

Use of Molecular Haplotypes Specific for the Human pro α 2(I) Collagen Gene in Linkage Analysis of the Mild Autosomal Dominant Forms of Osteogenesis Imperfecta

CATHERINE T. FALK,¹ ROBIN C. SCHWARTZ,² FRANCESCO RAMIREZ,³
AND PETROS TSIPOURAS^{2,4}

SUMMARY

Autosomal dominant osteogenesis imperfecta (OI) is a heterogeneous group of disorders. Molecular haplotypes associated with the pro α 2(I) gene of human type I procollagen were used for genetic linkage studies in a group of 10 families with OI. The clinical phenotypes of the families studied were those of OI type I and OI type IV. Evidence for linkage was highly suggestive in the four families with OI type IV ($Z = 3.91$ for $\hat{\theta} = 0$). In contrast, little or no indication for linkage was found in the six families with OI type I ($Z = .055$ for $\hat{\theta} = .415$). Heterogeneity between the two groups of families was highly significant ($\chi^2 = 11.14$, $P = .0008$), suggesting that at least two separate gene defects may be the cause of the autosomal dominant forms of OI.

INTRODUCTION

Osteogenesis imperfecta (OI) is a highly heterogeneous group of heritable systemic disorders of the connective tissue [1]. The clinical heterogeneity of these syndromes has been adequately defined recently [2], and the biochemical and molecular characterization of certain mutants has been correlated to specific

Received June 18, 1985.

This work was supported in part by grants GM29177 (C.T.F.) and AM032380 (F.R.) from the National Institutes of Health, and grant 6-411 from the March of Dimes-Birth Defects Foundation. P. T. is the recipient of Clinical Investigator Award AM-01224 from the National Institutes of Health.

¹ The New York Blood Center, New York, N.Y.

² Department of Pediatrics, UMDNJ-Rutgers Medical School, Piscataway, N.J.

³ Department of Obstetrics-Gynecology, UMDNJ-Rutgers Medical School.

⁴ Department of Biochemistry, UMDNJ-Rutgers Medical School.

clinical phenotypes [3]. In the mild forms, biochemical studies have suggested alteration in synthesis of the pro α 1(I) chains of type I procollagen (OI type I) [4] or a structural defect in the pro α 2(I) chain (OI type IV) [5].

The availability of specific DNA probes and their use in establishing restriction fragment length polymorphisms (RFLPs) have proved to be useful tools in the characterization of heterogeneous monogenic disorders [6–10]. Using human pro α 2(I) collagen gene probes, we identified two RFLPs, which we subsequently used for linkage studies in several families with mild variants of OI. One family, with OI type IV, showed a possibility of close linkage [11], whereas three families with OI type I were clearly not closely linked to the pro α 2(I) collagen gene [12]. The molecular heterogeneity, following a pattern of segregation similar to that of specific clinical phenotypes, prompted us to expand the study to include more families with mild OI.

We report here another RFLP in the pro α 2(I) gene and the use of three markers and the resulting molecular haplotypes for linkage studies in 10 families with mild autosomal dominant OI. Our data strongly suggest that at least two different gene defects are the cause in the mild forms of OI.

MATERIALS AND METHODS

Subjects

Affected and unaffected individuals from 10 families with mild dominant OI were tested for one or more of three polymorphisms associated with the human pro α 2(I) gene. Six of these families had OI type I according to clinical criteria [2], while the remaining four families had OI type IV. Pedigrees of the 10 families studied, together with relevant marker phenotypes, are given in the APPENDIX.

Additionally, DNA was obtained from a random sample of 91 individuals. All individuals were tested for one or more of the RFLPs, and 42 were tested for all three.

