

Erin Dietzsch · M. Iqbal Parker

CA repeat polymorphism in the promoter region of the *COL1A2* gene

Received: May 6, 1999 / Accepted: July 13, 1999

Abstract A polymorphic CA dinucleotide repeat sequence has been identified within the promoter of the human $\alpha 2(I)$ procollagen gene, located at 7q21.3–q22.1. Nine alleles have been identified in unrelated individuals and the observed heterozygosity for the polymorphism was 0.66. This marker may be useful in the prenatal diagnosis of inherited connective tissue diseases in which the *COL1A2* gene is involved. Furthermore, it may potentially improve the usefulness of the *COL1A2* genetic system as an anthropogenetic marker.

Key words CA repeat polymorphism · *COL1A2* gene promoter

Introduction

The collagens are a large and diverse family of structural proteins, and type I collagen is one of the primary components of the extracellular matrix. Of the 19 known types of collagen, type I collagen is the most abundant and widely expressed collagen in vertebrates. The $\alpha 1(I)$ and $\alpha 2(I)$ chains of type I collagen are encoded by the *COL1A1* and *COL1A2* genes, respectively (Dalglish, 1997). *COL1A2* is approximately 38 kb in length, consists of 52 exons (Dalglish, 1997), and has been localized to 7q21.3–q22.1 (Retief et al. 1985). This article describes a CA repeat polymorphism located within the promoter region of the *COL1A2* gene.

Ethical approval for this project was obtained from the Ethics Committee of the Faculty of Health Sciences, University of Cape Town.

Source and description

During the sequence analysis of the 5' region of the $\alpha 2(I)$ procollagen gene (Akai et al. 1998), a novel dinucleotide repeat was observed at position –1,400. Polymerase chain reaction (PCR) primers were designed to amplify the region of interest.

Primer sequences

COL 1 (forward): 5'-AGAGCAGCACTCTAACTAGG-3'
COL 2 (reverse): 5'-GGGAGGCCAAAGGTCCTGTG-3'

PCR conditions

PCR reactions were performed in a volume of 25 μ l containing 50 ng genomic DNA, 100 ng of each primer, 125 μ M of each dNTP, 1.25 μ Ci [α^{32} P]-dCTP, 2U *Taq* polymerase (Roche Molecular Biochemicals, previously known as Boehringer Mannheim, Germany), and *Taq* buffer. Amplification conditions consisted of an initial denaturation at 95°C for 5 min, followed by 35 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 45 s. Final extension was at 72°C for 10 min. PCR products of approximately 200 bp were resolved on denaturing 6% polyacrylamide gels and visualized by autoradiography.

Polymorphism and allele frequency

A total of nine alleles was detected. The allele frequencies were estimated from 47 unrelated individuals, on 94 chromosomes (Table 1). The observed heterozygosity was 0.66 and the polymorphic information content (PIC) 0.807. Codominant segregation was demonstrated in three three-generation CEPH (Centre d'Etude du Polymorphisme Humain) families, namely, family numbers 1349, 1362 and 1377.

E. Dietzsch · M.I. Parker (✉)
Department of Medical Biochemistry, CANSA/MRC Oesophageal Cancer Research Group, Faculty of Health Sciences, University of Cape Town, Observatory 7925, South Africa
Tel. +27-21-406-6259; Fax +27-21-447-3762
e-mail: mparker@physio.uct.ac.za

Table 1. Allele frequencies of the CA repeat in the COL1A2 promoter

| Allele | Size (bp) | Frequency |
|--------|-----------|-----------|
| A1 | 193 | 0.01 |
| A2 | 195 | 0.17 |
| A3 | 197 | 0.07 |
| A4 | 199 | 0.28 |
| A5 | 201 | 0.09 |
| A6 | 203 | 0.2 |
| A7 | 205 | 0.05 |
| A8 | 207 | 0.1 |
| A9 | 209 | 0.03 |

Several polymorphisms in *COL1A2* have been reported, few of which have been characterized in detail. In addition to silent base substitutions in the coding region of the gene (Dalglish 1997), restriction fragment length polymorphisms (RFLPs) (Mottes et al. 1989; Pepe et al. 1994, 1995), a 38-bp length variant (Watson and Dalglish 1990), a variable nucleotide of tandem repeats (VNTR; Pepe 1993), and a GT dinucleotide repeat (Chi et al. 1992) have been described. The microsatellite described by Chi et al. (1992) consists of at least 20 repeats and is located in the first intron, at position +1400. No recombination was observed between the promoter polymorphism described here, and the first intron polymorphism in the three CEPH families mentioned above.

