Am. J. Hum. Genet. 62:98–110, 1998

Analysis of the COL1A1 and COL1A2 Genes by PCR Amplification and Scanning by Conformation-Sensitive Gel Electrophoresis Identifies Only COL1A1 Mutations in 15 Patients with Osteogenesis Imperfecta Type I: Identification of Common Sequences of Null-Allele Mutations

Jarmo Körkkö,[1,4] Leena Ala-Kokko,[1,4] Anne De Paepe,[3] Lieve Nuytinck,[3] James Earley,[1,2] and Darwin J. Prockop[1]

[1]Center for Gene Therapy, Allegheny University of the Health Sciences, MCP–Hahnemann School of Medicine, and [2]Department of Biochemistry and Molecular Biology, Jefferson Medical College, Philadelphia; [3]Collagen Research Unit, Biocenter and Department of Medical Biochemistry, University of Oulu, Oulu, Finland; and [4]Department of Medical Genetics, University of Gent, Gent, Belgium

Summary

Although >90% of patients with osteogenesis imperfecta (OI) have been estimated to have mutations in the COL1A1 and COL1A2 genes for type I procollagen, mutations have been difficult to detect in all patients with the mildest forms of the disease (i.e., type I). In this study, we first searched for mutations in type I procollagen by analyses of protein and mRNA in fibroblasts from 10 patients with mild OI; no evidence of a mutation was found in 2 of the patients by the protein analyses, and no evidence of a mutation was found in 5 of the patients by the RNA analyses. We then searched for mutations in the original 10 patients and in 5 additional patients with mild OI, by analysis of genomic DNA. To assay the genomic DNA, we established a consensus sequence for the first 12 kb of the COL1A1 gene and for 30 kb of new sequences of the 38-kb COL1A2 gene. The sequences were then used to develop primers for PCR for the 103 exons and exon boundaries of the two genes. The PCR products were first scanned for heteroduplexes by conformation-sensitive gel electrophoresis, and then products containing heteroduplexes were sequenced. The results detected disease-causing mutations in 13 of the 15 patients and detected two additional probable disease-causing mutations in the remaining 2 patients. Analysis of the data developed in this study and elsewhere revealed common sequences for mutations causing null alleles.

Received September 5, 1997; accepted for publication November 6, 1997; electronically published January 16, 1998.

Address for correspondence and reprints: Dr. Darwin J. Prockop, Center for Gene Therapy, Allegheny University of the Health Sciences, MCP–Hahnemann School of Medicine, 245 North 15 Street, 10118 NCB, Mail Stop 421, Philadelphia, PA 19102-1192. E-mail: prockop@auhs.edu

© 1998 by The American Society of Human Genetics. All rights reserved.

0002-9297/98/6201-0016$02.00

Introduction

Osteogenesis imperfecta (OI) is a heritable disorder that produces varying degrees of bone fragility associated with defects in several other tissues rich in type I collagen. Almost 200 mutations in the two genes for type I procollagen (COL1A1 and COL1A2) have been detected in probands with OI (Prockop 1990; Byers 1993; Prockop and Kivirikko 1995; Kuivaniemi et al. 1997). The most severe variants of the disease (OI types II–IV) are caused primarily by single-base substitutions that convert a codon for an obligate glycine in the triple helix of the protein to a codon for an amino acid with a bulkier side chain that distorts the conformation of the triple helix. Most of the mutations in the mildest variant (OI type I) cause decreased expression of proα1(I) chains because of either premature-termination codons or RNA-splicing defects in the COL1A1 gene (Willing et al. 1992, 1994, 1996; Redford-Badwal et al. 1996; Körkkö et al. 1997). DNA linkage studies have suggested that >90% of probands with OI have a mutation in either the COL1A1 gene or the COL1A2 gene (Sykes et al. 1990). However, most unrelated probands have different mutations, and it has been difficult to define the mutations in a number of patients with OI type I by means of previously employed protocols, which have sought to assay type I procollagen synthesized by either cultured skin fibroblasts or mRNAs extracted from such cells.

In the present study, we have developed protocols to amplify all 103 exons and exon boundaries of both the COL1A1 gene and the COL1A2 gene by PCR and to scan for mutations by conformation-sensitive gel electrophoresis (CSGE) (Ganguly et al. 1993; Ganguly and Prockop 1995). To develop the protocols, it was necessary to complete 90% of the structure of the 38-kb COL1A2 gene and to resequence extensive regions of the 18-kb COL1A1 gene. We used the protocols to analyze the COL1A1 and COL1A2 genes in 15 patients

Table 1

Clinical Summary of OI Type I Phenotypes in Probands

Proband (Code)	Year of Birth	Fractures	Bone Deformity/ Short Stature[a]	Blue Sclerae[a]	Dentinogenesis Imperfecta[a]	Hearing Loss[a]	Family History[a]
1 (421)	1970	Multiple	−	ND	ND	−	−
2 (262	1964	2	−	+	−	−	+
3 (207)	1963	6	−	+	+	−	+
4 (395)	1974	>100	−	+	+	+	+
5 (491)	1990	Multiple	−	+	?	−	+
6 (185)	1974	12	−	+	−	−	+
7 (292)	1967	Multiple	−	ND	−	−	+
8 (394)	1952	Multiple	−	+	−	−	+
9 (283)	1956	Multiple	−	ND	?	−	+
10 (286)	1989	Multiple	−	+	±	−	+
11 (LE)	1963	Multiple	−	+	ND	−	+
12 (ROMA)	1980	Multiple	−	+	−	−	ND
13 (J198)	1970	Multiple	−	+	−	ND	−
14 (J210)	1963	Multiple	ND[a]	+	ND	ND	+
15 (J218)	1976	Multiple	−	+	±	−	+

[a] A plus sign (+) denotes presence; a minus sign (−) denotes absence; and a plus-or-minus sign (±) denotes possible presence. ND = not definitively defined.

with mild OI. Disease-causing mutations were found in 13 of the 15 patients, and two probable disease-causing mutations were found in the remaining 2 patients. Analysis of the data, together with analysis of previously published mutations causing OI, indicate that there are common sequences for mutations that produce null alleles of the COL1A1 gene.

Subjects and Methods

Subjects

All probands presented the typical phenotype of OI type I. The major clinical manifestations for each of them are summarized in table 1.

Protein Analysis

A skin biopsy was obtained from probands 1–10, and fibroblast cultures were established under standard conditions. Cells were seeded at 35,000 cells/cm^2. Labeling of the fibroblasts and purification of the collagen molecules were performed as reported elsewhere (Nuytinck et al., in press). In brief, for labeling of the cells, Basal Medium Eagle (Life Technologies) supplemented with 1 μCi ^{14}C proline/ml, 5% dialyzed FCS, 0.05 mg βAPN/ml, and 0.025 mg ascorbic acid/liter was added. After 20 h, the medium was removed and supplemented with proteinase inhibitors (0.1 mg phenylmethyl sulfonyl flouride/ml, 0.1 mg N-ethylmaleimide/ml, and 2 mM EDTA pH 7.5). The cell layer was trypsinized, and the cells were collected by centrifugation and were lysed in 0.5% Triton X-100 in 0.5 M acetic acid. The supernatant was used for collagen analysis.

