Interruption of a human nuclear sequence homologous to mitochondrial DNA by a member of the *KpnI* 1.8 kb family

Hisayuki Nomiya, Teruhisa Tsuzuki, Shoji Wakasugi, Makoto Fukuda and Kazunori Shimada

Department of Biochemistry, Kumamoto University School of Medicine, Kumamoto 860, Japan

Received 9 May 1984; Accepted 7 June 1984

ABSTRACT

We reported that several DNA sequences homologous to mitochondrial DNA (mtDNA) are present in the human nuclear genome (Tsuzuki *et al.* (1983) *Gene* 25, 223-229). Detailed Southern blot analyses revealed that one of such sequences is interrupted by a repetitive sequence about 1.8 kb long, and that the insert is one member of the dispersed repeated DNA sequences of the *KpnI* 1.8 kb family. Nucleotide sequence analysis showed that the *KpnI* 1.8 kb DNA is flanked with imperfect 15-base pair (bp) direct repeats of mtDNA. This *KpnI* 1.8 kb DNA has an A-rich sequence at its 3'-end, and has a considerable homology with one of the published cDNA sequences homologous to one of the human *KpnI* families and also to one of the African green monkey *KpnI* families, *KpnI*-LS1. These structural features suggest that the *KpnI* 1.8 kb DNA is a movable element and is inserted within the mtDNA-like sequence by an RNA-mediated process.

INTRODUCTION

Recent studies revealed that the nuclei of human cells¹ and other organisms²⁻⁷ contain sequences homologous to mitochondrial DNAs (mtDNAs). This would suggest that with evolution, exchange of genetic information occurred between nuclear and mtDNAs.

Complete nucleotide sequence of the human mtDNA has been determined by Anderson *et al.*⁸. The human mtDNA is organized in an extremely economical fashion and carries relatively few intergenic nucleotides and no introns. Therefore, by analyzing the molecular structure of each one of the mtDNA-like sequences present in the human nuclear DNA and comparing findings with those in the corresponding regions of mtDNA, information can be obtained on the exchange of genetic materials naturally occurring between mitochondria and nuclei in human cells.

We cloned several human nuclear DNA sequences homologous to the mitochondrial ribosomal RNA (rRNA) genes and analyzed the

structures of some of these sequences, in hybridization experiments¹. We now report that one of these clones carries long, dispersed repeated DNA sequences of the KpnI 1.8 kb family within a mtDNA-like sequence. Nucleotide sequence analysis revealed that the inserted KpnI 1.8 kb DNA is flanked with duplicated mtDNA-like sequences. These results show the mobility of the KpnI 1.8 kb family within the human cells, and suggest the extensive plasticity of the human nuclear DNA.

MATERIALS AND METHODS

We isolated and characterized the phage clone, Lm A-1 and its subclones¹. Plasmid clones carrying DNA fragments derived from the authentic KpnI families, each one of which is inserted at a KpnI site of pBK5⁹, were obtained from Dr. J. J. Maio. Human nuclear DNA was prepared from human placenta, as described by Blin and Stafford¹⁰ and mtDNA from human placenta was purified by the method of Drouin¹¹. Bidirectional Southern blots to two nitrocellulose filters were done as described¹². Prehybridization, hybridization and washing conditions were as described previously¹ except that dextranulphate was omitted from the hybridization buffer. Nick translation was performed according to the method of Rigby *et al*¹³. Nucleotide sequences were determined by the method of Maxam and Gilbert¹⁴, and were analyzed by the GENAS system¹⁵ at Kyushu University Computer Center. Restriction enzymes and T4 polynucleotide kinase were obtained from Takara Shuzo Co., Japan, and calf intestine phosphatase was from Boehringer Mannheim. (γ -³²P)ATP and (α -³²P)dCTP (3,000 Ci/mmol) were purchased from Amersham.

Propagation of E. coli cells carrying recombinant plasmids was carried out in accordance with the guidelines for recombinant DNA research issued by the Ministry of Education, Science and Culture, Japan.

