Am. J. Hum. Genet. 60:48–56, 1997

Duplication of Seven Exons in the Lysyl Hydroxylase Gene Is Associated with Longer Forms of a Repetitive Sequence within the Gene and Is a Common Cause for the Type VI Variant of Ehlers-Danlos Syndrome

Jari Heikkinen,[1] Tarja Toppinen,[1] Heather Yeowell,[3] Thomas Krieg,[4] Beat Steinmann,[5] Kari I. Kivirikko,[2] and Raili Myllylä[1]

[1]Biocenter and Department of Biochemistry, and [2]Biocenter, Collagen Research Unit and Department of Medical Biochemistry, University of Oulu, Oulu; [3]Division of Dermatology, Duke University Medical Center, Durham; [4]Department of Dermatology and Venerology, University of Cologne, Cologne; and [5]Division of Metabolic and Molecular Diseases, Department of Pediatrics, University of Zurich, Zurich

Summary

The type VI variant of the Ehlers-Danlos syndrome (EDS) is a recessively inherited connective tissue disorder which, in most families, is due to a deficiency in lysyl hydroxylase activity. We have recently characterized a homozygous duplication of 8.9 kb in the lysyl hydroxylase gene (PLOD) in two EDS VI families. The duplication is caused by a homologous recombination of Alu sequences in introns 9 and 16. Using PCR, we have analyzed 26 additional EDS VI families from various countries and found that 7 of them have this duplication. Our data has shown a frequency of 19.1% for this mutant allele among 35 EDS VI families studied by us so far. Our haplotype analysis shows a variation in the sequence of DNA region surrounding the duplication. There is an association between a particular allele size class, the long form, at the dinucleotide repeat within intron 16 and the duplication mutation in PLOD. Screening of a general population revealed one positive finding among 582 alleles tested. An abnormal sequence in exon 17 of the gene, which generated a stop codon in the exon sequence and aberrant mRNA processing, was responsible for the nonfunctionality of the other allele in one of the compound heterozygous patients.

Introduction

The type VI variant of the Ehlers-Danlos syndrome (EDS) is a recessively inherited connective tissue disorder characterized by muscular hypotonia, kyphoscoliosis, ocular manifestations such as ocular fragility, retinal detachment, and microcornea, joint hypermobility, skin hyperextensibility (Krane 1984; Steinmann et al. 1993; Byers 1995; Prockop and Kivirikko 1995), and osteoporosis (Steinmann et al. 1993). Cardiovascular complications are also found, and dissecting aneurysm of the aorta and rupture of large arteries have been described as a cause for death in some EDS VI patients (Beighton 1993; Steinmann et al. 1993; McKusick 1994). The syndrome has been divided into two subtypes: the type VIA variant is due to a deficiency in lysyl hydroxylase activity measured in cultured skin fibroblasts, whereas patients with the VIB variant have normal levels of this enzyme activity (Beighton 1993; Steinmann et al. 1993; McKusick 1994; Prockop and Kivirikko 1995). Intrafamilial variability of the clinical symptoms is considerable, as has been shown in sibling pairs (Steinmann et al. 1993). This suggests that mutations in the lysyl hydroxylase gene (PLOD) are not the only factors determining the clinical phenotype of the disease.

Lysyl hydroxylase (procollagen-lysine 5-dioxygenase; EC 1.14.11.4) catalyzes the formation of hydroxylysine in collagens and other proteins with collagen-like amino acid sequences by the hydroxylation of lysine residues in X-Lys-Gly sequences (Kivirikko and Myllylä 1980; Kivirikko et al. 1992). Recent cloning of chick (Myllylä et al. 1991) and human (Hautala et al. 1992) lysyl hydroxylase has made it possible to characterize mutations leading to the deficiency in lysyl hydroxylase activity in patients with EDS VIA. So far, only a few mutations that cause this syndrome have been analyzed at the genomic level. These include a homozygous point mutation creating a premature translational stop codon in two siblings (Hyland et al. 1992), an apparently homozygous duplication of seven exons that has been found in two unrelated families (Hautala et al. 1993; Pousi et al. 1994), and a point mutation and a triplet deletion in the two gene alleles in a compound heterozygous patient (Ha et al. 1994). In addition, we have preliminary data from a splicing mutation caused by an insertion of two thymidines at the 5′ splice site of an exon-intron boundary of PLOD in one patient (L. Pajunen, M. Suokas, T. Hautala, S. Kellokumpu, B. Tebbe, K. I. Kivirikko, and R. Myllylä, unpublished data).

Received June 4, 1996; accepted for publication October 24, 1996.

Address for correspondence and reprints: Dr. Raili Myllylä, Department of Biochemistry, University of Oulu, Linnanmaa, FIN-90570 Oulu, Finland. E-mail: rmyllyla@koivu.oulu.fi

0002-9297/97/6001-0009$02.00

The seven-exon duplication in PLOD is 8.9 kb in size and is caused by an Alu-Alu recombination in introns 9 and 16 of the gene (Hautala et al. 1993; Pousi et al. 1994). The duplication does not disturb the splicing of introns, and sequencing of the cDNA for lysyl hydroxylase from the affected cells revealed a lengthened mRNA containing the nucleotides for amino acids 326–585 duplicated in tandem (Hautala et al. 1993). The apparent homozygosity of this duplication in two unrelated families with nonconsanguineous parents suggested that this mutation might also be found in other EDS VI families. Although our initial screening of five other EDS VI families gave no positive finding for the duplication (Pousi et al. 1994), the presence of this duplication was recently suggested to be in other EDS VI patients (Marshall et al. 1994). In this paper, we describe an analysis of 26 additional EDS VI families for this duplication. From these 26 families, we have identified 7 new families with this mutation and have analyzed each case by chromosome-specific polymorphic markers. In four of the families, patients have the duplication in both gene alleles, whereas in the remaining three families, of which one is characterized in detail in this study, the patients are compound heterozygotes for the duplication. Our sequence analysis of the region of the duplication in PLOD indicates an association between a particular allele class at the dinucleotide repeat within intron 16 and the duplication mutation and may suggest an ancestral single mutational event for the duplication.