Collagen samples were isolated from the medium by alcohol precipitation and were redissolved in 0.5 M acetic acid. For the conversion of procollagen to collagen, the samples were digested with 50 μg pepsin/μl (Boehringer) for 4 h at 15°C. The digestion was stopped by the addition of 0.5 μg pepstatin/μl (Boehringer). SDS-electrophoresis was performed by the Laemmli (1975) system, with 3% stacking and 5% separation gel. Prior to being loaded, the samples were lyophilized, redissolved in sample buffer (Tris-HCl pH 6.8, 2 M urea, and 0.04% bromophenol blue), and denatured for 20 min at 55°C. Electrophoresis was performed at 8°C overnight (3.5 V/cm). The gels were processed for fluorography by use of 20% 2.5-diphenyloxazole in 100% acetic acid and then were dried and exposed to a hyperfilm MP (Amersham).

Analysis of Polymorphisms in cDNAs and Genomic DNAs

Total RNA was isolated from cultured skin fibroblasts by Trizol (Life Technologies). Prior to cDNA synthesis, RNA samples were treated with RNase-free DNase (Life Technologies) to avoid genomic DNA contamination in the reverse transcriptase–PCR experiments. For the conversion to cDNA, Moloney murine leukemia virus reverse transcriptase was used in combination with random hexanucleotide primers. For detection of the *Mnl*I polymorphisms in the COL1A1 gene, the primers and conditions used were those given by Sokolov et al. (1991). After enzymatic digestion, the fragments were evaluated either by agarose electrophoresis or by sepa-

ration on polyacrylamide gels by automated laser detection (ALF; Pharmacia).

For the detection of the 4-bp insertion polymorphism in the 3′ end of the COL1A1 gene (Nuytinck et al., in press), the following primers were used: 5′-CCT TTC TGC TCC TTT CTC CA (sense primer) and 5′-AGC AAC ACA GTT ACA CAA GG (antisense primer). Approximately 500 ng of genomic DNA was amplified under the following conditions: 94°C for 1 min, 56°C for 1 min, and 72°C for 1 min, for 25 cycles. The products were separated on a 6% polyacrylamide gel by an automated laser fluorescent DNA sequencer (ALF; Pharmacia). A fragment of 430 bp (allele A1) and/or a fragment of 434 bp (allele A2) was obtained by use of size markers for correct fragment-length evaluation.

Defining the Consensus Sequences of the 5′ Half of the COL1A1 Gene

To define the consensus sequences, genomic DNAs from eight unrelated probands with OI were used as separate templates for PCR with primers based on published sequences of the COL1A1 gene (Chu et al. 1985; D'Alessio et al. 1988; Määttä et al. 1991; Westerhausen et al. 1991). The PCR reactions were performed with a commercial DNA polymerase (AmpliTaq® Gold; Perkin-Elmer) in a 40-μl volume, with thermal cycling at 95°C for 10 min, for one cycle, followed by 95°C for 40 s, 60°C for 40 s, and 72°C for 50 s, for 35 cycles. The PCR products ranged in size from 1 kb to 2.5 kb. Sequences were defined by automated sequencing (ABI PRISM™ 377 Sequencer; Perkin-Elmer; and ABI PRISM Dye Therminator Cycle Sequencing Ready Kit, with AmpliTaq DNA polymerase, FS; Perkin-Elmer). Prior to the sequencing, the samples were treated with exonuclease I, to degrade the residual PCR primers, and with shrimp alkaline phosphatase, to dephosphorylate the residual nucleotides (Hanke and Wink 1994; Werle et al. 1994).

Completing the Structure of the Human COL1A2 Gene

To sequence the 5′ end of the human COL1A2 gene, a 14-kb fragment spanning introns 1–21 was obtained from an *Eco*RI/*Eco*RI genomic fragment that had previously been cloned into bacteriophage in the course of the definition of one mutation that had been shown to cause OI (Vasan et al. 1991). The fragment was broken randomly by sonication, and the fragments were size separated by gel electrophoresis. The selected fragments were subcloned into a plasmid (pUC18), and 60 clones were isolated. Plasmid DNA from 13 color-selected colonies indicated that the inserts ranged in size from 500 to 4,000 bp. Approximately 90% of the sequences were recovered by shotgun sequencing of the clones. The remaining gaps were closed by sequencing the original bacteriophage clone containing the 14-kb genomic fragment, by manual or automated procedures.

To complete the structure of the COL1A2 gene, a genomic P1 clone containing the complete human gene was obtained by PCR screening of human P1 library (Genome Systems). Screening was done by PCR amplification using primers designed on the basis of the published human cDNA sequences (de Wet et al. 1987; Kuivaniemi et al. 1988). The PCR primers were C1PF1 (5′-GTA CAT TTC CTA GAG AAC TTG) and C1PR1 (5′-CTA CTC TCA GCC CAG GAG GTC CTG), corresponding to sequences in intron 19 and exon 21/intron 21. Three positive clones were obtained: DMPC-HFF1 1250-E2, GS control 7403; DMPC-HFF1 1365-B1, GS control 7404; and DMPC-HFF1 1473-F6, GC control 7405. Because P1 clone 7407 was found to contain the entire coding sequences of the human COL1A2 gene, the clone was selected for detailed characterization of the gene.

To increase the yield of DNA, P1 clone 7407 was transferred from *Escherichia coli* strain NS3529 to *E. coli* strain NS3516 via transduction, as suggested by Genome Systems. The P1 plasmid DNA was isolated by the method of Birnboim and Doly (1979), with modifications suggested by Genome Systems. The isolated DNA was dissolved in water and was further purified, by spot dialysis with a membrane (VSWP 02500; Millipore), against water.

Nucleotide sequencing was performed by cycle sequencing of the P1 clone (dsDNA Cycle Sequencing System; Life Technologies; and Cycle Sequencing Kit; Pharmacia Biotech). Sequencing primers for the COL1A2 gene were designed on the basis of published cDNA and genomic sequences (Bernard et al. 1983; de Wet et al. 1987; Kuivaniemi et al. 1988). Additional primers were designed on the basis of sequences obtained during the study. The 5′ end of the primer was labeled with T4 polynucleotide kinase (U.S. Biochemical) and [γ-^{33}P]ATP (DuPont NEN). For cycle sequencing, 0.5–2.0 μg of P1 DNA was used as template, and thermal cycling was performed with a commercial instrument (either GeneAmp 9600; Perkin Elmer; or PTC 225 DNA Engine Tetrad; MJ-Research). Some of the nucleotide sequences were obtained by the Terminator Cycle Sequencing Ready Reaction kit and ABI PRISM 377 DNA Sequencer (Perkin-Elmer). The data were analyzed by means of the Wisconsin Sequence Analysis Package versions 8.0 and 8.1 UNIX (Genetics Computer Group) and the Editseq program from the Lasergene software package (DNAStar).