Defects in either *COL1A1* or *COL1A2* cause dominant osteogenesis imperfecta (OI) or Ehlers-Danlos syndrome (EDS) types VIIA and VIIB. Multiple amino acid substitutions resulting in OI, as well as exon skipping mutations giving rise to OI or EDS, have been documented (Dalglish 1997). Prenatal diagnoses of these genetic disorders relies on identifying the *COL1* gene that is affected. This is achieved by excluding the unaffected gene, and determining the haplotype associated with the disorder. In most instances, both these requirements are achieved using polymorphic markers (Pepe et al. 1994), including the VNTR in intron 12 (Pepe 1993) and the dinucleotide repeat in intron 1 (Chi et al. 1992).

Polymorphisms of the two genes encoding the type I collagen peptides have proved to be valuable anthropogenetic markers (Pepe et al. 1995). The polymorphism described here will improve the usefulness of the *COL1A2* genetic system as an anthropogenetic marker, enabling the reconstruction of genetic relationships among different human populations. In addition, the identification of a highly polymorphic marker in this region will improve prenatal diagnosis of disorders such as OI and EDS.

Acknowledgments This study was supported by grants from the University of Cape Town. The authors are members of the Oesophageal Cancer Research Group, funded by the Medical Research Council and Cancer Association of South Africa. We thank L. Warnich for providing DNA of CEPH families, and M.J. Kotze for providing DNA of normal control individuals.

References

- Akai J, Kimura A, Arai K, Uehara K, Hata R-I (1998) Fine structural analysis of the unique 5' region of the human *COL1A2* gene containing two regions of dinucleotide repeats adjacent to the transcriptional start site. *Connect Tissue Res* 30:1-6
- Chi DD, Helms C, Steinbrueck, Mishra SK, Donis-Keller H (1992) Two chromosome 7 dinucleotide repeat polymorphisms at gene loci epidermal growth factor receptor (EGFR) and pro α 2(I) collagen (*COL1A2*). *Hum Mol Genet* 1:135
- Dalglish R (1997) The human collagen mutation database. *Nucleic Acids Res* 25:181-187
- Mottes M, Cugola L, Pignatti PF (1989) Haplotype frequencies of the collagen type-I genes in the Italian population. *Hum Genet* 83:369-372
- Pepe G (1993) A highly polymorphic (ACT)_n VNTR (variable nucleotide of tandem repeats) locus inside intron 12 of *COL1A2*, one of the two genes involved in dominant osteogenesis imperfecta. *Hum Mutat* 2:300-305
- Pepe G, Bue C, Orlandi P, Ena F, Meloni T, Modiano G (1994) *COL1A2* gene (α 2 gene of type I collagen) at the haplotype level as a new valuable anthropogenetic marker: a study on Sardinians. *Hum Biol* 66:613-623
- Pepe G, Rickards O, Jodice C, Modiano G (1995) Allele and haplotype frequency distribution of the *EcoRI*, *RsaI* and *MspI* *COL1A2* RFLPs among various human populations. *Hum Biol* 67:905-920
- Retief E, Parker MI, Retief AE (1985) Regional chromosome mapping of human collagen genes alpha 2 (I) and alpha 1 (I) (*COL1A2* and *COL1A1*). *Hum Genet* 69:304-308
- Watson LW, Dalglish R (1990) PCR detection of a 38 bp length variant in the *COL1A2* gene. *Nucleic Acids Res* 18:5925