Patients, Material, and Methods

Patients

All probands display characteristic features of the disease, including skin and joint hyperextensibility, skin fragility, muscular hypotonia, microcornea, and kyphoscoliosis. The families in the study are unrelated. In two families (AR and TD), the parents of the proband are first cousins; in the other families, there is no known consanguinity between the parents. All available EDS VI families were screened, and each patient diagnosed with EDS VI and indicated to have a reduced level of lysyl hydroxylase activity was used in the study.

Cell Cultures

Control human skin fibroblasts and fibroblasts of the individuals from the affected families were cultured at 37°C in Dulbecco's modified Eagle's medium (Gibco) under 5% CO_2. The medium was supplemented with 10% (v/v) FCS (Gibco) and ascorbic acid (50 μg/ml). The cells were grown to confluency, washed with 0.14 M NaCl in 20 mM phosphate buffer, pH 7.4, and stored at −70°C until used.

Enzyme Assays

About 5×10^6 cultured skin fibroblasts were sonicated for 5 s with a 20-kHz cell disrupter in a buffer containing 0.2 M NaCl, O.1 M glycine, 0.1% (v/v) Triton X-100, and 0.02 M Tris-HCl, pH 7.5, at 4°C. The homogenate was centrifuged at 15,000 *g* for 30 min at 4°C, and aliquots of the supernatants were used for enzyme assays. Prolyl 4-hydroxylase and lysyl hydroxylase activity were determined by measuring radioactively labeled 4-hydroxyproline and hydroxylysine from (^{14}C)proline– or (^{14}C)lysine–labeled nonhydroxylated procollagens, respectively (Kivirikko and Myllylä 1982). The enzyme activities were expressed per milligram of soluble cell protein.

Southern and Northern Blot Analysis

High-molecular-weight genomic DNA was isolated from cultured skin fibroblasts as described by Sambrook et al. (1989). In some families, DNA of the parents was isolated from blood cells. DNA was digested with restriction enzymes as indicated by Sambrook et al., and DNA fragments were fractionated electrophoretically on a 0.8% agarose gel, transferred to a nitrocellulose filter, and hybridized to a full-length lysyl hydroxylase cDNA probe as described by Sambrook et al. for northern analysis.

Total RNA was isolated from the cultured human skin fibroblasts using the guanidine isothiocyanate method (Sambrook et al. 1989). RNA was electrophoresed in a 0.7% agarose gel containing 2 M formaldehyde and transferred to a nitrocellulose filter. The northern filters were hybridized at 42°C overnight with a radioactively labeled full-length human lysyl hydroxylase cDNA probe, LH31 (Hautala et al. 1992), under conditions of 50% (v/v) formamide, 5 × SSC (1 × SSC is 0.15 M NaCl, 15 mM sodium citrate buffer, pH 6.8), 1% (w/v) polyvinylpyrroline, 0.250 mg denatured salmon sperm DNA/ml, and 0.1% (w/v) SDS. The filters were washed with 0.5 × SSC, 0.1% (w/v) SDS at 55°C.

cDNA Synthesis and PCR Amplification

cDNA was synthesized (Sambrook et al. 1989) using Moloney murine leukemia virus reverse transcriptase. Total RNA, 5 μg, was reverse transcribed in the presence of 50 mM Tris-HCl, pH 8.3, 40 mM KCl, 6 mM $MgCl_2$, 1 mM dithiothreitol, 0.4 mM of each dNTP, 20 units RNasin (Promega), and 300 units of reverse transcriptase using lysyl hydroxylase–specific primers. The reaction was carried out at 37°C for 1 h, after which the reaction mixture was heated to 95°C for 10 min and treated with GeneClean II (Bio 101).

The PCR amplification of the cDNA and genomic DNA was performed with a Perkin-Elmer Cetus thermal cycler using *Taq* DNA polymerase (Promega) in the presence of 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM $MgCl_2$, and 0.01% (w/v) gelatin. Various lysyl hydroxylase intron- or exon-specific primers (0.2

μM) were used under empirically determined conditions for the amplification. In the screening of the duplication junction in the genome of the patients an oligonucleotide JH71 from intron 16 (5′-CAGTTTTACAACATTCATATC-3′) and an oligonucleotide JH55 from intron 9 of the lysyl hydroxylase gene (5′-GATCACGAGGTCAGGAGAT-3′) were used as primers (Pousi et al. 1994). The reaction mixture was denatured at 94°C for 1 min, annealed at 56°C, and the extension was carried out at 72°C for 1 min. The annealing temperature was decreased by 2°C every two cycles for three times and was maintained at 50°C for the final 30 cycles. The PCR products were electrophoresed on a 1.5% gel and purified with GeneClean II. *Taq* cycle sequencing kit (United States Biochemical) and Perkin-Elmer Cetus thermal cycler were used in direct nucleotide sequencing.

In analysis of the cDNA of patient HR, an oligonucleotide TH16 (5′-CTCTGGGAGGTGTTCAG-3′) from exon 14 sequence and an oligonucleotide JH11 (5′-AATCCACCCCGACTCGG-3′, biotin labeled) of exon 18 sequence (reverse) were used in the amplification in the presence of radioactive dCTP. The annealing temperature was decreased by 2°C every two cycles for three times and was maintained at 51°C for the final 30 cycles. The strands were separated by Dynabeads (Dynal) biomagnetic separation system. The products were fractionated on a 6% sequencing gel, and the radioactivity of the fragments was quantitated using a PhosphoImager (Molecular Dynamics).