Restriction Endonuclease Analysis of Genomic DNA

Nuclear DNA was isolated from the leukocytes contained in 10–15 ml of EDTA-anticoagulated blood. Ten to 15 μ g of DNA were digested to completion under conditions recommended by the commercial supplier. Digested DNA and appropriate DNA size markers were separated by electrophoresis in 0.6% or 1.0% (w/v) agarose gels. The DNA fragments were transferred to nitrocellulose filters [13] and hybridized with the human pro α 2(I) probes for 24–48 hrs as described [11]. The filters were then washed for 10 min at 68°C with each of the following solutions: 2 \times SSC, 1 \times SSC, 0.5 \times SSC and 0.1 \times SSC (SSC, buffer containing 0.15 M NaCl in 0.015 M sodium citrate, pH 6.8). The probes used in the experiments were labeled to a specific activity of 2–5 \times 10⁸ cpm/ μ g by nick-translation.

DNA Probes for the Human pro α 2(I) Gene

The genomic probes used in these experiments have been described [12]. The *Eco*RI RFLP specific probe consisted of 6.75 kilobases (kb) of genomic DNA extending downstream from the codon of amino acid residue 19 of the pro α 2(I) chain [14]. The *Msp*I and *Stu*I RFLP specific probes consisted of 4.1 kb of genomic DNA containing coding sequences for the triple helical domain and the C-propeptide of the pro α 2(I) chain [14].

Nomenclature

The recommendations of the Seventh International Workshop on Human Gene Mapping were followed in naming the alleles generated by the presence or absence of the

restriction endonucleases *EcoRI*, *MspI*, and *StuI* [15]. Thus, the allele generated by the presence of the *EcoRI* site is designated as *A1* and the one generated by the absence of it as *A2*. Accordingly, we designated as *B1*, *B2* and *C1*, *C2* the alleles in which *MspI* and *StuI* sites were present or absent, respectively.

Linkage Analysis

Linkage analysis was carried out for OI vs. *EcoRI* in all 10 families using the computer program LIPED [16]. Because of the suggestion by Tshipouras et al. [12] that the clinical heterogeneity of the mild forms of OI might reflect different linkage relationships to the human *pro α 2(I)* gene RFLPs, we looked for heterogeneity of lod scores between the type I and type IV families using Morton's test for heterogeneity [17]. Linkage analysis between the *EcoRI* and *MspI* RFLPs was also carried out in the four OI families informative for both DNA markers to determine whether any recombinants were detectable. In the absence of recombinants, linkage analysis between OI and RFLP haplotypes could be performed, possibly increasing the total amount of linkage information.

Estimates of Gene and Haplotype Frequencies and Allelic Associations in Unrelated Individuals

Gene frequencies were obtained by gene counting for the three RFLPs. Haplotype frequencies were estimated using maximum likelihood methods for all pairwise haplotypes as well as for the three-locus haplotypes. The gene and pairwise haplotype frequencies were used to estimate allelic associations due to linkage disequilibrium.

RESULTS

Polymorphic Restriction Sites in the pro α 2(I) Gene

A random sample of 91 individuals as well as members from 10 families with mild autosomal dominant OI were tested for one or more of three DNA polymorphisms. Two of those polymorphisms have been described [11, 12]. The third, generated by the restriction endonuclease *StuI*, has been recently identified. Nuclear DNA from 49 unrelated individuals and also from members of three families was cleaved with restriction endonuclease *StuI*. A polymorphic site was detected after hybridization with the region specific probe. Individuals with three different genotypes were identified (fig. 1). Segregation analysis in three families (data not shown) demonstrated that the polymorphic site segregated as an autosomal codominant trait. A total of 98 chromosomes was examined. The allelic frequencies were .92 for the *C1* and .08 for the *C2* allele. The frequencies of the three genotypes were compatible with Hardy-Weinberg equilibrium. This RFLP is most likely generated by a single base change since no size variations were observed in fragments generated by other restriction endonucleases that cleave in the same region of the gene. Allele frequencies for all three RFLPs based on our random sample of 91 unrelated individuals are given in table 1.