Mutation Analysis

Genomic DNA was extracted from blood samples or from cultured skin fibroblasts. The exons and the flank-

ing sequences of the 51 exons of the COL1A1 gene and the 52 exons of the COL1A2 gene were amplified by a series of specific primers (table 2). Genomic DNA was amplified in a 40-μl volume, by thermal cycling at 95°C for 10 min, for one cycle, followed by 95°C for 40 s, 60°C for 40 s, and 72°C for 50 s, for 35 cycles. This was followed by heteroduplex formation steps: 95°C for 5 min and 68°C for 30 min. CSGE analysis was performed as described elsewhere, with the precaution that the taurine buffer was not autoclaved (Ganguly et al. 1993; Ganguly and Prockop 1995). The gels were examined with a hand-held UV illuminator (short wave), to identify regions containing bands of homoduplexes and heteroduplexes. Appropriate regions were cut from the gels, transferred to a filter paper, and analyzed further in a UV-image analyzer with a CCD camera (DOC-IT™ Gel Documentation System; UVP). The image from the monitor was recorded by a thermal printer. PCR products containing heteroduplexes were sequenced either manually (Sequenase PCR Product Sequencing Kit; USB) or by automated instrument (ABI PRISM 377 Sequencer; Perkin-Elmer; and ABI PRISM Dye Therminator Cycle Sequencing Ready Kit with AmpliTaq DNA polymerase, FS; Perkin-Elmer) after treatment with exonuclease I and shrimp alkaline phosphatase (Hanke and Wink 1994; Werle et al. 1994). PCR products that contained deletions in one allele were purified from agarose (QIAEX II Gel Extraction Kit; Qiagen). Approximately 60 ng of purified product was cloned into a plasmid (pT7 Blue T-Vector Kit; Novagen) and was sequenced.

Results

Detection of Mutations by Assays of Protein and mRNAs

In initial experiments in the present study, fibroblasts from 10 probands with type I OI were assayed for the ratio of newly synthesized chains of type I and type III procollagen. Also, the fibroblasts were assayed for the presence of polymorphic markers in mRNA that reflect expression of both alleles of the COL1A1 gene. As is shown in table 3, a reduced ratio of newly synthesized type I to type III procollagen was found in 7 of the 10 probands, and a slight reduction was seen in an 8th proband (probands 1–8); the ratio was normal in fibroblasts from 2 of the probands (probands 9 and 10). An assay for a polymorphic base in the mRNA indicated that one COL1A1 allele either was not expressed or was expressed at a reduced level in fibroblasts from five probands (probands 1–4 and 9), including one in whom the protein assays were normal (proband 9). The mRNA assay indicated that both alleles were expressed in one proband with a decreased ratio of type I to type III collagen (proband 8). The mRNA assays for four probands were noninformative.

Sequences of the COL1A1 and COL1A2 Genes

To develop PCR primers for analysis of the COL1A1 and COL1A2 genes, new sequences were defined. In the case of the COL1A1 gene, several groups contributed to the complete structure of the gene (Chu et al. 1985; D'Alessio et al. 1988; Määttä et al. 1991; Westerhausen et al. 1991). However, in the present study, our attempts to develop PCR primers based on published sequences of the 5′ half of the COL1A1 gene were largely unsuccessful; apparently the published sequences were not from the most frequent alleles of the gene. Therefore, about 12 kb of the COL1A1 gene were re-sequenced from genomic DNA of 8 individual probands. A new consensus sequence was developed that in some regions differed considerably from the previously published sequences. For example, comparison of 3.4 kb of sequences from introns 6–24 indicated that 9.6% of the nucleotides differed. Except for 13 single-base polymorphisms in the introns, the same consensus sequence was found in 16 alleles from eight unrelated individuals. The revised consensus sequence of the complete 18-kb COL1A1 gene have been submitted to GenBank (accession number AF017178).

In the case of the COL1A2 gene, previous publications provided the complete cDNA of ~5 kb coding for the proα2(I) chain (Bernard et al. 1983; de Wet et al. 1987; Kuivaniemi et al. 1988), ~4 kb of genomic sequences that included ~400 bp of the promoter and 5′ UTR, and 3.5 kb of introns (Myers et al. 1983; Dickson et al. 1984, 1985; Kuivaniemi et al. 1988; Tromp and Prockop 1988; Sherwood et al. 1990; Ganguly et al. 1991; Vasan et al. 1991). In the present study, we sequenced >30 kb to provide the complete sequence of the 38-kb gene. Approximately 11 kb of new sequences from the 5′ end of the gene were obtained from a 14-kb *Eco*RI/*Eco*RI genomic fragment (Vasan et al. 1991). An additional 19 kb of sequences of the COL1A2 gene were obtained from P1 clone 7407. The results defined all the intron sizes and sequences (fig. 1). The sequences of the COL1A2 gene have been submitted to GenBank (accession number AF004877).

The intron sizes of the human COL1A2 gene were approximately the same as the previously defined introns of the corresponding 30-kb gene from chick (Boedtker et al. 1985). As noted elsewhere, there were major differences between the intron sizes of the human and chick COL2A1 genes (Ala-Kokko et al. 1995).

Table 2

Oligonucleotide Primers for PCR Amplification of Promoter Regions, 103 Exons, and Flanking Sequences and Polyadenylation Signals of the COL1A1 And COL1A2 Genes

	COL1A1			COL1A2		
Region	Sequence	Position	Size (bp)	Sequence	Position	Size (bp)
Promoter:						
5′	atcctgaggacccagctgcac	−387	344	gccacgtcccttcccccattc	−312	340
3′	gttagcgtccgctcatgcgtg	−44		aagtccgcgtatccacaaagctgagcat	28	
Exon:						
1:						
5′	gacgggagtttctcctcggggtc	−115	321	agttggaggtactggccacgactg	−108	315
3′	gagtctccggatcatccacgtc	103		gcgttttcccacatgcctgag	137	
2:						
5′	gctgatgaggagcaggcgag	−161	333	tgctgatccctgccatacttttgac	−189	280
	atccaagtgtgcctcttagac	−90		tcttcccttccaagagaagacatc	80	
3′	gtttgctaatgctgctcccgtc	108				
3:						
5′	gctggaggcctctgccgacgggagcagc	−71	240	gtgaaggtatatttgtatactacac	−97	279
3′	ggcctcgggggccagtgtctc	133		tgttatcttaaacatcaaagctac	167	
4:						
5′	gcctctgccgacgggagcagc	−201	372	cattgtagttacatcagtcttacc	−112	294
3′	aggctgtccagggatgccatc	135		gcttcttctgcagtgcattacctg	146	
5:						
5′	acctggcctcttgtttcttctc	−162	386	tccaccctacttgcacatagaaagg	−110	385
3′	ctgtaggattcttgcaacttttct	122		gacaagggctcacaaagagaatggg	182	
6:						
5′	cacaccaggaagtgcatgatgtcag	−99	261	tcggccaagtttttgacgtacagct	−208	338
3′	ctcccaagctgtctataccagccgc	90		tggcgtggtaaaatgtgacataaaa	76	
7:						
5′	atacgcggctggtatagacag	−166	300	gggaggaataaaaactatggaatc	−125	316
3′	tctctgagcatctctcctgccctca	89		gaccagcttcaccaggctcac	146	
8:						
5′	tggagggaagactgggatgag	−111	247	gtactgaaagcttgtaatgcctc	−88	223
3′	aagacccaggcctgggagttcttct	82		ggagacccatcatttcactaagg	81	
9:						
5′	cccctggtgagcctggcgag	−194	347	tgaacctggtcaaactgtgagtac	−109	273
3′	ctgagtatcgttcccaaatgtg	97		tgtcaggcatattcagcttttggca	110	
10:						
5′	ctggggccccccaaacctgacctgc	−120	262	accaagattcccccatttgtgctga	−70	344
3′	ggccattagaacacactcactg	88		gttgcttatggtatgcttgctgtc	220	
11:						
5′	ctgaacctgggcttcactgcac	−98	292	tttgtcgctctgtgcttagagg	−143	321
3′	gatgtccactctctggcccttg	140		cttccctttggcaaactccagggat	124	
12:						
5′	caaagggatggcggtgatgac	−111	253	gctgggacctggaacactggacttc	−145	260
3′	ctgtagatcagagaataatgag	88		tggaggtcatggggaatttcaatca	61	
13:						
5′	gtaagaggctgtctgaacatc	−88	228	gaacctggatatgtggtactatctg	−212	361
3′	gtcagatgagatgggagacagc	95		gaatacaatgctgaaggatacagtg	104	