Haplotype Analysis

The following microsatellite repeat markers were used in the haplotype analysis: D1S160, D1S548, D1S244, and D1S434. One biotin-labeled primer was used in the amplification for each locus, and the strands were separated by using Dynabeads (Dynal) biomagnetic separation system. The strand was 5′ labeled and run on a 6% sequencing gel. In a lysyl hydroxylase allele analysis, oligonucleotides TH57B (5′-CCGGATGTCTATTGGTT-3′) and RE16LH2 (5′-TGCCTGGTTCACAACCAGCCTG-3′) were used to amplify part of intron 16 (GenBank/European Molecular Biology Laboratory databases U16302) of PLOD (Heikkinen et al. 1994). The single-strand fraction of the amplified DNA was used as a template in a primer extension reaction, which was carried out as one cycle PCR reaction by using 5′ labeled RE16LH8 (5′-GGTGCTAGGCCCTGTGTTCTCT-3′) as a primer.

Results

Frequency of the Duplication of Exons 10–16 of PLOD among Patients with EDS VI

The probands of the 26 new EDS VI families included in this study had characteristic features of the disease. Lysyl hydroxylase activity was determined in skin fibroblasts of all patients, and the activity of prolyl 4-hydroxylase, another enzyme involved in collagen synthesis (Kivirikko and Myllylä 1982), served as a reference activity. The ratio of lysyl hydroxylase to prolyl 4-hydroxylase activity varied from ~5% to 20%, in skin fibroblasts of the patients to those in controls, thus establishing the diagnosis of EDS VIA (data not shown). A low level of hydroxylation of lysine residues was also verified in dermal specimens of patients AR, AB, and TD (not shown).

A PCR assay (Pousi et al. 1994) to detect the seven-exon duplication was performed on cDNA and/or genomic DNA samples of the probands. The homozygosity/heterozygosity of the duplicated PLOD allele in the affected individuals was determined by using *Eco*RI digestion and Southern analysis. As reported elsewhere, *Eco*RI digestion produces a DNA fragment of ~20 kb from the duplicated allele, whereas a corresponding fragment of 13.2 kb is obtained from the normal allele (Hautala et al. 1993; Pousi et al. 1994). A homozygous duplication was found in four of the new families in addition to the two families reported earlier (Hautala et al. 1993; Pousi et al. 1994). Patients from three separate families were found to be compound heterozygotes for the duplication (table 1), as shown by Southern analysis of one of these patients (HR, fig. 1).

DNA was also analyzed from the parents in five of the families (table 1). The data indicated that the duplicated allele is found in parents of the patients from Iran, Russia, Yugoslavia, Turkey, and the United States. The consanguinity of the parents of the Iranian (AR) or Turkish (TD) patients explains the presence of the duplicated allele in both parents. To the best of our knowledge, the parents of the Yugoslavian, Russian, and American (544 in table 1) patients are not related to each other. In the AR family, the patient has two clinically healthy sisters, of whom one was found to be a carrier for the duplicated allele in our assay (fig. 2). In the Yugoslavian family of patient AB, two affected children inherited the duplicated allele from their father (fig. 2). In the present study, the seven-exon duplication has been found in 9 of the 50 genetically independent alleles of 26 families, i.e., in 18% of the alleles. Previously, our analysis of 18 PLOD alleles in nine probands (Hyland et al. 1992; Hautala et al. 1993; Ha et al. 1994; Pousi et al. 1994) has shown a homozygous duplication in two probands (4 alleles) (Hautala et al. 1993; Pousi et al. 1994) (table 1). Therefore, of a total of 68 gene alleles, the duplication has been found in 13 alleles, to give an overall frequency of 19.1%.

We have sequenced the junction fragment of the duplication (Pousi et al. 1994) from all the patients. The sequence covers ~300 nt, part of the nucleotides originating from intron 16 and part from intron 9 of PLOD. Our data indicate that the recombination of PLOD is identical in all individuals. The fragment sequences are identical in all patients (not shown).

Table 1

EDS VI Patients with Duplication of Seven Exons in the Lysyl Hydroxylase Gene

Patient	Sex	No. of the Duplicated Allele	Carrier of the Duplicated Allele[a]		Nationality	Consanguinity of the Parents[a,b]	Reference
			Mother	Father			
AR	F	2	+	+	Iranian	First cousins	P. Heim, M. Raghunath, L. Meiss, U. Heise, R. Myllylä, A. Kohlschütter, and B. Steinmann, unpublished data
HR[c]	F	1	nd	nd	German	nc	
AB[c]	M	1	−	+	Yugoslav	nc	
TD[c]	M	2	+	+	Turk	First cousins	
544[c]	F	2	+	+	American	nc	
558	M	2	+	+	Russian	nc	
948	M	1	nd	nd	American	nd	
CRL1195[d]	F	2	nd	nd	American	nc	Sussman et al. (1974); Pousi et al. (1994)
GM01790[c,d]	F	2	nd	nd	American	nc	Krane et al. (1972); Pinnell et al. (1972); Hautala et al. (1993)

[a] nd = not determined.
[b] nc = no close consanguinity.
[c] Two affected members in the family.
[d] The duplication has been reported previously (Hautala et al. 1993; Pousi et al. 1994).

