Restriction Fragment Length
Polymorphism Associated with the Pro\(\alpha_2(I)\) Gene of
Human Type I Procollagen

APPLICATION TO A FAMILY WITH AN
AUTOSOMAL DOMINANT FORM OF OSTEOGENESIS IMPERFECTA

Petros Tsipouras, Jeanne C. Myers, Francesco Ramirez, and
Darwin J. Prockop, Departments of Biochemistry, Pediatrics, and Obstetrics
and Gynecology, University of Medicine and Dentistry of New Jersey-
Rutgers Medical School, Piscataway, New Jersey 08854

Abstract One cloned complementary DNA and
one genomic subclone were used to detect restriction
fragment length polymorphism associated with the
pro\(\alpha_2(I)\) gene for human type I procollagen. The re-
striction fragments obtained from examination of 30-
122 chromosomes confirmed previous indications that
the pro\(\alpha_2(I)\) gene is found in a single copy in the hu-
man haploid genome. One highly polymorphic site was
detected with EcoRI in the 5'-half of the gene. The
restriction site polymorphism at the site had an allelic
frequency of 0.38, and it generated two fragments of
10.5 and 3.5 kilobase in homozygous individuals. The
restriction fragment length polymorphism generated
at the EcoRI site was used to study affected and non-
affected individuals in four generations of a family
with an autosomal dominant form of osteogenesis im-
perfecta. The data demonstrated a linkage of the phe-
notype to a pro\(\alpha_2(I)\) allele with a lod score of 2.41
at a recombination fraction \(\theta\) of 0. The data therefore
provided presumptive evidence that osteogenesis im-
perfecta in this family is caused by a mutation in the
pro\(\alpha_2(I)\) gene or some contiguous region of the ge-
nome. The relatively high frequency of polymorphism
at the EcoRI site makes it useful for studying a broad
range of genetic disorders in which mutations in type
I procollagen are suspected. In addition, the poly-
morphic site should provide useful markers for linkage
studies with other loci located on human chromo-
some 7.

INTRODUCTION

Variations in base sequences in and around genes gen-
erate DNA fragments of varying length when genomic
DNA from different individuals is cleaved with re-
striction endonucleases. The detection of such frag-
ments of varying lengths, known as RFLP\(^1\), was first
used to examine the inheritance of specific alleles of
mitochondrial and yeast genes (1-3). Within the past
several years, it has proven to be a powerful tool for
the study of human genetic disorders such as sickle
cell disease (4-5), thalassemias (6), and human growth
hormone deficiency (7, 8).

Recently, clones of complementary (c)DNA and ge-
nomic DNA for type I procollagen have become avail-
able (9-19). Studies with these DNA have revealed
some unique characteristics of the collagen genes such
as a complex intron-exon structure (9-11), a high de-
gree of nucleotide conservation in an area coding for
the carbohydrate attachment site in the \(C\)-propeptide
of the pro\(\alpha_2(I)\) chain (16), and transcription of each
gene into several different messenger RNA (19). Fur-
thermore, it has been established that the two human
genes for pro\(\alpha_1(I)\) and pro\(\alpha_2(I)\) chains are not syntenic

\(^1\) Abbreviations used in this paper: kb, kilobase; OI, os-
teo genesis imperfecta; RFLP, restriction fragment length
polymorphism; SSC, buffer containing 0.15 M NaCl in 0.015
M sodium citrate, pH 6.8.
but they are located on chromosomes 17 (17) and 7 (18), respectively. Here we have used one cloned cDNA and a genomic subclone to detect RFLP associated with the proα2(I) gene of type I procollagen.

METHODS

Materials. Restriction endonucleases and other enzymes were purchased from New England Nuclear, Boston, MA, and Boehringer-Mannheim Biochemicals, Indianapolis, IN. α3P-labeled deoxynucleotides were obtained from New England Nuclear and nitrocellulose paper was obtained from Schleicher & Schuell, Inc., Keene, NH.