(*continued*)

Table 2 (continued)

Region	COL1A1 Sequence	COL1A1 Position	COL1A1 Size (bp)	COL1A2 Sequence	COL1A2 Position	COL1A2 Size (bp)
14:						
5′	ggtgagtgtgcccagttccag	−117	263	cttgtacaggttggaaactgaac	−94	258
3′	cgttaagtccactgagcactg	92		ccacgggcaccctaagaaga	110	
15:						
5′	gatccctgagctctggaaggggctc	−83	279	ccgtgggcttcctggtgagag	−148	298
3′	gagatggcagctgcaagtcac	145		tggtaaagtgtctgaaatgatgc	105	
16:						
5′	gggcgaggttatgttggtctg	−102	229	caccctggataccatgaatgtc	−187	318
3′	tttggggaacagggagacatgaacc	73		ctgcaaacacagttccaatctttca	77	
17:						
5′	ctgatcattgctctcctgtccctgt	−103	320	cagtagccaagatggcagaatc	−191	462
3′	accaggctgtccatcagcac	118		ccagtaaggccgtttgctccag	172	
18:						
5′	taagtgtccccgactcagtgtc	−88	220	cgttggacctcctgtaagtag	−156	306
3′	agccagggcgtgacgtaggag	87		aaaatgcagtgtggtccattagg	105	
19:						
5′	aagtatcctgccaggcttcag	−101	319	taatgtgtgctgcctctacagc	−64	330
3′	gaagaggatgagctgagagtc	119		catatagcagacgggagtgtac	167	
20:						
5′	caagggtaacagcgtgagtac	−144	348	cttgagcttctctttaccttgac	−106	325
3′	tgaggctgggcctccagtgtc	150		caccactgggaccaggaggac	165	
21:						
5′	ggctctgaggctggcacaggatg	−65	292	cgtaagtagctctatcatcac	−126	363
3′	ggaaaccacggctaccaggtc	119		aaggcagatggaaagcagatg	129	
22:						
5′	ccggaccccctggcgagcgtg	−114	278	gggttgggtgaagtgttttggcttg	−135	268
3′	cacaggaacagttagggtctc	110		gaggatgctaaagctaatgacac	79	
23:						
5′	cccaaggtaacctctccttgc	−131	309	gctgtctatcacttacttcctag	−109	268
3′	gatccggaacgcctcatcccaagac	79		tcaaaaatgcaactgtcagcaagac	136	
24:						
5′	gtcttgggatgaggcgttccggatc	−110	283	aaaaagtcgggggaaaaggtgcctt	−101	349
3′	gtccggggcgaccatcttgac	119		tctcccctgctctgctttcagtcct	194	
25:						
5′	gccctggcagccctggtcctg	−129	368	tttcatccgtggcagcatcataagc	−81	276
3′	tagggaggctgaggtccagaaagtg	140		ctgagactggactgattcgcag	96	
26:						
5′	agggcccagcaagaagcacctgc	−160	309	tggagctgcatggtgatggatc	−165	298
3′	gctgaggaccgtggcctctagc	95		tatcagatggtgtaaaaaaaaaagtgtggttcttagatg	79	
27:						
5′	cctgcaggaggggtgctagag	−83	239	gctttcgtgggaacccacaatgagt	−139	315
3′	cacagagagaacactacagtcac	100		tagcaacgtatgtcaccactg	122	

(*continued*)

Table 2 (continued)

Region	COL1A1 Sequence	COL1A1 Position	COL1A1 Size (bp)	COL1A2 Sequence	COL1A2 Position	COL1A2 Size (bp)
28:						
5′	ctgctgtgagtgtccctgatg	−108	247	tggccatctccattttcagtc	−92	257
3′	ggagggaaggtttagaatctg	85		tgcttcagtcctgaaatcatgt	111	
29:						
5′	ggtgaggcctcatggctgtc	−112	251	gagctgtaaatcaccataccgtac	−115	333
3′	tggctgtctgattagctaggaggcgg	85		tggctcattctctccatcagcac	164	
30:						
5′	gggttcctctctaatcacggccagac	−110	246	tgcactcatgtagatactgccaggt	−121	264
3′	agaagggaaggacagggcatgtgaag	91		gacttgttgcagggtcatcagtggc	100	
31:						
5′	cctctggagcaagagtaagtag	−107	318	aaaccagggctcggaagctacacaa	−102	359
3′	accccacaccctatctccatg	112		ggtccactggaatcggattgctgtt	158	
32:						
5′	tttctcaaggcttgtcgttggccttg	−120	291	tctccctcctttcaatagcccagcc	−152	336
3′	gattcaaaggaggcagagatgggagc	63		gtgaaaacttgggcatccttgtgca	75	
33:						
5′	cctctcaggaaacccagacacaagca[a]	−87	318	gaatggtaaggaatcgagacattgc	−148	276
3′				aatttggaaaattctcaattcaacataaaaaaaaaatccaagtacgaag	74	
34:						
5′				ggagtaccctccttctgagagtggc	−198	330
3′	gttcccaggttgacagctcag[a]	123		attgctggggctctttgggactagg	78	
35:						
5′	gtcctgccaaactgagctgtc	−129	327	actctgtgagatgtgcgtcag	−137	320
3′	attggagagatgcgtctgacaggagg	144		gttggggccagcaggaccgac	140	
36:						
5′	ccctgtctgtgccttcacccccttgc	−97	249	gaagccctgtaagtaagaacctg	−112	246
3′	cttctcccctgaggatggctgac	98		gttacagctctggtattccgac	80	
37:						
5′	tgcctccattactgctcctcc	−81	276	gatttgctggtccggctgtgag	−113	348
3′	tgtaggagagcacagacgcatcaagc	87		cttccgttattttccatcttctatc	127	
38:						
5′	tgagtggcttggccctctgtg	−97	240	tgcgggaatgatccacttgaagaaa	−85	283
3′	agagggagaacagccaactcatccg	99		tcggaattgctctgaatagaatgaa	144	
39:						
5′	gagtatcacccgcctctctgttgagc	−124	259	gaaattcccatcttacccaaattcttg	−70	263
3′	tcagtcagccccaccatccttctg	81		gaaaagctgacttcagaccaggag	139	
40:						
5′	gtgggggctgccagaaggatg	−94	337	cataaaggaagacaggagttgc	−192	447
3′	tgaggtgccagacagcagcacag	81		catcaacaaatagatgccacttg	93	
41:						
5′	agtgccagctcagatctctgcagctc	−98	309	ctcacaatcttcaagccaacctgtg	−75	258
3′	gtccgctggagtcatctctac	103		tctgtcacatttgaagtggcagctt	75	