A Duplicated Allele, Together with an Allele for an Abnormal Shortened Sequence in the Exon 17 Region, Causes the Disease in One Patient

Patient HR (table 1), who was shown to be a compound heterozygote for the duplication mutation by Southern blotting, was analyzed in detail to characterize the mutation in the nonduplicated allele. The cDNA for lysyl hydroxylase was amplified in three partially overlapping parts (Hautala et al. 1993). The 5′ end of the cDNA was amplified to give the same size fragment as that from the control cells (not shown), and the middle part gave two fragments, one as from the control cells and the other as from the cell line containing the homozygous duplication (not shown). The 3′ end of the cDNA, nucleotides corresponding to sequence of exon 15 to exon 19 (Heikkinen et al. 1994), produced two fragments (fig. 3*A*). Sequencing of the smaller fragment

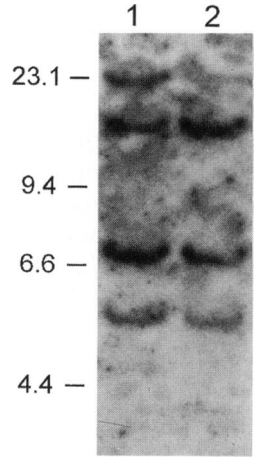

Figure 1 Southern analysis of DNA from HR (lane 1) and the control (lane 2). Genomic DNA isolated from skin fibroblasts was digested with *Eco*RI. Migration of DNA markers (kb) is indicated.

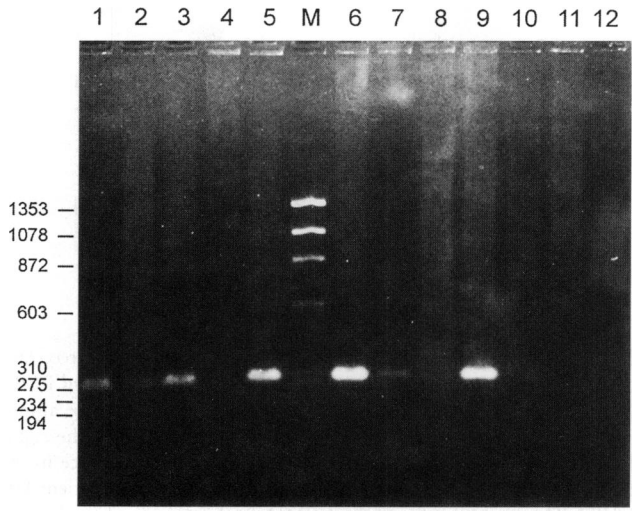

Figure 2 Amplification of the junction fragment of the duplication in PLOD. Genomic DNA was amplified by oligonucleotide primers JH71 and JH55 (Pousi et al. 1994); AR (lane 1), the mother of AR (lane 2), the father of AR (lane 3), sister I of AR (lane 4), sister II of AR (lane 5), AB (lane 6), the sister of AB (lane 7), the mother of AB (lane 8), the father of AB (lane 9), normal individuals (lanes 10 and 11), and water control (lane 12). The molecular-weight sizes (bp) (XHae III; lane M) are shown on the left.

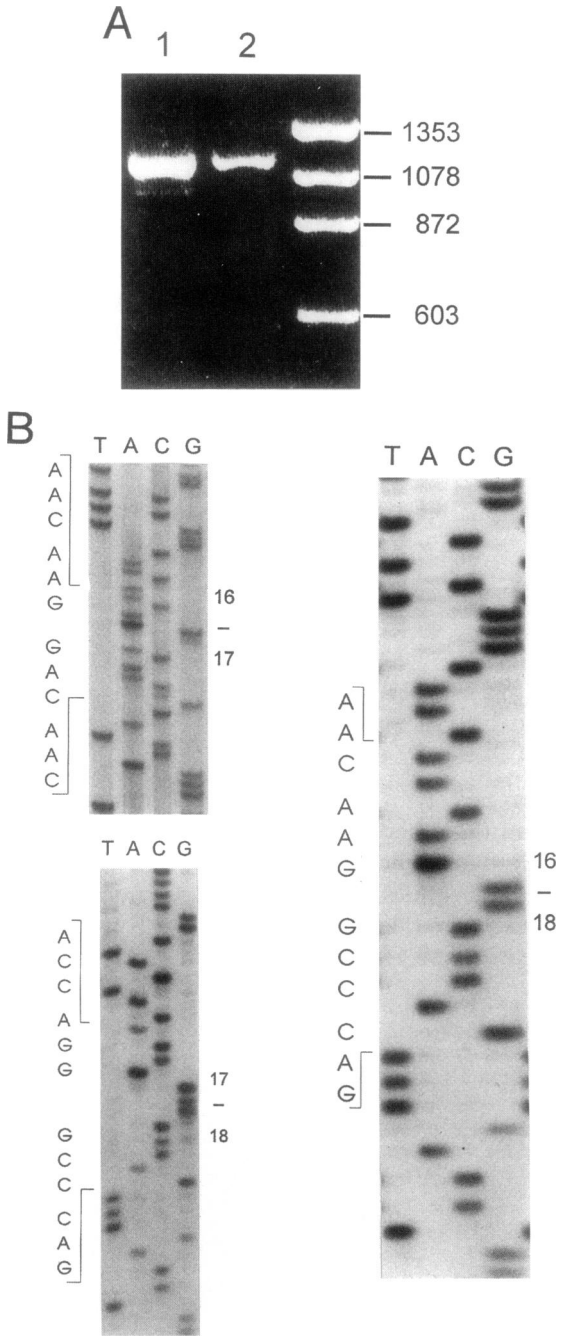

Figure 3 *A*, Amplification of the 3′ end of lysyl hydroxylase cDNA from HR (lane 1) and control (lane 2) cells. PCR1 and PCR2 oligonucleotide primers (Hautala et al. 1993) were used in the amplification. The molecular-weight markers (bp) are shown on the right. *B*, Deletion of nucleotides corresponding to exon 17 sequence in the cDNA sequence in one allele for lysyl hydroxylase in the patient HR (*right*). The sequences for control cDNA in the regions of the boundaries of exons 16 and 17, 17, and 18 are shown (*left*).

indicated that the nucleotides of exon 17 (fig. 3*B*) were deleted from the sequence.