Nuclear DNA preparation. High-molecular-weight DNA was prepared (20) from peripheral blood obtained from randomly selected normal individuals and from affected and nonaffected members of one family with autosomal dominant forms of osteogenesis imperfecta (OI).

Preparation of DNA probes. The cloned cDNA Hf-32 was used here as an insert in pBR322 (14). The probe specific for the 5'-end of the gene (19) consisted of two contiguous EcoRI genomic subclones (NJ-3' and NJ-3'' in Fig. 1). The EcoRI site at the 5'-terminus of NJ-3' is an artificial site, generated by the linker in the course of preparing a lambda phage library with an Alul/HaeIII digest of genomic DNA (22). The probes were labeled to a specific activity of 2 to 5 × 10⁶ cpm/μg by nick-translation.

Analysis of RFLP. Samples of nuclear DNA prepared from leukocytes were digested to completion under conditions recommended by the commercial supplier. Digested DNA and appropriate DNA size markers were separated by electrophoresis in 0.6–1.0% (wt/vol) agarose gels. The DNA fragments were transferred to nitrocellulose filters and hybridized with the human procα2(I) probes for 24–48 h at 40°C (20). The filters were then washed for 10 min at 68°C with each of the following solutions: 2 × SSC, 1 × SSC, 0.5 × SSC, and 0.1 × SSC (SSC, buffer containing 0.15 M NaCl in 0.015 M sodium citrate, pH 6.8) (20).

RESULTS

DNA probes for the human proα2(I) gene. The cDNA probe (Hf-32) used in these experiments contained 1,443 nucleotides coding for the α2(I) domain and 597 nucleotides coding for the C-propeptide, and it included about half of the coding sequences of the gene (16). Because of the large number and large size of the intervening sequences in the gene, the Hf-32 probe hybridized to fragments spanning over 12 kilobase (kb) of the gene (Fig. 1). The probe for the 5'-end of the gene consisted of two DNA fragments (NJ-3' and NJ-3''), which contained 6.75 kb of genomic DNA extending downstream from the codon for the amino acid residue 19 of the α2(I) chain (Fig. 1).

RFLP associated with the human proα2(I) gene. To search for RFLP associated with the 3'-half of the proα2(I) gene, leukocyte DNA from 61 individuals was cleaved with 11 restriction endonucleases and hybridized with Hf-32. The 61 individuals were randomly selected among members of our laboratory staff and other volunteers. The number of chromosomes examined with each restriction endonuclease varied from 82 to 122. As indicated in Table I, no RFLP was detected with 10 of the restriction endonucleases. One individual had a polymorphic site for PstI. In individuals lacking the site, three PstI fragments of 5.7, 8.2, and 15.5 kb were generated. Restriction mapping of genomic clones for procα2(I) indicated that the PstI fragments hybridizing with Hf-32 are arranged in the 5'- to 3'-order of 5.7, 8.2, and 15.5 kb. Hf-32 itself contains a single PstI site (16), which corresponds to

![Figure 1: Schematic map of the proα2(I) gene of human type I procollagen and the DNA probes used here. The 5'-boundary of the gene has not been precisely defined but is ~38 kb from the 3'-end (10, 19). E and P indicate EcoRI and PstI sites, respectively. The locations of the two polymorphic sites studied here are indicated by the asterisks. The polymorphic site for PstI may or may not be in an exon. (E) indicates an artificial EcoRI site in the genomic probe.](image-url)
the site generating the 8.2- and 15.5-kb fragments. With DNA from the individual displaying the polymorphism, fragments of 4.7 and 3.5 kb were detected in addition to the fragments of 5.7, 8.2, and 15.5 kb. Therefore, it appeared that the RFLP in this individual consisted of an additional PstI site within the 8.2-kb fragment. Based on the data provided in Table 1, the frequency of the polymorphic site for PstI is \(~0.01\).