Linkage and Heterogeneity Analysis

The results of linkage analysis for OI vs. *EcoRI* are given in table 2. The total lod scores for all 10 families suggest the possibility of linkage between OI and *EcoRI*, but the combined total does not reach the generally accepted level of 3.0 (see table 2). Because of the suggestion of linkage heterogeneity reported by Tshipouras et al. [12], the 10 families were separated into two clinical groups,

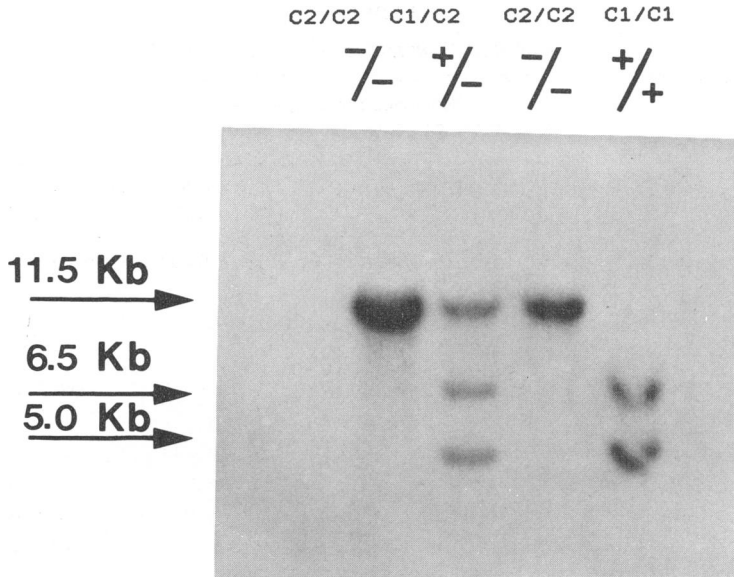


FIG. 1.—Molecular genotypes generated by the restriction endonuclease *Stu*I

those with OI type I or OI type IV (table 3). Morton's test for heterogeneity of linkage was then carried out on these two groups. Linkage heterogeneity is evident in this extended analysis that includes the five families previously reported [12] and five additional ones. The χ^2 for heterogeneity between the two sets of families is highly significant ($\chi^2 \times 11.14$, 1 d.f.). There is no evidence for heterogeneity within the families of each OI type (table 4). With the demonstration of heterogeneity between the two clinical subtypes, linkage analysis can be carried out independently for the two groups. There is good evidence for close linkage between OI type IV and the *Eco*RI RFLP ($Z = 3.91$ at $\hat{\theta} = 0$). In contrast, in all six families with OI type I, at least one recombinant is detectable between OI and either *Eco*RI, *Msp*I, or the *Eco*RI/*Msp*I haplotype (table 2 and APPENDIX).

TABLE 1

GENE FREQUENCIES OF THE THREE RFLPS, *Eco*RI, *Msp*I, AND *Stu*I, ASSOCIATED WITH THE PRO α 2(I) COLLAGEN GENE

RFLP	Sample size	Presence of site	Absence of site
<i>Eco</i> RI	84	.339	.661
<i>Msp</i> I	88	.886	.114
<i>Stu</i> I	49	.918	.082

TABLE 2
LOD SCORES FOR LINKAGE BETWEEN OI AND *EcoRI*

OI TYPE	No. FAMILIES	RECOMBINATION FREQUENCIES						$\hat{\theta}$	Z
		.05	.10	.20	.30	.40	.45		
Type I	6	-5.276	-2.927	-0.904	-0.144	0.052	0.039	0.415	0.055
Type IV	4	3.558	3.187	2.391	1.525	0.638	0.257	0.000	3.910
Total . . .	10	-1.718	0.260	1.487	1.381	0.690	0.296	0.237	1.545

In four OI families, information was obtained for both the *EcoRI* and *MspI* RFLPs. Because there was no recombination between the two DNA markers ($Z = 3.19$, $\hat{\theta} = 0$), we were able to determine joint *EcoRI/MspI* haplotypes for individuals in those families and carry out linkage analysis between the haplotypes and OI. In two families (both with OI type I), additional linkage information was obtained. In one, definite recombinants were detected between the molecular haplotype and OI due to informative *MspI* phenotypes in individuals uninformative for *EcoRI*. The overall result in type I families was to shift $\hat{\theta}$ from .415 to .35 and increase Z from .055 to .280. The test for heterogeneity between type I and type IV families remained highly significant ($\chi^2 = 9.95$, 1 d.f.).