Southern analysis of the genomic DNA that had been digested with *Eco*RI and *Xba*I restriction enzymes indicated the presence of a band indicative of the duplicated allele; otherwise the band pattern was normal (fig. 1, for *Eco*RI digestion). Amplification of the genomic DNA by oligonucleotides covering exon 17 and part of the surrounding introns resulted in two bands, the upper giving the normal sequence and the lower band the abnormal sequence. In both sequences (fig. 4) 102 nt from the 3′ end of exon 17 were identical, whereas 45 nt at the 5′ end of the normal exon were replaced in the abnormal allele by 31 nt. The abnormal allele contained three substitutions and a deletion of 14 nt, which introduced a frameshift. Because of this abnormal nucleotide sequence, a translational stop codon was generated after the codons of nine amino acids.

Sequences corresponding to exon 17 were reamplified from the cDNA. Oligonucleotides of exon 14 and 18 were used in the amplification. The single-stranded radioactively labeled products were fractionated on a denaturing urea polyacrylamide gel. The data indicate (not shown) that the cDNA of HR patient contains fragments of 466, 452, and 319 nt in the ratios of 65:15:20, respectively. The fragment of 466 nt was present in the control cDNA. It represents a product of the duplicated allele in HR patient. The fragment of 452 nt corresponds to a normally spliced product of nonduplicated allele with an abnormal sequence in the exon 17 region. A fragment of 319 corresponds to a product lacking exon 17 sequences. Truncation of mRNA by the sequences of exon 17 does not change the reading frame of the transcript. The data indicate that the abnormality in the sequence of exon 17 causes either a reduction of expression of the abnormal allele or affects the stability of the gene product. Furthermore, the data shows that the altered sequence of the exon 17 region causes partial aberration of mRNA processing.

Occurrence of the Duplicated Allele in General Population

We have previously screened 40 Finnish blood samples and 52 blood samples from U.S. citizens in order to see whether the duplication is a common rearrangement in the genome (Pousi et al. 1994). A negative result was not surprising, owing to the rarity of the disease. We have continued our screening by analyzing 239 new DNA samples from anonymous healthy U.S. citizens. One positive finding was obtained in the test. Taken together, our screening from U.S. citizens revealed one positive among 582 alleles tested.

Haplotype Analysis of the Patients with the Duplication in PLOD

In order to determine genetic homogeneity/heterogeneity of the duplicated cases, a haplotype analysis of 1p36 was carried out. Four microsatellite markers, D1S160, D1S548, D1S244, and D1S434 were used for haplotyping (table 2). The framework map of the mark-

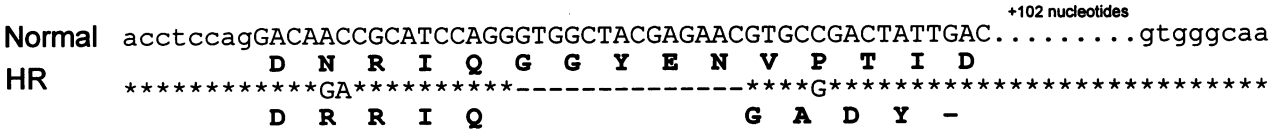

Figure 4 Nucleotide sequence of exon 17 region from genomic DNA. First line, the sequence of the normal allele in the region of exon 17; second line, cDNA deduced amino acid sequence of exon 17; third line, the abnormal sequence of HR; and fourth line, deduced amino acids from the abnormal sequence of exon 17 of HR. The nucleotides of HR identical to those in the normal sequence are indicated by an asterisk (*). The translation-termination codon is designed by a short line (-). A dashed line (- - -) indicates the deleted nucleotides. The exon sequence is indicated by capital letters, the intron sequences by lowercase letters.

ers is thought to be the following: tel-D1S160-D1S548-D1S244-D1S434-cen, with 10 cM separating D1S160 and D1S244 and 9 cM separating D1S244 and D1S434 (Dracopoli et al. 1994). A high-resolution cytogenetic map for DNA markers assigned PLOD between ENO1 and TNFR2 (Van Roy et al. 1993). It is known that TNFR2 is physically linked to D1S434 (Beltinger et al. 1996). Furthermore, it is thought that ENO1 is physically linked to D1S160 and that PLOD locates closer to TNFR2/D1S434 than D1S244 (P. White, personal communication). It is clear from our haplotype analysis that for each marker there are several alleles associated with the duplication.

We have found two tandem dinucleotide repeats in the intron 16 of PLOD located ~1,000 nt 5′ to the duplication starting point. TG nucleotides occur ~10 times and TA repeats occur ~20 times in the sequence. There is a variation in the length of the repeats in different individuals; the number of TA nucleotides is especially variable from one person to another, as shown by the identification of 15 different alleles in 44 samples from anonymous U.S. citizens. We have assayed the length of the tandem repeat from all the duplicated cases and a group of anonymous controls including the person giving a positive finding in our PCR test. It was clearly seen that an allele having a size of 226 bp in our test strongly dominated among the alleles in the controls (fig. 5), with the others comprising a minority of them. It was interesting to find that, among the minority alleles, the longest ones (>258 bp) were dominant among the patients (fig. 5). The long allele was also present in the gene of the control with a positive signal in the PCR test. These findings strongly suggest that the long form of the lysyl hydroxylase allele and the duplication of the gene are associated with each other.

