Refinement of human chromosome 7 map around the proalpha2(I)collagen gene by long-range restriction mapping

Juha Kere*, Ritva Tolvanen, Helen Donis-Keller¹ and Albert de la Chapelle

Department of Medical Genetics, University of Helsinki, 00290 Helsinki, Finland and ¹Department of Genetics, Washington University School of Medicine, St Louis, MO 63110, USA

Received December 21, 1990; Revised and Accepted April 15, 1991

ABSTRACT

The physical proximity of the closely linked proalpha2(I)collagen (COL1A2) and erythropoietin (EPO) genes and five loci with no known function was studied by long-range restriction mapping experiments using pulsed-field gel electrophoresis. COL1A2 and D7S64 were found to be within 100 kb of each other, providing a new informative marker for linkage studies with respect to COL1A2. D7S15 and D7S79 were within 350 kb of each other. The physical distance between COL1A2 and EPO was determined to be at least 600 kb. Two CpG rich islands were recognized within 600 kb of COL1A2, suggesting that other genes might lie in the vicinity of COL1A2.

INTRODUCTION

Chromosome 7 is one of the most extensively studied human autosomes, being covered with cloned markers which form a densely spaced genetic linkage map (1-3). Detailed fine restriction maps of the proalpha2(I)collagen gene (COL1A2) located in 7q21.3-q22.1 are available (4), whereas long-range restriction maps around COL1A2 have not been published. Moreover, the location of COL1A2 in the linkage map is somewhat ambiguous (1,2). The erythropoietin gene (EPO) maps to the same chromosomal region and is very closely linked with COL1A2 ($Z=4.81$ at $\theta_{\max}=0$) (5,6), but the physical distance between these genes is not known. Of interest in this region is also the localization of a recurrent leukemia-associated chromosome 7 deletion breakpoint close to these genes (7,8).

In this study we performed physical mapping of the region around COL1A2 and EPO by pulsed-field gel electrophoresis (PFGE), with the aims of 1) refining the map of chromosome 7 by physical means, and 2) determining physical distances between a number of genetically closely linked markers in this region.

MATERIALS AND METHODS

The mapping experiments were performed on DNA prepared from freshly isolated white blood cells from a healthy male. Red blood cells were removed by dextran sedimentation, and the plasma containing white blood cells was recovered. The cells were washed with saline, embedded in low-melting point agarose plugs, and deproteinized as described (8,9). To facilitate double digestions, most restriction digestions were performed in potassium glutamate buffer (BssHII, MluI, NaeI, NarI, NruI, PaeR7I, SacII, SfiI) (10). Buffers recommended by the manufacturers were used with enzymes requiring high salt concentrations (EagI, NotI, SalI).

The electrophoreses were run in workshop-made tanks designed according to the contour-clamped homogeneous electric field (CHEF) principle (11). Agarose gels (1%, run in $0.25 \times$ Tris-borate electrophoresis buffer) were calibrated with *Saccharomyces cerevisiae* DNA preparations or bacteriophage lambda concatamers applied to several lanes in each gel (9). The gels were blotted to nitrocellulose filters after photography and UV irradiation. The probes were labelled with ³²P-dCTP (Oligolabelling kit, Pharmacia, Uppsala, Sweden, or Nick-translation kit, Amersham, Buckinghamshire, England). Probes giving high background hybridization were prereassociated with sheared, unlabelled human DNA (12). Between repeated hybridizations to the same filter, the probes were removed by washing several times with boiling water.

The probes used in the mapping studies are listed in Table 1.

RESULTS

Detection of physical linkage between markers

Physical linkage between the markers listed in Table 1 was sought by hybridizing them to CHEF filters containing BssHII, MluI, and NotI digested samples. These enzymes were selected because in preliminary experiments they were found to create the longest repeatedly well-cut fragments with probes for COL1A2 and EPO.