(*continued*)

Table 2 (continued)

	COL1A1			COL1A2		
Region	Sequence	Position	Size (bp)	Sequence	Position	Size (bp)
42:						
5′	gagaacagatttggtagagatgac	−84	329	ggaggggaaggttagcattccatcg	−64	351
3′	caggggaaccttcggcaccag	135		aaagcccattctttggcctaagcaa	179	
43:						
5′	cccatgccagtaccctcagcatggc	−90	242	agggttcgttactgagcactg	−110	323
3′	gggagagcaggggaatatgggtcag	98		ccacggggccatgaggaccag	159	
44:						
5′	gcaacactccatgaccacagc	−96	304	atggtcaacccggacacaag	−106	331
3′	cctgcctgggtgaagtccgac	100		caacttagctaggcccaagatac	117	
45:						
5′	ggagagagagatccagcagagggga	−94	246	caacccagattgatgctaagcttc	−288	480
3′	gggacaaactgtcaggcggaagttc	98		gtatcaattctcagcatggactg	138	
46:						
5′	catgccttcagaactctacag	−90	297	gcagattaccagcagaggtgagagc	−145	320
3′	ggggaaagaatgactatccag	99		tgaaaatccttctgagctgaaggcc	67	
47:						
5′	gttgcccacactgcccttgtc	−92	249	tcccattgaatttggaaaaaaaaaaaaatatgtctcttgac	−81	257
3′	aacccttctccagagaggcaaaggg	103		gacaccaggtacatgtgagctg	114	
48:						
5′	ccgtggggccagagccagcag	−102	299	caagagaagacagttcatctctg	−115	326
3′	gcacagagagggaagagagtgggga	89		tggggctaactttaatgggttgtc	103	
49:						
5′	gctggtcctgttgtatgtagc	−103	475	gaacatgcttccgtgtgaagctc	−102	535
3′	ccagcaccatatggtaggggcacat	89		agggaaatgaggttgggtgctggtt	177	
50:						
5′	ccagggtccccatgcccatatgtgc	−87	344	tgttacttatgagagtcagtatctttc	−98	459
3′	catgtcccttctgagcactgggcta	66		gtccaatcaatccatcttctaatgtg	176	
51:						
5′	ggaccctggacaggaaggccagcagg	−74	399	cccttttcctaagcttggatctgag	−132	471
3′	gatggagagagggcactatggc	82		ttaaccccccttagacccccccttg	96	
52:						
5′	gggctttttggccaggccatagtgcc	−84	319	ggacagacatcttcagaatgac	−120	401
3′	gagggggttcagtttgggttgcttgtctg	88		ttgcccacaatttaagcaagtag	134	
Polyadenylation signal:						
A1:						
5′	tttttcctttgcattcatctctc	−138	303	ccttccatttcttctgcacatctac	−91	323
3′	cattgtttcctgtgtcttctgg	440		tgtggatcacactcatgggaaagtg	413	
A2:						
5′	agagacaacttcccaaagcac	−1256	217	gttcataatacaaaggtgctaat	−525	273
3′	aggccccttccccatgtctac	1472		gaaacaaagcttctgtggaacc	737	

[a] Exons 33 and 34 were fused for the COL1A1 gene.

Table 3

Mutations Detected in OI Type I

Proband (Code)	Reduced Ratio, Procollagen I: Procollagen III[a]	Absence of COL1A1 Allele in mRNA[b]	Mutation in COL1A
1 (421)	+	+	A^{+3}IVS22→G
2 (262)	+	+	G^{-12}IVS20→A
3 (207)	+	+	A^{-2}IVS5→G
4 (395)	+	+	del T nt 927 (E12)
5 (491)	+	NI	Arg^{183}→stop
6 (185)	+	NI	del T nt 2192 (E31)
7 (292)	±	NI	del G nt 3198 (E43)
8 (394)	+	0	G^{+1}IVS12→A
9 (283)	0	+	del T nt 2732 (E38)
10 (286)	0	NI	Arg^{42}→stop
11 (LE)	NA	NA	ins C nt 1787 (E24)
12 (ROMA)	NA	NA	Arg^{519}→stop
13 (J198)	NA	NA	ins AC nt 1838 (E25)
14 (J210)	NA	NA	G^{-1}IVS25→A
15 (J218)	NA	NA	G^{-1}IVS18→A

[a] A plus sign (+) denotes presence; a minus sign (−) denotes absence; and a plus-or-minus sign (±) denotes possible presence. NA = not assayed.

[b] A plus sign (+) denotes presence. NI = noninformative assay (because proband was homozygous for polymorphism); and NA = not assayed.

Primers for PCR Amplification of the Exons and Exon Boundaries

The sequences of the human COL1A2 gene that have been defined in the present study, together with the previously published sequences of both the COL1A1 gene and the COL1A2 gene, were used to design primers for amplification of each of the 103 exons and of ⩾80 bp of the 5′ and 3′ flanking sequences (table 2). Conditions for PCR amplification were optimized so that each PCR product gave a discrete band detectable by agarose gel electrophoresis and by PAGE. In addition, the primers were designed so that the PCR products were not >500 bp, so that the presence of single-base substitutions could be detected by heteroduplex analysis using CSGE (Ganguly and Prockop 1995; J. Körkkö, S. Annunen, D. J. Prockop, and L. Ala-Kokko, unpublished data).