Molecular Haplotypes

In the absence of sufficient family data, population data can be used to estimate haplotype frequencies using maximum likelihood (ML) methods. In those classes where the phase of the haplotypes cannot be observed (i.e., in double heterozygotes), the ML estimate provides the most likely distribution of the haplotypes from those ambiguous phenotypic classes, based on the sample data available. Using the sample of 91 random individuals typed for two or

TABLE 3
CLINICAL PHENOTYPES

Family no.	Onset	Hearing loss	Dentinogenesis imperfecta	Scleral hue	OI type
1	Postnatal	+	-	Blue	I
2	Postnatal	+	-	Blue	I
3	Postnatal	+	-	Blue	I
4	Postnatal	+	-	Blue	I
5	Postnatal	+	-	Blue	I
6	Postnatal	+	+	Blue	I
7	Postnatal	+	-	Not blue	IV
8	Postnatal	+	+	Not blue	IV
9	Postnatal	+	+	Not blue	IV
10	Postnatal	+	+	Not blue	IV

TABLE 4
HETEROGENEITY TEST FOR LINKAGE: TYPE I VS. TYPE IV FAMILIES

Comparison	d.f.	χ^2	P-value
Type I vs. Type IV	1	11.144	.0008
Intra-Type I	5	1.612	.8998
Intra-Type IV	3	0	. . .
Total	9	12.756	.1729

three of the RFLPs, the ML estimates of the haplotype frequencies for the three RFLP locus pairs and for all three loci were calculated. The ML haplotype frequencies are given in table 5. There is strong linkage disequilibrium between alleles for all locus pairs. The positive associations (δ values) are shown in table 6 along with D' values, showing what fraction of the maximum δ value is represented by the estimated δ [18]. The haplotype frequencies for the

TABLE 5
MAXIMUM LIKELIHOOD (ML) HAPLOTYPE FREQUENCIES OF THE RFLP POLYMORPHISMS
ASSOCIATED WITH THE *pro α 2(I)* GENE

A. Two-locus haplotype frequencies		
Locus pair (no.)	Haplotype	ML frequency
<i>ECO-MSP</i> (168)	A1B1	.228
	A1B2	.112
	A2B1	.653
	A2B2	.007
<i>ECO-STU</i> (84)	A1C1	.254
	A1C2	.080
	A2C1	.651
	A2C2	.015
<i>MSP-STU</i> (92)	B1C1	.879
	B1C2	.001
	B2C1	.033
	B2C2	.087
B. Three-locus haplotype frequencies		
	Haplotype	ML frequency
	A1B1C1	.215
	A1B1C2	.001
	A1B2C1	.035
	A1B2C2	.081
	A2B1C1	.651
	A2B1C2	.001
	A2B2C1	.001
	A2B2C2	.015

TABLE 6
POSITIVE ALLELIC ASSOCIATIONS DUE TO LINKAGE DISEQUILIBRIUM

Locus pair	Alleles	δ value	D' value*
<i>MSP-STU</i>	B1C1	.076	.894
	B2C2	.077	.906
<i>ECO-MSP</i>	A1B2	.072	.911
	A2B1	.071	.899
<i>ECO-STU</i>	A1C2	.048	.762
	A2C1	.047	.746

NOTE: δ values are calculated using haplotype frequencies and allele frequencies from the samples indicated in table 5A.

* D' is defined as the actual linkage disequilibrium (δ) divided by the maximum value that the disequilibrium could take on with unchanged allele frequencies [18].

three-locus data are also given in table 5. It can be seen that two of the eight haplotypes account for over 86% of the sample.