* To whom correspondence should be addressed at Department of Molecular Microbiology, Washington University School of Medicine, Box 8230, St Louis, MO 63110, USA

Probes for COL1A2, D7S15, D7S64, and D7S79 seemed to detect at least one shared fragment, whereas the other probes hybridized to different fragments with all three enzymes (Table 2). As a more rigorous test of physical closeness, the putative linked probes were hybridized to CHEF filters with DNA from several different single and double restriction digested samples. The smallest common fragment detected by probes for COL1A2 and D7S64 was < 100 kb, but SfiI produced additional fragments hybridizing to the D7S64 probe above 100 kb (Fig. 1). Probes for D7S15 and D7S79 hybridized to identically sized fragments with several enzymes (the smallest common fragment was a 350 kb SfiI fragment) (Fig. 1). All these fragments except a 1100 kb NotI fragment were of different size than the fragments detected with probes for COL1A2 and D7S64.

Construction of long-range restriction maps

Altogether 11 restriction enzymes were used in double digestion experiments, but only some of them yielded information which allowed the placing of their cleavage sites with respect to each other. The restriction enzymes NotI, MluI, and BssHII cut DNA

into large (> 800 kb) fragments and also yielded double digestion and partial digestion products that allowed the construction of long-range restriction maps around the linked probes. SalI, NarI, and NruI either cut only some of the available molecules or the products were too large to be resolved under our experimental conditions. PaeR7I fragments were generally < 200 kb in size, and the sites were densely spaced, as illustrated by several partial digestion bands. These results did not help in map construction.

The core of the long-range restriction map around COL1A2 was based on the following observations (Fig. 2, Table 2). COL1A2 hybridized to distinct 820 kb BssHII and 1100 kb MluI fragments. BssHII/MluI double digestion created an additional 790 kb fragment, which placed these two restriction sites at a distance of about 30 kb. NotI easily produced partial digestion, with COL1A2 hybridizing to fragments of 1100, 1200, and 1300 kb. BssHII/NotI double digestion produced a novel fragment of 790 kb, and MluI/NotI double digestion produced two novel fragments of 770 kb and 900 kb (partial digestion). The double digestion products allowed the unambiguous placement of the BssHII, MluI, and NotI sites with respect to each other, and NotI partial digestions allowed the orientation of fragments flanking the COL1A2 NotI fragment.

Similar principles were applied in the construction of maps around EPO, D7S15, and D7S79, using data presented in Fig. 1 and Table 2.

At the construction of the restriction maps it became evident that the 1100 kb NotI fragments detected by probes for COL1A2 or D7S64 and by probes for D7S15 or D7S79 were different although similar in size. Further evidence against the maps overlapping was obtained by double digestion results with other restriction enzymes (not shown). The pattern of restriction sites at the borders of the maps did not suggest that they are directly juxtaposed but did not exclude it either.

Table 1. Characteristics of the DNA probes. Probe type: G, genomic; C, cDNA.

Locus	Probe	Type	Reference
COL1A2	NJ-1	G (3')	4
COL1A2	NJ-3	G (5')	4
EPO	epo	C	5
D7S15	CRI-L917	G	1,2
D7S64	CRI-L1238	G	1,2
D7S79	CRI-S25	G	1,2
D7S82	CRI-S56	G	1,2
D7S90	CRI-S130	G	1,2

Table 2. Restriction fragments detected by genetically linked probes in single and double digestions. Fragment sizes are in kb. Figures in parentheses denote weak or partial digestion fragments. X, double digestion created no novel fragments. Enzyme names are shortened to three first letters.