Detection of Mutations in OI Type I by PCR and CSGE Scanning

The new protocols were used to search for mutations in the initial 10 probands with type OI and in 5 additional probands (table 1). Twenty-five polymorphisms in the COL1A1 gene and 18 polymorphisms in the COL1A2 gene were detected with frequencies ranging between .01 and ∼.5 (not shown). Most of the polymorphisms were found within the intron sequences analyzed. In addition, the results detected the presence of unique heteroduplexes in each of the 15 probands (see fig. 2). Sequencing of the PCR products that generated unique heteroduplexes identified 13 mutations. Five of the mutations were single-base deletions or insertions (probands 4, 6, 7, 9, and 11; table 1), one mutation was a 2-bp insertion (proband 13), and three of the mutations were single-base substitutions that converted a co-

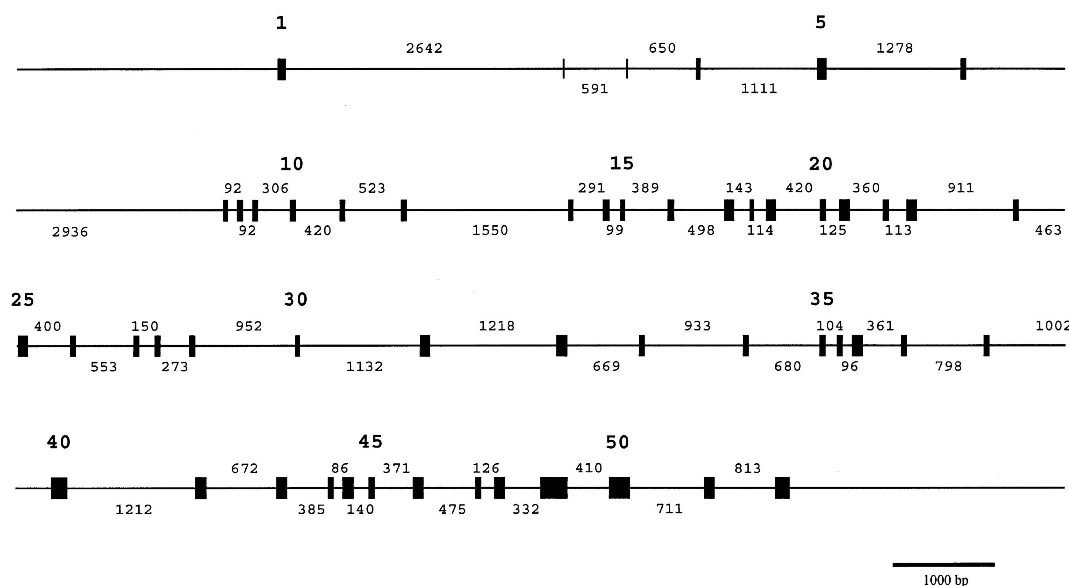

Figure 1 Schematic representation of the genomic organization of the human COL1A2 gene. The exons and introns are drawn to scale; the exact sizes are given for the introns.

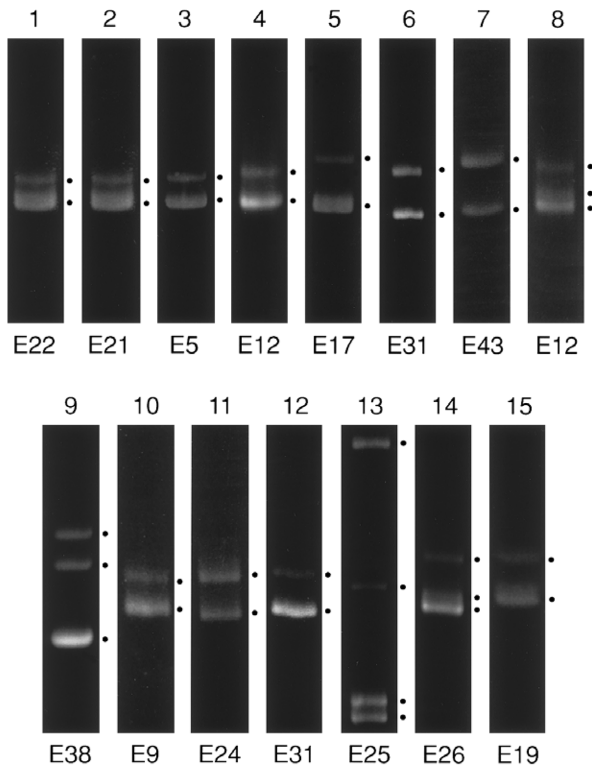

Figure 2 CSGE assays of PCR products for heteroduplexes reflecting the presence of single-base changes in the COL1A1 gene. Samples are from probands described in tables 1 and 2. Dots indicate homoduplex and heteroduplex bands that are readily visible by UV illumination of the gels but that are difficult to reproduce by the image provided by the thermal printer of the UV-image analyzer.

don for arginine to a premature-termination codon (probands 5, 10, and 12). Four mutations altered consensus sites of RNA splicing in the first two or last two bases of an intron (probands 3, 8, 14, and 15). Two additional changes in the two remaining probands were also likely to alter RNA splicing, one at position +3 of intron 22 (proband 1) and the other at position −12 of intron 20 (proband 2). Although the bases at these two positions are not conserved in all introns (Nakai and Sakamoto 1994), the mutations were not found in 100 other COL1A1 alleles. Also, no other changes were found in the analysis of all the other PCR products from the COL1A1 and COL1A2 genes of these two patients. In addition, the protein data and the mRNA data both indicated that one COL1A1 allele was not expressed at a normal level in both probands. Two other observations supported the conclusion that the single-base change at position −12 of intron 20 in proband 2 was a disease-causing mutation. One observation was that the same change was found in the proband's affected father and sister but not in his unaffected brother, and the second observation was that the single-base change at −12 produced a 3′ consensus sequence for RNA splicing—that is, it converted a sequence of $(Pyr)_{11}$CGG to $(Pyr)_{11}$CAG. In summary, definitive disease-causing mutations were found in 13 probands, and probable disease-causing mutations were found in the other 2 probands. As indicated in table 3, all the mutations detected were in the COL1A1 gene.

Identification of Common Sequences for Mutations Causing COL1A1 Null Alleles

Comparison between the data generated in the present study and those reported for previously defined mutations in OI identified common sequences for null-allele mutations in the COL1A1 gene (Willing et al. 1994, 1996; Redford-Badwal et al. 1996; Körkkö et al. 1997). For the sake of comparison, we first considered single-base substitutions that converted a codon for arginine, CGA, to a premature-termination codon, TGA. There were nine mutations that met this criterion (table 4). All were found in the sequence context of C/GCC <u>C</u>GA GG/T in the COL1A1 gene. They were found in five of the six such sequences in the wild-type COL1A1 gene. In contrast, no such mutations have yet been reported that convert codon of CGA for arginine in seven other sequence contexts of the COL1A1 gene—that is, in which the CGA is not preceded by G/CCC.

A second category of mutations were mutations that created frameshifts and premature-termination codons because of single-base deletions or insertions. A total of 14 such mutations have been reported in the COL1A1 gene (table 5). Nine of the mutations were in the sequence context of CCC CCT. Only 5 of these 14 mutations were found in other sequences. In the COL2A1 gene, six such mutations were reported, and two are in the sequence context of CCC CCT (table 5).

