduced accumulation of the collagen matrix *in vitro* probably reflects this breakdown of mutant collagen [39]. It is likely however, that while the mutant is not stably incorporated into the matrix, it does exert a deleterious effect on matrix formation or organization, possibly indicating some form of transient association or incorporation. In another OI fibroblast (heterozygous for a helical Gly-to-Arg mutation), mutant chains were selectively degraded after incorporation into the matrix [39]. If the OI4 mutant collagen chains were fully degraded before matrix assembly it would be equivalent to a functionally null allele and the clinical phenotype could be expected to be less severe, similar to the case for OI type I, where half-normal levels of type I collagen lead to a clinically milder phenotype.

This work was supported by grants from the National Health and Medical Research Council of Australia and the Royal Children's Hospital Research Foundation.

REFERENCES

- 1 Byers, P. H. (1990) Trends Genet. **6**, 293–300
- 2 Kuivaniemi, H., Tromp, G. and Prockop, D. J. (1991) FASEB J. **5**, 2052–2060

- 3 Carstens, R. P., Fenton, W. A. and Rosenberg, L. R. (1991) Am. J. Hum. Genet. **48**, 1105–1114
- 4 Marvit, J., DiLella, A. G., Brayton, K., Ledley, F. D., Robson, K. J. H. and Wood, S. L. C. (1987) Nucleic Acids Res. **15**, 5613–5628
- 5 Green, P. M., Montadon, A. J., Bentley, D. R. and Giannelli, F. (1991) Blood Coagulation Fibrinolysis **2**, 539–565
- 6 De Boer, M., Bolscher, B. G., Dinuer, M. C., Orkin, S. H., Smith, C. I., Ahlin, A., Weening, R. S. and Roos, D. (1992) Blood **80**, 1553–1558
- 7 Berg, L. P., Grundy, C. B., Thomas, F., Millar, D. S., Green, P. J., Slomski, R., Reiss, J., Kakkar, V. V. and Cooper, D. N. (1992) Genomics **13**, 1359–1361
- 8 Eisensmith, R. C. and Woo, S. L. C. (1992) Hum. Mutat. **1**, 13–23
- 9 Sakaraba, H., Eng, C. M., Desnick, R. J. and Bishop, D. F. (1992) Genomics **12**, 643–650
- 10 Andrews, L. G. and Markert, M. L. (1992) J. Biol. Chem. **267**, 7834–7838
- 11 Dworniczak, B., Aulehla-Scholz, C., Kalaydjieva, L., Bartholomse, K., Grudka, K. and Horst, J. (1991) Genomics **11**, 242–246
- 12 Parks, W. C. and Deak, S. B. (1990) Am. J. Respir. Cell. Mol. Biol. **2**, 399–406
- 13 Saitta, B., Stokes, D. G., Vissing, H., Timpl, R. and Chu, M.-L. (1990) J. Biol. Chem. **265**, 6473–6480
- 14 Ryan, M. C. and Sandell, L. J. (1990) J. Biol. Chem. **265**, 10334–10339
- 15 Jackson-Grusby, L., Kuo, A. and Leder, P. (1992) Genes Dev. **6**, 29–37
- 16 Buckingham, M. E. (1989) Br. Med. Bull. **45**, 608–629
- 17 Smith, C. W., Patton, J. G. and Nadal-Grinard, B. (1989) Annu. Rev. Genet. **23**, 527–577
- 18 Fulle, P. J. (1991) FASEB J. **5**, 3092–3099
- 19 Strehler, E. E., Heim, R. and Carafoli, E. (1991) Adv. Exp. Med. Biol. **307**, 251–261
- 20 Schwarzbauer, J. E. (1991) BioEssays **13**, 527–533
- 21 D'Alessio, M., Bernard, M., Pretorius, P. J., De Wet, W. and Ramirez, F. (1988) Gene **67**, 105–115
- 22 Silience, D. O., Senn, A. S. and Danks, D. M. (1979) J. Med. Genet. **16**, 101–116
- 23 Bateman, J. F., Mascara, T., Chan, D. and Cole, W. G. (1984) Biochem. J. **217**, 103–115
- 24 Bateman, J. F., Chan, D., Mascara, T., Rogers, J. G. and Cole, W. G. (1986) Biochem. J. **240**, 699–708
- 25 Bateman, J. F., Mascara, T., Chan, D. and Cole, W. G. (1987) J. Biol. Chem. **262**, 4445–4451
- 26 Chan, D., Lamande, S. R., Cole, W. G. and Bateman, J. F. (1990) Biochem. J. **269**, 175–181
- 27 Baker, A. T., Ramshaw, J. A. M., Chan, D., Cole, W. G. and Bateman, J. F. (1989) Biochem. J. **261**, 253–257
- 28 Wake, S. A. and Mercer, J. F. B. (1985) Biochem. J. **228**, 425–432
- 29 Cotton, R. G. H., Rodrigues, N. R. and Campbell, R. D. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 4397–4401
- 30 Dahl, H.-H. M., Lamande, S. R., Cotton, R. G. H. and Bateman, J. F. (1989) Anal. Biochem. **183**, 263–268
- 31 Lamande, S. R., Dahl, H.-H. M., Cole, W. G. and Bateman, J. F. (1989) J. Biol. Chem. **264**, 15809–15812
- 32 Bateman, J. F., Lamande, S. R., Dahl, H.-H., Chan, D., Mascara, T. and Cole, W. G. (1989) J. Biol. Chem. **264**, 10960–10964
- 33 Tromp, G., Kuivaniemi, H., Stacey, A., Shikata, H., Baldwin, C., Jaenisch, R. and Prockop, D. J. (1988) Biochem. J. **253**, 919–922
- 34 Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988) Science **239**, 487–491
- 35 Kogan, S. C., Doherty, M. and Gitschier, J. (1987) New Eng. J. Med. **317**, 985–990
- 36 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning, A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- 37 Bachmann, B., Luke, W. and Hunsmann, G. (1990) Nucleic Acids Res. **18**, 1309
- 38 Bateman, J. F., Chan, D., Lamande, S., Mascara, T. and Cole, W. G. (1988) Ann. N.Y. Acad. Sci. **543**, 95–105
- 39 Bateman, J. F. and Golub, S. B. (1994) Matrix Biol. **14**, 251–262
- 40 Robberson, B. L., Cote, G. J. and Berget, S. M. (1990) Mol. Cell Biol. **10**, 84–94
- 41 Penotti, F. E. J. (1991) Theor. Biol. **150**, 385–420
- 42 Iida, Y. (1990) J. Theor. Biol. **145**, 523–533
- 43 Padgett, R. A., Grabowski, P. J., Konarski, M. M., Seiler, S. and Sharp, P. A. (1986) Annu. Rev. Biochem. **55**, 1119–1150
- 44 Talerico, M. and Berget, S. M. (1990) Mol. Cell Biol. **10**, 6299–6305
- 45 Shapiro, M. B. and Senapathy, P. (1987) Nucleic Acids Res. **15**, 7155–7174
- 46 Bonadio, J., Ramirez, F. and Barr, M. (1990) J. Biol. Chem. **265**, 2262–2268
- 47 Bachinger, H.-P. and Davis, J. M. (1991) Int. J. Biol. Macromol. **13**, 152–156
- 48 Morris, N. P., Watt, S. L., Davis, J. M. and Bachinger, H.-P. (1990) J. Biol. Chem. **265**, 10081–10087