To search for RFLP associated with the 5'-end of the gene, similar experiments were carried out with the combined probe NJ-3' and NJ-3". A highly polymorphic site for EcoRI was detected at the 5'-end of the gene. DNA of individuals homozygous for the absence of the polymorphic site (\(-/-\)) generated one fragment of 14 kb (Fig. 2). DNA from individuals homozygous for the presence of the polymorphic site (+/+ ) generated two fragments of 10.5 and 3.5 kb. DNA from individuals heterozygous for the polymorphic site (+/- ) generated three fragments of 14, 10.5, and 3.5 kb. Thus, we identified three different genotypes. Analysis of DNA from a normal family demonstrated that the polymorphic site segregated as an autosomal dominant trait (Fig. 2). A total of 100 chromosomes were examined from randomly selected individuals, and the allelic frequency for the presence of the polymorphic site was 0.38. The three genotypes were in Hardy-Weinberg equilibrium with observed values close to the theoretical (\(P < 0.10\)).

To confirm the polymorphic nature of the EcoRI site, we digested DNA from individuals homozygous for the presence (+/+ ) or absence (\(-/-\) ) of the polymorphic site with other enzymes mapped in the same region. Single digestions with the restriction endonucleases HindIII and XbaI, and double digestions with the enzymes EcoRI and HindIII produced fragments with sizes consistent with the map distances of the recognition sites in the same region of the gene (19).

Analysis of a family with OI. The RFLP generated at the EcoRI in the 5'-region of the gene was used to study a family with an autosomal dominant form of OI. Leukocyte DNA was obtained from 18 individuals in four generations (Fig. 3). To a variable degree, affected individuals presented with fractures, dentinogenesis imperfecta, hearing loss, small joint laxity, and white sclerae. In generation I, the affected individual (I-2) was heterozygous (\(-+/+\) ) for the presence of the EcoRI site and the nonaffected (I-1) was homozygous for absence of the site (\(-/-\) ). In the subsequent three generations the phenotype cosegregated with a pro2(1) allele bearing the EcoRI recognition site. The lod score was 2.41 at a recombination fraction (\(\theta\) ) of 0, indicating strong linkage (21).

DISCUSSION

Several human genes have now been surveyed for DNA polymorphism, readily detectable as RFLP. In

<table>
<thead>
<tr>
<th>Number of chromosomes screened</th>
<th>Restriction endonuclease</th>
<th>Cleavage sites screened</th>
</tr>
</thead>
<tbody>
<tr>
<td>88</td>
<td>BamHI</td>
<td>2</td>
</tr>
<tr>
<td>94</td>
<td>BglI</td>
<td>3</td>
</tr>
<tr>
<td>86</td>
<td>BglII</td>
<td>5</td>
</tr>
<tr>
<td>122</td>
<td>EcoRI</td>
<td>4</td>
</tr>
<tr>
<td>106</td>
<td>HindIII</td>
<td>5</td>
</tr>
<tr>
<td>82</td>
<td>HpaI</td>
<td>3</td>
</tr>
<tr>
<td>94</td>
<td>KpnI</td>
<td>3</td>
</tr>
<tr>
<td>100</td>
<td>PstI</td>
<td>3-41</td>
</tr>
<tr>
<td>88</td>
<td>PvuII</td>
<td>4</td>
</tr>
<tr>
<td>88</td>
<td>SacI</td>
<td>6</td>
</tr>
<tr>
<td>84</td>
<td>XbaI</td>
<td>5</td>
</tr>
</tbody>
</table>

* Hybridization with Hf-32, a 2.2-kb cDNA (see Fig. 1).
1 Polymorphic site found in one chromosome examined.
FIGURE 3  Family with an autosomal dominant form of OI. Symbols: Individuals heterozygous 
(−/+ ) and homozygous for the absence (−/− ) and presence (+/+ ) of the polymorphic EcoRI 
site; horizontal bars, individuals examined by one of us (Dr. P. Tsipouras). (■ ●): affected by 
OI; (+): examined clinically; arrow: proband.