DISCUSSION

Genetic linkage has been used to unravel the heterogeneity of hereditary disorders. It can be particularly powerful if applied to disorders with known or suspected clinical heterogeneity, such as OI. This group of genetic disorders is characterized by bone fragility and other systemic manifestations of the connective tissue. OI can be classified in at least four different groups based on the clinical phenotype and mode of inheritance [2]. Autosomal dominant OI (OI type I and OI type IV) is characterized by postnatal onset of fractures, absent, mild, or moderate skeletal deformity, and presenile hearing loss. Individuals affected with OI type IV differ clinically from individuals affected with OI type I in several ways. They lack blue sclerae and tend to be short-statured adults. A significant number of them are also born with skeletal fractures and bowed tibiae and many develop dentinogenesis imperfecta [19]. Defects in the structure of the α -chains of type I procollagen have been shown in a few variants with mild dominant, severe, or lethal perinatal OI [20–23].

The presence of several RFLPs within the pro α 2(I) gene enabled us to study the linkage relationship between autosomal dominant OI and that gene. We previously reported on linkage heterogeneity in five families with autosomal dominant OI [12]. In our expanded sample of 10 families, the results of linkage analysis suggest the possibility of linkage between OI and the *Eco*RI RFLP. A general test for linkage heterogeneity among all 10 families along the line suggested by Smith [24] and by Ott [25] is consistent with the possibility of heterogeneity, but does not quite reach a significant level (see table 7). However, because of the clearly defined clinical differences between type I and type IV OI, it is reasonable to use the "predivided sample test," as described by Hodge et al. [26], on the two clinically distinct sets of families (table 3). Here Morton's test for heterogeneity is highly significant with $P = .0008$ (table 4).

TABLE 7
 HETEROGENEITY TEST FOR LINKAGE BETWEEN OI AND *EcoRI* AMONG 10 FAMILIES
 WITH OI TYPE I OR OI TYPE IV [24]

Hypothesis	Alpha	Z	$\hat{\theta}$
H0: No linkage	(0)	0	0.50
H1: Linkage in all families	1.00	1.54	0.25
H2: Linkage in a fraction (alpha) of families no linkage in (1-alpha) of families35	1.95	0

Components of χ^2 :			
Source	d.f.	χ^2	P-value (one-tail)
H2 vs. H1 heterogeneity	1	1.92	0.083
H1 vs. H0 linkage	1	7.07	0.004
H2 vs. H0 total	2	8.99	0.005

Our results indicate that the two different clinical types of autosomal dominant OI (OI type I and OI type IV) are also etiologically different. This conclusion is supported by at least two studies. Barsh et al. [4] reported that cultured fibroblasts from members of some families with OI type I synthesized a reduced amount of pro α 1(I) chains, suggesting a nonfunctional allele. Also, the biochemical analysis of collagens synthesized by cells of affected individuals from a family with OI type IV previously reported [11] revealed a small peptide deletion in the pro α 2(I) chain (R. J. Wenstrup, P. Tsipouras, and P. H. Byers [27] and unpublished data, 1985).

The previous observation strengthens the contention that the observed linkage heterogeneity reflects etiological heterogeneity in the mild forms of OI. It has been suggested that autosomal dominant OI can be further subdivided in two groups according to the presence or absence of dentinogenesis imperfecta [28, 29]. Our results indicate that linkage studies fail to distinguish between the two subtypes of OI type IV (tables 3 and 4), suggesting that those two clinical phenotypes may result from different mutations of the pro α 2(I) chains.

We reported here on another RFLP associated with the pro α 2(I) gene. Analysis of our data indicates that the three DNA markers are in linkage disequilibrium and two molecular haplotypes are predominant in our sample. Definition of haplotypes in individual genes or gene clusters is essential because specific mutations in certain ethnic groups are often fixed in particular molecular haplotypes, as shown in the β -thalassemias [10].

Genetic analysis of a larger number of families with OI or other genetic disorders of the connective tissue with parallel definition of the biochemical and molecular defects could prove that a similar situation exists in gene systems other than the β -globin.

The use of DNA markers associated with the various human collagen genes for linkage studies in families with connective tissue disorders will prove to be a very powerful method to dissect the genetic heterogeneity and define the molecular defects in this group of disorders.