Discussion

Assays of procollagens and mRNAs in cultured skin fibroblasts have been useful in the definition of mutations that cause OI. However, as illustrated here, the assays do not detect all mutations, particularly in OI type I. Also, the assays require biopsies of skin and culture of fibroblasts for ⩾3 wk. For these reasons, we developed protocols for PCR amplification of the exon and exon boundaries of all 103 exons of the COL1A1 and COL1A2 genes.

To develop PCR-amplification primers directly from genomic DNA, it was necessary to perform extensive resequencing of the 5′ half of the COL1A1 gene and to sequence 30 kb of the 38-kb COL1A2 gene. Comparison between our data for the human COL1A2 gene and

Table 4

Common Sequences in COL1A1 for Mutation of CGA Codons to Premature-Translation-Termination Codons

Sequence Context	No. of Mutations	Exon Reported	Amino Acid Number	Reference(s)
CCC CGA GG	1	9	42	Present study
CCC CGA GG	1	17	183	Present study
GCC CGA GG	2	19	237	Redford Badwal et al. (1996); Willing et al. (1996)
CCC CGA GG	2	31	519	Willing et al. (1996); present study
CCC CGA gt	3	47	963	Willing et al. (1994); Körkkö et al. (1997)
GAC CGA GA	0	2		
CCC CGA GG	0	4		
GGC CGA GA	0	5		
GCT CGA GG	0	11		
AAG CGA GG	0	21		
GAG CGA GG	0	26		
GGC CGA GT	0	39		
GGA CGA GA	0	43		

previously reported data on the chick COL1A2 gene (Boedtker et al. 1985) indicate that the intron sizes are conserved. In contrast, the intron sizes of the COL2A1 gene are not conserved (Ala-Kokko et al. 1995). The primers for PCR amplification of the exons were developed so that the PCR products could be scanned, by CSGE, for single-base differences and other mutations (Ganguly et al. 1993; Ganguly and Prockop 1995). The availability of the scanning technique greatly reduced the amount of DNA sequencing necessary to search for mutations in the genes of patients with OI.

With the new protocols, definitive disease-causing mutations were found in 13 probands with type I OI, including 2 probands in whom protein assays were negative and 1 proband in whom the RNA assay was not informative. In addition, probable disease-causing mutations were found in the remaining 2 probands.

One of the initial hypotheses informing the present study was that probands in whom no mutations were found in the COL1A1 gene were likely to have mutations in the COL1A2 gene. However, no mutations in the COL1A2 gene were found. Therefore, mutations in the COL1A2 gene appear to be a rare cause of OI type I, even though mutations in the gene can cause some severe variants of OI (see Kuivaniemi et al. 1997). Individuals and mice heterozygous for nonfunctional proα2(I) chains probably have decreased bone density (Dickson et al. 1984; Chipman et al. 1993; Saban and King 1996). Therefore, individuals with nonfunctional COL1A2 genes may have phenotypes that are milder than OI type I and that overlap with that of osteoporosis.

As suggested by previous reports (see Byers 1993; Kuivaniemi et al. 1997), most mutations found in patients with OI type I are mutations that introduce either premature-termination codons for translation or aberrant RNA splicing and that thereby reduce the expression of the COL1A1 gene. Analysis of such mutations detected in the present study and reported previously indicates that they tend to occur in common sequence contexts. Of a total of nine mutations that converted the arginine codon CGA to the premature-termination codon TGA, all were found in the sequence context of G/CCC CGA GG/T of the COL1A1 gene (Willing et al. 1994, 1996; Redford-Badwal et al. 1996; Körkkö et al. 1997). None were found in seven CGA codons for arginine in other sequence contexts of the COL1A1 gene. Therefore, the mutability of the CpG sequence (see Ollila et al. 1996; Körkkö et al. 1997) is apparently enhanced by the context. Ollila et al. (1996) have noted that CpG is more mutable in the context of YYCGRR—and especially in the context of YYCGRY—than in other contexts. The results of the present study agree with this suggestion. As noted here, the COL1A1 gene has six G/CCC CGA GG/T sequences, whereas the COL1A2 has none. A similar analysis has indicated that single-base insertions or deletions in collagen genes are frequently in the context of CCC CCT. As noted by Jego et al. (1993), such mu-

Table 5

Null-Allele Mutations Produced by Single-Nucleotide Insertions or Deletions

Gene	No. of CCC CCT Sequences	Reported Single-Base Insertions or Deletions: Total	Reported Single-Base Insertions or Deletions: In CCC CCT Sequence
COL1A1	30	14[a]	9
COL1A2	6	0	0
COL2A1	16	6[b]	2
COL3A1	7	0	0

[a] Source: Dalgleish (1997); present study.

[b] Source: Ahmad et al. (1995); Brown et al. (1995); Kuivaniemi et al. (1997); A. De Paepe (unpublished data).

tations are likely to occur in duplicated sequences. In the case of collagens, the sequence CCC CCT, coding for Pro-Pro or Pro-Hyp, is a common sequence in the COL1A1 gene, and 9 of the 14 reported single-base insertions or deletions in the COL1A1 gene are in such a sequence. As is seen in table 5, such sequences are less common in the COL1A2, COL2A1, and COL3A1 genes, but two of the six null alleles reported in the COL2A1 gene are in the sequences.

Acknowledgments

We are indebted to the following collaborating clinical geneticists and physicians who referred the patients: I. Liebaers, M. Bonduelle, M. Freund, Boris A. Kousseff, L. Van Malderghem, Algelo Serra, and M. N. Van Thienen. We would like to thank A. Harju, A. Jokinen, K. Wettinck, J.P. Renard, and M. Van Thielen for technical assistance. This work was partially funded by grants from the Academy of Finland, the University of Gent, and the Fund for Scientific Research, Flanders.

References

Ahmad NN, Dimascio J, Knowlton RG, Tasman WS (1995) Stickler syndrome: a mutation in the nonhelical 3′ end of type II procollagen gene. Arch Ophthalmol 113:1454–1457

Ala-Kokko L, Kvist A-P, Metsäranta M, Kivirikko KI, de Crombrugghe B, Prockop DJ, Vuorio E (1995) Conservation of the sizes of 53 introns and over 100 intronic sequences for the binding of common transcription factors in the human and mouse genes for type II procollagen (COL2A1). Biochem J 308:923–929

Bernard MP, Myers JC, Chu M-L, Ramirez F, Prockop DJ (1983) Structure of a cDNA for the proα2 chain of human type I procollagen: comparison with chick cDNA for proα2 (I) identifies structurally conserved features of the protein and the gene. Biochemistry 22:1139–1145

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513–1523

Boedtker H, Finer M, Aho S (1985) The structure of the chicken α2 collagen gene. Ann N Y Acad Sci 460:85–116

Brown DM, Vandenburgh K, Kimura AE, Weingeist TA, Sheffield VC, Stone EM (1995) Novel frameshift mutations in the procollagen 2 gene (COL2A1) associated with Stickler syndrome (hereditary arthro-ophthalmopathy). Hum Mol Genet 2:141–142