the cluster of β-globin genes, seven sites of relatively 
high frequency have been identified in ~60 kb of 
DNA (23). In addition, RFLP generated by either of 
two restriction endonucleases, DdeI or MstII, has been 
used to identify directly the mutation in the β'-globin 
gene (4, 5). In the case of the gene for human growth 
hormone, five polymorphic restriction sites with allelic 
frequencies ranging from 0.30 to 0.44 have been 
detected within ~40 kb of genomic DNA (7, 8). With 
the human insulin gene, extensive RFLP is found in 
the 5'-flanking region as a result of either insertions or 
deletions at a site ~800 bases upstream from the gene 
(24). The most common alteration is a 1.6-kb insertion, 
which has an allelic frequency of ~0.20 in normal 
individuals, and a significantly higher frequency in 
diabetics who are noninsulin dependent.

The probes used here made it possible to survey 
~25 kb of the proα2(I) gene for RFLP (Fig. 1). Only 
one polymorphic site in one individual was found with 
the cDNA Hf-32, even though this probe hybridized 
to fragments spanning over 12 kb of the gene and a 
total of 210 bases were surveyed with 11 restriction 
endonucleases. It is possible, however, that additional 
polymeric sites will be found as the same region is 
explored more extensively with additional restriction 
endonucleases or with genomic probes. In contrast, the 
polymeric EcoRI site in the 5'-half of the gene had 
the very high allelic frequency of 0.38.

The data developed here rigorously confirm previ 
sious indications that the proα2(I) gene is found in a 
single copy in the human haploid genome (11, 25, 26). 
All the fragments detected with the genomic probe 
were accounted for by the linear map developed from 
overlapping phage clones of the human proα2(I) gene 
(19). Since the intervening and flanking sequences of 
highly homologous genes, such as the two α-globin 
genes (27) are dissimilar, a different restriction pattern 
would have been expected if more than one copy of 
the proα2(I) procollagen gene existed per haploid ge 
nome. The three genotypes for the polymorphic EcoRI 
site are in equilibrium. Therefore it is likely that be 
cause of a founder effect, the mutation creating or 
abolishing the recognition site for the EcoRI was fixed 
in the human genome before any mutations producing 
-genetic disorders of collagen appeared.

The EcoRI site in the 5'-end of the proα2(I) gene 
made it possible to demonstrate linkage of an OI pheno 
type with the presence of a proα2(I) allele in four 
generations of a family with an autosomal dominant 
form of OI. Since a lod score of 2.41 is generally re 
garded as significant if derived from a single family 
(21), the data obtained here indicate linkage. The re 
-sults therefore provide presumptive evidence that the 
OI phenotype in this family is produced by a mutation 
in the proα2(I) gene or in some contiguous region of 
the genome.

It should be noted that use of RFLP to establish 
linkage is probably more important for examining ge 
etic disorders of connective tissue than for most other 
genetic diseases. A variety of observations suggest that
mutations in genes for type I procollagen are the cause of many forms of OI, of Ehlers-Danlos syndrome, and of Marfan syndrome (28-30). However, it has been difficult to generate definitive data about the molecular defects, because of the large size of both the genes and the proteins. Also, the assembly of the procollagen molecule involves at least 11 posttranslational enzymes, and defects in one or more of these can produce disorders of connective tissue. In addition, defects of genes for other components of connective tissue may well produce similar phenotypes. Therefore, data demonstrating linkage of clinical phenotypes to specific alleles should provide invaluable information for systematically defining the molecular basis of a variety of disorders of the extracellular matrix. It should also be noted that the relatively high frequency of the EcoRI RFLP associated with the proα2(I) gene can be used as molecular marker for linkage studies with other loci assigned on chromosome 7, a chromosome for which only a few markers are currently available. A partially characterized α1-like collagen gene with a HindIII RFLP has recently been located on chromosome 7 (31).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the expert technical assistance of Ms. Ann Pizanis.

This work was supported, in part, by National Institutes of Health grants Am-16516 and Am-06913 and by grants from the Hunterdon Health Fund, the Foundation of the University of Medicine and Dentistry of New Jersey, and the March of Dimes Birth Defects Foundation.

REFERENCES