Discussion

We have shown in this report, in which we have screened 50 genetically independent lysyl hydroxylase

Table 2

Haplotype Analysis of the Duplicated Alleles

Allele	No. of the Duplicated Allele	Haplotype for[a]			
		D1S160	D1S548	D1S244	D1S434
AR	2	2	3	3	5
		4	3	8	5
AB	1	2	3	3	5
TD	2	2	1	5	5
		2	1	8	5
TD sister	2	2	1	8	5
		6	1	5	5
948	1	6	2	7	2
CRL1195[b]	2	6	3	7	2
GMO1790	2	6	3	7	2
HR[c]	1	5,5	2,2	4,12	5,5
544[c]	2	2,2	2,3	3,8	1,5
558[c]	2	1,5	2,3	3,3	1,5
No. of different alleles		≥6	4	12	6

[a] The number of haplotype refers to Généthon/CEPH marker database.

[b] Reexamination of the result of our earlier paper in the case of D1S160 (Pousi et al. 1994) confirms this result.

[c] Instead of haplotype, genotype is presented.

Am. J. Hum. Genet. 60:48–56, 1997

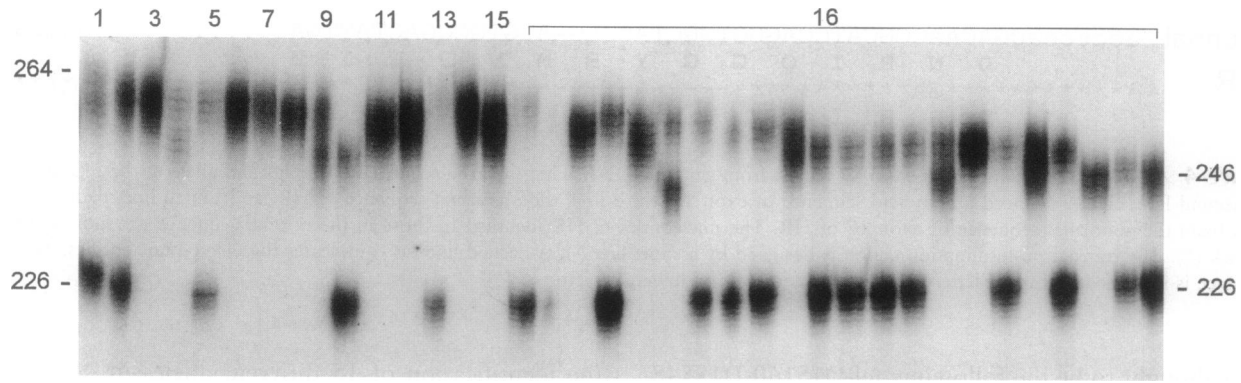

Figure 5 Analysis of dinucleotide repeat sequences in intron 16 of PLOD. Lane 1, a carrier found in general population. Lanes 2–4, the father, the patient, and the mother in the AR family. Lanes 5–7, the father, the patient, and the mother in the TD family. Lanes 8–10, the father, the patient, and the mother in the AB family. Lane 11, GMO1790. Lane 12, CRL1195. Lane 13, HR. Lane 14, 558. Lane 15, 948. Lane 16, examples of samples from anonymous U.S. citizens (44 samples analyzed); the examples indicate a variation in the length of the repeats in different individuals. The size of some alleles is indicated (bp). The data indicate that all patients homozygous with the duplication generated a band >258 bp from both alleles, whereas samples from the general population (including the carrier found in the study) generated a band of similar size from 4 (4.5%) of 88 alleles determined. It should be noted that, because of the duplication, the repeat is present two times in the mutant allele.

alleles from 26 affected families, that a duplication of seven exons is a common cause for EDS VI. About one fifth of the abnormalities of all the patients studied by us so far are due to this duplication. The duplication has occurred identically in all these cases. It is caused by a homologous recombination of two identical 44-nt regions of Alu sequences in introns 9 and 16 in PLOD (Hautala et al. 1993; Heikkinen et al. 1994; Pousi et al. 1994). Primers for the amplification of the junction fragment of the duplication (Pousi et al. 1994) have allowed a rapid screening for detection of this rearrangement among the EDS VI patients.

Southern analysis of the DNA from the German patient (HR) indicated that she was a compound heterozygote for the gene mutation. The parents of this patient were unavailable for studies. Further analysis of the patient indicated that, in addition to the duplication described above, there is an abnormality in the sequence covering the exon 17 region in the other allele of the gene. This mutation makes the other allele functionally abnormal leading to a premature translational stop codon within the exon 17 sequence. The abnormality also leads to aberrant splicing of mRNA. This finding suggests that, in addition to consensus sequences located on the intron side of each splice site, the exon sequences also have an important role for the splice-site selection.