Restriction enzyme	5'-COL1A2-3' NJ-3	NJ-1	EPO epo	D7S15 L917	D7S64 L1238	D7S79 S25	D7S82 S56	D7S90 S130
Bss	820	820	≈ 100 (180)	780	820	780	240	580
Mlu	1100		230 (320,490)	1000			850	≈ 900–950
Nae	400(530, 660,820)	400(530, 660,820)	150(220)	780	400(530, 660,820)	780	250	600
Not	1100 (1200,1300)		200	1000				≈ 600
Sac	460 (520,610)	460 (520,610)	200	900–1000	460 (520,610)			
Sfi	70 (100)	110	100	350	70–110	350		
Bss + Mlu	790	790		X	790			
Bss + Nae	530 (660)	530 (660)	X	X	530 (660)	X		
Bss + Not	790		X	X		X		
Bss + Sac	620	620		X	620	X		
Bss + Sfi	X	X		X	X	X		
Mlu + Nae	620	620	X	X	620	X		
Mlu + Not	770 (900)	770 (900)		900	770 (900)			
Mlu + Sac	610	610			610			
Mlu + Sfi	X	X						
Nae + Not	X				X			
Nae + Sac	340 (560)	340 (560)			340 (560)			
Nae + Sfi	X	X		X	X	X		
Not + Sac	X							
Not + Sfi	X	X	X	X		X		
Sac + Sfi	X	70						

EPO was found to reside in a region rich in rare cutter restriction sites, and no total or partial digestion restriction fragments shared with any other probe were detected (Fig. 3). A minimum distance of about 600 kb between COL1A2 and EPO could be inferred (Fig. 4).

Orientation of COL1A2

The use of the restriction enzymes BssHII, MluI, NaeI, NotI, SacII, and SfiI allowed the orientation of COL1A2 within the long-range map (Fig. 2). The orientation was provided by the BssHII, MluI, and NotI sites which have a different pattern at each end of the COL1A2-containing segment. The pattern and sizes of NaeI and SacII restriction fragments detected by probes for COL1A2 strongly suggested the interpretation presented below.

NaeI produced partial digestion patterns which were repeatedly observed in different blots. The partial digestion products were

modified in double digestions with BssHII and MluI. Double digestion with BssHII strongly intensified all partial NaeI digestion bands, and NaeI/MluI double digestion created a new fragment which was 30–40 kb smaller than the 660 kb NaeI partial digestion band. Thus, a NaeI site which gave rise to this partial digestion band was likely to be close to a BssHII site which is 30–40 kb away from an MluI site. These sites were readily identified in the core map (Fig. 2). Furthermore, NaeI/NotI double digestion did not create novel bands but intensified a third partial digestion band which was very weak in most digestions and most likely resulted from the incomplete cleavage of a NaeI site at the other end of the fragment. Digestion with SacII resulted in an incomplete digestion pattern as well, and double digestion with NaeI created a novel 340 kb NaeI/SacII fragment which hybridized to probes for COL1A2. This fragment helped in the localization of the SacII and NaeI sites in relation to each other. SacII/MluI double digestion resulted in the intensification of a

Figure 1. Detection of physical linkage. The sizes of restriction fragments are given in Table 2. A and B, a filter was successively hybridized to probes CRI-L1238 (A) and NJ-3 (B). All but SfiI fragments were identical. PFGE was at 120 V and 120 s pulse time for 60 h. C and D, a filter was successively hybridized to probes CRI-L917 (C) and CRI-S25 (D). Most single and double digestion fragments were identical, but the loci mapped to different PaeR7I fragments. PFGE was at 85 V and 120 s pulse time for 90 h.

Figure 2. PFGE data used for the construction of a restriction map around COL1A2. A, probe NJ-3 hybridized to large BssHII, MluI, and NotI fragments which were used to construct a core map. PFGE was at 85 V and 900 s pulse time for 140 h. The film was overexposed to show partial digestion products. B and C, probes NJ-1 (B) and NJ-3 (C) were hybridized to filters with NaeI and SacII partial and double digestion products which allowed the orientation of these sites with respect to the core map. D and E, a filter was successively hybridized to probes NJ-3 (D) and NJ-1 (E). The 5' and 3' ends of COL1A2 were found to be separated by an SfiI site, and a SacII site was localized within 70 kb of the 3' end. PFGE was at 130 V and 20 s pulse time for 40 h.

weak partial digestion product, whereas SacII/BssHII double digestion produced a novel partial digestion fragment which was 20–30 kb larger than the SacII/MluI fragment. Double digestion with NotI did not modify the SacII digestion products. This pattern was compatible with a SacII site located close to an MluI site and 20–30 kb away from a BssHII site, and oriented the SacII partial digestion products in relation to these sites.