NOTE ADDED IN PROOF: The *StuI* RFLP was also found to be in linkage disequilibrium with the two other COL1A2-related RFLPs by Børresen et al. (BØRRESEN A-L, BERG K, TSIPOURAS P, DICKSON LA, PROCKOP DJ, RAMIREZ F: *DNA Polymorphisms in Collagen Genes: Potential Use in the Study of Disease in Medical Genetics: Past, Present, Future*. New York, Alan R. Liss, 1985).

ACKNOWLEDGMENTS

We gratefully acknowledge Drs. L. Stefan Levin and C. R. Paterson for referring three of the families included in the study, Dr. Anne-Lise Børresen for helpful discussions, Dr. Peter Byers for reviewing the manuscript, and Mrs. Lois Bertha for secretarial assistance.

REFERENCES

1. SMITH R, FRANCIS MJO, HOUGHTON GR: *The Brittle Bone Syndrome*. London, Butterworths, 1983
2. SILLENCE DO, SENN A, DANKS DM: Genetic heterogeneity in osteogenesis imperfecta. *J Med Genet* 16:101-116, 1979
3. RAMIREZ F, SANGIORGI F, TSIPOURAS P: Human collagens: biochemical, molecular and genetic features in normal and diseased states, in *Human Genes and Diseases*, edited by BLASI F, New York, John Wiley, 1985
4. BARSH GS, DAVID KE, BYERS PH: Type I osteogenesis imperfecta: a non-functional allele for pro α 1(I) chains of type I procollagen. *Proc Natl Acad Sci USA* 79:3838-3842, 1982
5. BYERS PH, SHAPIRO JR, ROWE DW, DAVID KE, HOLBROOK KA: Abnormal α 2-chain in type I collagen from a patient with a form of osteogenesis imperfecta. *J Clin Invest* 71:689-697, 1983
6. BOTSTEIN D, WHITE RL, SKOLNICK M, DAVIS RW: Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314-331, 1980
7. PHILLIPS JA, HJELLE BL, SEEBURG PH, ZACHMANN M: Molecular basis for familial isolated growth hormone deficiency. *Proc Natl Acad Sci USA* 78:6372-6375, 1981
8. PROCHOWNIK EV, ANTONARAKIS S, BAUER KA, ROSENBERG RD, FEARON EF, ORKIN SH: Molecular heterogeneity of inherited antithrombin III deficiency. *N Engl J Med* 308:1549-1552, 1983
9. CAVENEE WK, DRYJA TP, PHILLIPS RA, ET AL.: Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* 305:779-784, 1982
10. ORKIN SH, KAZAZIAN HH JR, ANTONARAKIS SE ET AL.: Linkage of β -thalassemia mutations and β -globin gene polymorphisms with DNA polymorphisms in human β -globin gene cluster. *Nature* 296:627-631, 1982
11. TSIPOURAS P, MYERS JC, RAMIREZ F, PROCKOP DJ: Restriction fragment length polymorphism associated with the pro α 2(I) gene of human type I procollagen. *J Clin Invest* 72:1262-1267, 1983
12. TSIPOURAS P, BØRRESEN A-L, DICKSON LA, BERG K, PROCKOP DJ, RAMIREZ F: Molecular heterogeneity in the mild autosomal dominant forms of osteogenesis imperfecta. *Am J Hum Genet* 36:1172-1179, 1984
13. SOUTHERN EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98:503-517, 1975
14. MYERS JC, DICKSON LA, DE WET WJ, ET AL.: Analysis of the 3' end of the human pro α 2(I) collagen gene. *J Biol Chem* 258:10128-10135, 1983
15. SKOLNICK MH, WILLARD HP, MENLOVE LA: Report of the Committee on Human Gene Mapping by Recombinant DNA Techniques. *Cytogenet Cell Genet* 37:210-273, 1984
16. OTT J: Estimation of the recombination fraction in human pedigrees: efficient com-