Alu repeats are small transposable elements found in the human genome. More than half a million copies are present in the haploid genome and comprise ~5% of the bulk of the human genomic DNA. Alu repeats have been classified into subfamilies of different evolutionary ages. The oldest subfamily was estimated to be ~55 millions of years (Myr) and the group of younger subfamilies ~20 Myr in age (Schmid and Maraia 1992; Zietkiewich et al. 1994). The greatest expansion of Alu elements (subfamilies Sx, Sp, and Sq) took place prior to simian radiation, whereas the group of younger Alu sequences (subfamilies Sc and Sb) appeared at the time when the Alu amplification was being greatly decelerated. Although most Alu repeats are fixed in the genome, there are indications that some Alu repeats are still mobile (Wallace et al. 1991). The distribution of Alu elements throughout the human genome is not random; the LDL receptor gene, the C1 inhibitor gene, and the α-galactosidase A gene are examples of genes that are rich in Alu sequences (Hu and Worton 1992). PLOD also appears to belong to the family of Alu-rich genes. We have sequenced ~25% of the gene and have found 15 Alu elements in the sequence (Heikkinen et al. 1994), most of which are located in introns 9 and 16. The large number of Alu sequences generates a considerable homology between these introns, and a 44-nt identical sequence located in this region of Alu repeats generates the duplication of exons 10–16 (Heikkinen et al. 1994; Pousi et al. 1994). Alu sequences are also involved in the generation of a duplication of a subset of exons in other genes found in diseases that include familial hypercholesterolemia, Lesch-Nyhan syndrome, and Duchenne muscular dystrophy (Lehrman et al. 1987; Hu et al. 1991; Marcus et al. 1993).

The high frequency of homozygosity for the duplication and the finding that the parents of the homozygous patient are carriers for the duplicated allele confirm our suggestion (Pousi et al. 1994) that uniparental isodisomy inheritance is not involved in the penetrance of the disease. The identification of compound heterozygotes in which one allele has the duplication and the other has a different mutation suggests that the duplicated allele

may be present in the general population. However, because of the recessive type of inheritance, individuals carrying the duplication in one lysyl hydroxylase allele not accompanied by the duplication or a different mutation in the other allele do not develop the symptoms of EDS VI. The low prevalence of the disease suggests that identification of the duplicated allele in the general population requires the analysis of thousands of DNA samples. Our preliminary screening of 184 alleles, which was negative (Pousi et al. 1994), supports this statement. The one positive finding in the screening of 478 additional alleles leads further support to this theory. In order to estimate the frequency of the duplicated allele in the general population, it would be necessary to screen thousands of samples. It is also probable that the frequency varies in different countries.

There are two possible explanations of the duplication in patients with EDS VI. One could assume that there is one ancestral chromosome from which all the duplicated disease cases originate. The other possibility is that there are many different recombinations that can produce a similar duplication in PLOD. If one assumes that the duplication hits the ancestor chromosome, the surrounding DNA region should be identical in the descendant chromosomes. This was not the case in our haplotype analysis. It should be noted, however, that because of recombination this identity gradually disappears as a function of the number of generations. EDS VI is not a disease that is unique to one or two populations. Patients with the syndrome are globally distributed (Krane 1984; Beighton 1993; Steinmann et al. 1993; McKusick 1994; Byers 1995; Prockop and Kivirikko 1995). Our data from patients with the duplicated lysyl hydroxylase allele suggest that, similarly, this rearrangement is not restricted to one population or to a certain geographical region. So, it may well be that association of the genetic markers by the disease locus is found only with an intragenic marker, as indeed is found in this study. Therefore, the data may support an ancestral single-mutational event for the duplication. Furthermore, the duplication mutations are identical at the sequence level in all individuals, giving no evidence for the independent origins of the duplications.

It is known that a variety of DNA sequences play direct or indirect roles in recombination. They may influence recombination by their effects on DNA structure, for instance by forming intramolecular triplexes or stem-loop structures. Some dinucleotide repeats (Wahl and Moore 1990) have been shown to stimulate homologous recombination in transfected DNA, recombination enhancement being conferred over distance. Lysyl hydroxylase haplotype analysis indicated that the duplication was tightly associated with long forms of the dinucleotide repeats in intron 16 of PLOD. The repeats are located ~1 kb up to the breakpoint of the rearrangement. One can speculate that, by some as yet uncharacterized mechanism, a lengthening of the repeat may stimulate homologous recombination in the PLOD, which might lead to independent duplications of the gene.

Acknowledgments

This work was supported by grants from the Research and Science Foundation of Farmos and from the Sigrid Juselius Foundation and by grants from Research Councils for the Natural Sciences and Medicine, both within the Academy of Finland. This work was also supported in part by grant AG10215 from the National Institute of Aging (to H.N.Y.), by a basic research grant from the March of Dimes Birth Defects Foundation (to H.N.Y.), and by a grant from the Swiss National Science Foundation (32-42198.94). The authors thank Dr. Peter Byers, Department of Pathology, University of Washington, Seattle; Dr. Albert Chudley, Clinical Genetics, University of Manitoba, Winnipeg; Dr. M. H. K. Shokeir, Division of Medical Genetics, University of Saskatchewan, Saskatoon; Ann Muilenberg, Division of Medical Genetics, Department of Pediatrics, University of Iowa, Iowa City, for providing EDS VI fibroblasts. Dr. Jeff Murray, Department of Pediatrics, University of Iowa, is acknowledged for providing us with blood DNA samples. Dr. Peter White, Division of Oncology, Children's Hospital of Philadelphia, is acknowledged for his help with microsatellite markers.