Byers PH (1993) Osteogenesis imperfecta. In: Royce PM, Steinmann B (eds) Connective tissue and its heritable disorders. Wiley-Liss, New York, pp 317–351

Chipman SD, Sweet HO, McBride DJ Jr, Davisson MT, Marks SC Jr, Shuldiner AR, Wenstrup RJ, et al (1993) Defective proα2(I) collagen synthesis in a recessive mutation in mice: a model of human osteogenesis imperfecta. Proc Natl Acad Sci USA 90:1701–1705

Chu M-L, de Wet W, Bernard M, Ramirez F (1985) Fine structural analysis of the human proα1(I) collagen gene: promoter structure, AluI repeats, and polymorphic transcripts. J Biol Chem 260:2315–2320

D'Alessio M, Bernard M, Pretorious P, de Wet W, Ramirez F (1988) Complete nucleotide sequence of the region encompassing the first twenty-five exons of the human proα1(I) collagen gene (COL1A1). Gene 67:105–115

Dalgleish R (1997) The human type I collagen mutation database. Nucleic Acids Res 25: 181–187

de Wet W, Bernard M, Benson-Chanda V, Chu M-L, Dickson L, Weil D, Ramirez F (1987) Organization of the human proα2(I) collagen gene. J Biol Chem 262:16032–16036

Dickson LA, de Wet W, Di Liberto M, Weil D, Ramirez F (1985) Analysis of the promoter region and the N-propeptide domain of the human pro alpha 2(I) collagen gene. Nucleic Acids Res 13:3427–3438

Dickson LA, Pihlajaniemi T, Deak S, Pope FM, Nicholls A, Prockop DJ, Myers JC (1984) Nuclease S1 mapping of a homozygous mutation in the carboxylpropeptide-coding region of the proα2(I) collagen gene in a patient with osteogenesis imperfecta. Proc Natl Acad Sci USA 81:4524–4528

Ganguly A, Baldwin CT, Strobel D, Conway D, Horton W, Prockop DJ (1991) Heterozygous mutation in the G^{+5} position of intron 33 of the proα2(I) gene (COL1A2) that causes aberrant RNA splicing and lethal osteogenesis imperfecta. J Biol Chem 266:12035–12040

Ganguly A, Prockop DJ (1995) Detection of mismatched bases in double stranded DNA by gel electrophoresis. Electrophoresis 16:1830–1835

Ganguly A, Rock M, Prockop DJ (1993) Conformation-sensitive gel electrophoresis for rapid detection of single base differences in double-stranded PCR products and DNA fragments: evidence for solvent-induced bends in DNA heteroduplexes. Proc Natl Acad Sci USA 90:10325–10329

Hanke M, Wink M (1994) Direct DNA sequencing of PCR-amplified vector inserts following enzymatic degradation of primer and dNTPs. Biotechniques 17:858–860

Jego N, Thomas G, Hamelin R (1993) Short direct repeats flanking deletions, and duplicating insertions in p53 gene in human cancers. Oncogene 8:209–213

Körkkö J, Kuivaniemi H, Paassilta P, Zhuang J, Tromp G, De Paepe A, Prockop DJ, et al (1997) Two new recurrent nucleotide mutations in the COL1A1 gene in four patients with osteogenesis imperfecta: about one-fifth are recurrent. Hum Mutat 9:148–156

Kuivaniemi H, Tromp G, Chu, M-L, Prockop DJ (1988) Structure of a full-length cDNA clone for the preproα2(I) chain of human type I procollagen: comparison with the chick gene confirms unusual patterns of gene conservation. Biochem J 252:633–640

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

Laemmli UK (1975) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685

Määttä A, Bornstein P, Penttinen RP (1991) Highly conserved sequences in the 3′-untranslated region of the COL1A1 gene bind cell-specific nuclear proteins. FEBS Lett 279:9–13

Myers JC, Dickson LA, de Wet WJ, Bernard MP, Chu M-L, DiLiberto M, Pepe G, et al (1983) Analysis of the 3′ end of the human proα2(I) collagen gene: utilization of multiple polyadenylation sites in cultured fibroblasts. J Biol Chem 258:10128–10135

Nakai K, Sakamoto H (1994) Construction of a novel database containing aberrant splicing mutations of mammalian genes. Gene 141:171–177

Nuytinck L, Coppin A, De Paepe A. A 4 bp insertion polymorphism in the 3′ UTR of the COL1A1 gene is highly informative for null-allele testing in patients with OI type I. Matrix Biol (in press)

Ollila J, Lappalainen T, Vihinen M (1996) Sequence specificity in CpG mutation hotspots. FEBS Lett 396:119–122

Prockop DJ (1990) Mutations that alter the primary structure of type I collagen: the perils of a system for generating large structures by the principle of nucleated growth. J Biol Chem 265:15349–15352

Prockop DJ, Kivirikko KI (1995) Collagens: molecular biology, diseases and potentials for therapy. Annu Rev Biochem 64: 403–434

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

Saban J, King D (1996) PCR genotyping of mutant mice. Biotechniques 21:190–192

Sherwood AL, Bottenus RE, Martzen MR, Bornstein P (1990) Structural and functional analysis of the first intron of the human α2(I) collagen encoding gene. Gene 89:238–244

Sokolov BP, Prytkov AN, Tromp G, Knowlton RG, Prockop DJ (1991) Exclusion of COL1A1, COL1A2, and COL3A1 genes as candidate genes for Ehlers-Danlos syndrome type I in one large family. Hum Genet 88:125–129

Sykes B, Ogilvie D, Wordsworth P, Wallis G, Mathew C, Beighton P, Nicholls A, et al (1990) Consistent linkage of dominantly inherited osteogenesis imperfecta to the type I collagen loci: COL1A1 and COL1A2. Am J Hum Genet 46: 293–307

Tromp G, Prockop DJ (1988) Single base mutation in the proα2(I) collagen gene that causes efficient splicing of RNA from exon 27 to exon 29 and synthesis of a shortened but in-frame proα2(I) chain. Proc Natl Acad Sci 85:5254–5258

Vasan NS, Kuivaniemi H, Vogel BE, Minor RR, Wootton JAM, Tromp G, Weksberg R, et al (1991) A mutation in the proα2(I) gene (COL1A2) for type I procollagen in Ehlers-Danlos syndrome type VII: evidence suggesting that skipping of exon 6 in RNA splicing may be a common cause of the phenotype. Am J Hum Genet 48:305–317

Werle E, Schneider C, Renner M, Volker M, Fiehn W (1994) Convenient single-step, one tube purification of PCR products for direct sequencing. Nucleic Acids Res 22:4354–4355

Westerhausen AI, Constantinou C, Pack M, Peng M, Hanning C, Olsen A, Prockop DJ (1991) Completion of the last half of the structure of the human gene for the proα1(I) chain of type I procollagen (COL1A1). Matrix 11:375–379

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

Willing MC, Pruchno CJ, Atkinson M, Byers PH (1992) Osteogenesis imperfecta type I is commonly due to a COL1A1 null allele of type I collagen. Am J Hum Genet 51:508–515