The 5' and 3' ends of COL1A2 were found to be separated by an SfiI site, as the probes NJ-3 and NJ-1 detected fragments of different sizes in SfiI digests (but in no other digest). The SfiI fragment detected by NJ-1 was cut by double digestion with SacII, but not with BssHII, MluI, NaeI, or NotI, indicating the presence of a SacII site within 70 kb 3' of the sequence detected by NJ-1. These results suggested a likely location and orientation for COL1A2 within the NaeI and SacII map (Fig. 2, Fig. 4).

Figure 3. A filter was successively hybridized to the EPO probe (A) and probe NJ-3 (B). No shared total or partial digestion fragments were detected. PFGE was at 130 V and 150 s pulse time for 65 h.

Figure 4. Restriction maps around COL1A2, EPO, D7S15, D7S64, and D7S79. Restriction sites identified by partial digestions only are drawn with a thin line. The probable order of the loci is cen-[D7S15, D7S79]-[COL1A2, D7S64]-EPO-tel (1,2,8). A, NaeI; B, BssHII; E, EagI; F, SfiI; M, MluI; N, NotI; S, SacII. The scale is in kilobases. Thin horizontal bars under the locus names denote uncertainty as to their localization.

DISCUSSION

We used pulsed field gel electrophoresis (PFGE) to construct restriction maps of the region around COL1A2 in chromosomal region 7q21.3-q22.1 and to supplement the existing genetic linkage maps. The restriction maps were constructed using the same DNA in all experiments. Strictly, the maps are valid only in the DNA used for mapping, as there may be variation in fragment lengths between different individuals or tissues (13,14).

The position of COL1A2 within the contiguous genetic linkage map of chromosome 7 had not been unambiguously determined (1,2). The physical mapping data presented here resolve this ambiguity, provide information of minimum distance between COL1A2 and EPO, and allow the identification of CpG rich islands flanking COL1A2.

COL1A2 and D7S64 were found to lie very close to each other, hybridizing to the same restriction fragments in all but SfiI digests. This result places them within <100 kb of each other, providing a physical anchor site for the chromosome 7 linkage map (1,2). As the probe for D7S64 detects an RFLP with PIC 0.39, it can be used for genetic linkage studies in individuals who are homozygous for the less informative COL1A2 polymorphisms (4). Proximal of D7S64 in linkage maps, D7S15 and D7S79 were found to be located within 350 kb of each other. This is compatible with the genetic distance (0 cM, $Z=18.66$)(2). We have previously suggested that COL1A2, D7S64, D7S15, and D7S79 are at most 1.4 Mb apart (15), but further experiments with other restriction enzymes revealed incompatibilities in the physical map, suggesting that they may not be so closely linked. The recombination fraction between is 0.02 ($Z=22.30$) between D7S64 and D7S15 and 0.01 ($Z=22.31$) between COL1A2 and D7S15, whereas no recombinations have been observed between COL1A2 and D7S79 ($Z=14.45$)(2 and HDK, unpublished). These figures are compatible with measurable physical distances which however remain undetermined.

The short genetic distance between COL1A2 and EPO suggested that physical linkage might also be detectable between them. In our study they were not found to share any common total or partial digestion restriction fragments using 11 different rare cutter enzymes. The restriction maps imply a minimum distance of approximately 600 kb between them (Fig. 4). This may not be unexpected, as the confidence interval for the genetic linkage is still very broad at $Z=4.81$. Our earlier results on the mapping of a partially deleted chromosome 7 suggested that COL1A2 might be located centromeric of EPO (8).

The clustering of rare cutter restriction sites which occurs in CpG rich islands marks the 5' end of many genes (14). At least two such clusters were detected within 600 kb of COL1A2, but the 5' end of COL1A2 was not marked with densely located restriction sites (Fig. 4). The presence of CpG rich islands might suggest that there are other genes in the vicinity of COL1A2.