References

Beighton P (1993) The Ehlers-Danlos syndrome. In: Beighton P (ed) McKusick's heritable disorders of connective tissue. Mosby-Year Book, St Louis, pp 189–251

Beltinger CP, White PS, Maris JM, Sulman EP, Jensen SJ, LePaslier D, Stallard BJ, et al (1996) Physical mapping and genomic structure of the human TNFR2 gene. Genomics 35:94–100

Byers PH (1995) Disorders of collagen biosynthesis and structure. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) The metabolic and molecular bases of inherited diseases. McGraw-Hill, New York, pp 4029–4077

Dracopoli NC, Bruns GAP, Brodeur GM, Landes GM, Matise TC, Seldin MF, Vance JM, et al (1994) Report of the first international workshop on human chromosome 1 mapping 1994. Cytogenet Cell Genet 67:144–165

Ha VT, Marshall MK, Elsas LJ, Pinnell SR, Yeowell HN (1994) A patient with Ehlers-Danlos syndrome type VI is a compound heterozygote for mutations in the lysyl hydroxylase gene. J Clin Invest 93:1716–1721

Hautala T, Byers MG, Eddy RL, Shows TB, Kivirikko KI, Myllylä R (1992) Cloning of human lysyl hydroxylase: complete cDNA-derived amino acid sequence and assignment of the gene to chromosome 1p36.3-p36.2. Genomics 13: 62–69

Hautala T, Heikkinen J, Kivirikko KI, Myllylä R (1993) A large duplication in the gene for lysyl hydroxylase accounts for the type VI variant of the Ehlers-Danlos syndrome in two siblings. Genomics 15:399–404

Heikkinen J, Hautala T, Kivirikko KI, Myllylä R (1994) Structure and expression of the human lysyl hydroxylase gene (PLOD): introns 9 and 16 contain Alu sequences at the

sites of recombination in Ehlers-Danlos syndrome type VI patients. Genomics 24:464–471

Hu X, Ray PN, Worton RG (1991) Mechanisms of tandem duplication in the Duchenne muscular dystrophy gene include both homologous and nonhomologous intrachromosomal recombination. EMBO J 10:2471–2477

Hu X, Worton RG (1992) Partial gene duplication as a cause of human disease. Hum Mutat 1:3–12

Hyland J, Ala-Kokko L, Royce P, Steinmann B, Kivirikko KI, Myllylä R (1992) A homozygous stop codon in the lysyl hydroxylase gene in two siblings with Ehlers-Danlos syndrome type VI. Nat Genet 2:228–231

Kivirikko KI, Myllylä R (1980) The hydroxylation of prolyl and lysyl residues. In: Freedman RB, Hawkins HC (eds) The enzymology of post-translational modification of proteins. Academic Press, London, pp 53–104

——— (1982) Post-translational enzymes in the biosynthesis of collagen: intracellular enzymes. Methods Enzymol 82A: 245–304

Kivirikko KI, Myllylä R, Pihlajaniemi T (1992) Hydroxylation of proline and lysine residues in collagens and other animal and plant proteins. In: Harding JJ, Crabbe MJC (eds) Post-translational modifications of proteins. CRC Press, pp 1–51

Krane SM (1984) Genetic and acquired disorders of collagen deposition. In: Piez KA, Reddi AH (eds) Extracellular matrix biochemistry. Elsevier, New York, pp 413–463

Krane SM, Pinnell SR, Erbe RW (1972) Lysyl-protocollagen hydroxylase deficiency in fibroblasts from siblings with hydroxylysine-deficient collagen. Proc Natl Acad Sci USA 69: 2899–2903

Lehrman MA, Goldstein JL, Russell DW, Brown MS (1987) Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia. Cell 48:827–835

Marcus S, Hellgren D, Lambert B, Fällström SP, Wahlström J (1993) Duplication in the hypoxanthine phosphoribosyltransferase gene caused by Alu-Alu recombination in a patient with Lesch Nyhan syndrome. Hum Genet 90:477–482

Marshall ML, Walker L, Murad S, Pinnell S, Yeowell H (1994) Dermal fibroblasts from a subset of patients with Ehlers-Danlos syndrome type VI exhibit an abnormally large 4.2 Kb mRNA for lysyl hydroxylase. Matrix Biol 14:398

McKusick VA (1994) Mendelian inheritance in man: a catalog of human genes and genetic disorders. John Hopkins University Press, Baltimore, pp 1786–1787

Myllylä R, Pihlajaniemi T, Pajunen L, Turpeenniemi-Hujanen T, Kivirikko KI (1991) Molecular cloning of chick lysyl hydroxylase. Little homology in primary structure to the two types of subunit of prolyl 4-hydroxylase. J Biol Chem 266:2805–2810

Pinnell SR, Krane SM, Kenzora JE, Glimcher MJ (1972) A heritable disorder of connective tissue: hydroxylysine-deficient collagen disease. N Engl J Med 286:1013–1020

Pousi B, Hautala T, Heikkinen J, Pajunen L, Kivirikko KI, Myllylä R (1994) Alu-Alu recombination results in a duplication of seven exons in the lysyl hydroxylase gene in a patient with the type VI variant of Ehlers-Danlos syndrome. Am J Hum Genet 55:899–906

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

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

Schmid C, Maraia R (1992). Transcriptional regulation and transpositional selection of active SINE sequences. Curr Opin Genet Dev 2:874–882

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

Sussman M, Lichtenstein J, Nigra TP, Martin GR, McKusick VA (1974) Hydroxylysine-deficient skin collagen in a patient with a form of the Ehlers-Danlos syndrome. J Bone Joint Surg 56-A:1228–1234

Van Roy N, Laureys G, Versteeg R, Opdenakker G, Speleman F (1993) High-resolution fluorescence mapping of 46 DNA markers to the short arm of human chromosome 1. Genomics 18:71–78

Wahl WP, Moore PD (1990) Homologous recombination enhancement conferred by the Z-DNA motif $d(TG)_{30}$ is abrogated by simian virus 40 T antigen binding to adjacent DNA sequences. Mol Cell Biol 10: 794–800

Wallace MR, Andersen LB, Sailino AM, Gregory PE, Glover TW, Collins FS (1991) A de novo Alu insertion results in neurofibromatosis 1. Nature 353:861–866

Zietkiewics E, Richer C, Makalowski W, Jurka J, Labuda D (1994) A young Alu subfamily amplified independently in human and African great Apes lineages. Nucleic Acid Res 22:5608–5612