Physical mapping is likely to benefit from the use of yeast artificial chromosomes (YACs)(16). The construction of restriction maps of YAC DNA is much simpler than of genomic DNA for several reasons. The ends of a YAC can be used as reference points, the risk of interpreting polymorphisms or partially homologous sequences as partial digestions is diminished, and the masking of sites by methylation is avoided. Furthermore, overlapping YAC clones can be identified by YAC to YAC hybridization, allowing chromosome walking. We are currently isolating YACs covering this region, which should allow comparisons between YAC maps and maps constructed by PFGE on genomic DNA.

ACKNOWLEDGEMENTS

We thank Dr. Petros Tsipouras and Dr. Paul C. Watkins for COL1A2 and EPO probes. This work was supported by the Academy of Finland, the Sigrid Juselius Foundation, the Finnish Cultural Foundation, Finska Läkaresällskapet, the Paulo Foundation, and NIH grants HG00304 and HG00201 (HD-K). Part of the study was done at the Folkhälsan Institute of Genetics.

REFERENCES

1. Barker D, Green P, Knowlton R, Schumm J, Lander E, Oliphant A, Willard H, Akots G, Brown V, Gravius T, Helms C, Nelson C, Parker C, Rediker K, Rising M, Watt D, Weiffenbach B, Donis-Keller H (1987) *Proc Natl Acad Sci USA* **84**, 8006–8010.
2. Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES, Botstein D, Akots K, Rediker KS, Gravius T, Brown VA, Rising MB, Parker C, Powers JA, Watt DE, Kauffman ER, Bricker A, Phipps P, Muller-Kahle H, Fulton TR, Ng S, Schumm JW, Braman JC, Knowlton RG, Barker DF, Crooks SM, Lincoln SE, Daly MJ, Abrahamson J (1987) *Cell* **51**, 319–337.
3. Lathrop GM, O'Connell P, Leppert M, Nakamura Y, Farrall M, Tsui L-C, Lalouel J-M, White R (1989) *Genomics* **5**, 866–873.
4. Borresen AL, Moller P, Berg K (1988) *Hum Genet* **78**, 216–221.
5. Watkins PC, Eddy R, Hoffman N, Stanislovits P, Beck AK, Galli J, Vellucci V, Gusella JF, Shows TB (1986) *Cytogenet Cell Genet* **42**, 214–218.
6. Klinger KW, Winqvist R, Riccio A, Andreasen PA, Sartorio R, Nielsen LS, Stuart N, Stanislovits P, Watkins P, Douglas R, Grzeschik K-H, Alitalo K, Blasi F, Dano K (1987) *Proc Natl Acad Sci USA* **84**, 8548–8552.
7. Kere J, Ruutu T, Davies KA, Roninson IB, Watkins PC, Winqvist R, de la Chapelle A (1989) *Blood* **73**, 230–234.
8. Kere J (1989) *Nucl Acids Res* **17**, 1511–1520.
9. Schwartz DC, Cantor CR (1984) *Cell* **37**, 67–75.
10. McClelland M, Hanish J, Nelson M, Patel Y (1988) *Nucl Acids Res* **16**, 364.
11. Chu G, Vollrath D, Davis RW (1986) *Science* **234**, 1582–1585.
12. Sealey PG, Whittaker PA, Southern EM (1985) *Nucl Acids Res* **13**, 1905–1922.
13. Fulton TR, Bowcock AM, Smith DR, Daneshvar L, Green P, Cavalli-Sforza LL, Donis-Keller H (1989) *Nucl Acids Res* **17**, 271–284.
14. Bird AP (1986) *Nature* **321**, 209–213.
15. Kere J, Donis-Keller H, de la Chapelle A (1989) *Cytogenet Cell Genet* **51**, 1022.
16. Burke DT, Carle GF, Olson MV (1987) *Science* **236**, 806–812.