RESULTS

Integration of a Repetitive Sequence, KpnI 1.8 kb DNA, within a Nuclear mtDNA-Like Sequence

We isolated several phage clones carrying nuclear DNA sequences which have an extensive homology to human mtDNA se-

Fig. 1. Restriction map of the DNA fragment carried with a phage clone, Lm A-1. The nuclear DNA is indicated by a solid line and the regions containing repetitive sequences are indicated by open and solid boxes. The regions indicated by solid boxes correspond to *Kpn*I 1.8 kb DNA sequence (Fig. 2 and see text). The nuclear mtDNA-like regions are indicated by hatched boxes. The two lines above the map represent subcloned fragments, #212 and #1, respectively. These fragments were used in Southern blot analyses (Figs. 2 and 4) and nucleotide sequencing. The flanking regions of the *Kpn*I 1.8 kb DNA are enlarged and arrows under the enlarged maps represent the sequencing strategy. Each arrow indicates the length and direction of the sequences determined by the method of Maxam and Gilbert¹⁴. Restriction sites: *Eco*RI (E), *Bam*HI (B), *Bgl*II (Bg), *Hind*III (H), *Mlu*I (M), *Fok*I (F), *Dde*I (D).

quence, from a human gene library by screening with a cDNA clone of human mitochondrial 16S rRNA, pHIG-2¹⁶, as a probe¹. One of these clones, named Lm A-1 carries a mtDNA-like sequence extending from the mitochondrial rRNA gene to the unidentified reading frame 2 (URF2) regions and two different kinds of dispersed repeated nuclear DNA sequences. Southern blot analysis of this clone led to detection one of the repeated sequences within the mtDNA-like sequence (Fig. 1). From data obtained in foregoing hybridization experiments¹, we speculated that the left-hand junction of the repeated sequence is located near *Bam*HI site, as shown in Fig. 1. The results shown in Fig. 2A indicate that the 1.6 kilobase pairs (kb) *Bam*HI/*Mlu*I fragment derived from #212 DNA contains a highly repetitive nuclear sequence (panel b) and that the adjacent 0.6 kb *Mlu*I/*Eco*RI fragment hybridizes only with mtDNA (panel c). Therefore, the

Fig. 2. Localization and identification of the insert in the nuclear mtDNA-like sequence. A, *Eco*RI digested human nuclear DNA (lane 1) and *Xba*I digested human mtDNA (lane 2) were separated on a 0.8% agarose gel (panel a), transferred bidirectionally to two nitrocellulose filters and hybridized with 32 P-labelled *Bam*HI/*Mlu*I (panel b), or *Mlu*I/*Eco*RI fragments (panel c). These fragments were prepared from a subcloned #212 DNA. B, Clones containing various *Kpn*I families inserted at a *Kpn*I site of pBK5 were digested with *Kpn*I, separated on a 0.8% agarose gel (panel a) and transferred to a nitrocellulose filter. The filter was then hybridized with a 32 P-labelled *Bam*HI/*Mlu*I fragment, prepared as described above (panel b). Lanes 1, pBK(1.2)11; 2, pBK(1.5)54; 3, pBK(1.8)11; 4, pBK(1.8)20; 5, pBK(1.8)26.

right-hand junction of the repeat is probably around the *Mlu*I site (Fig. 1) and the size of the insert is about 1.6 kb.

There are various kinds of long, dispersed repeated DNA

sequences in the human genome, and the KpnI families have been most extensively studied^{9,17-22}. The individual KpnI family consists of different lengths of DNA units, each of which ranges from 1.2 kb to over 6 kb in length and does not cross-hybridize. The data shown in Fig. 2B indicate that only the KpnI 1.8 kb DNA cloned in pBK(1.8)11⁹ specifically hybridizes with the DNA segment present in the nuclear mtDNA-like sequence.

Structures of the Terminal Regions of the KpnI 1.8 kb DNA

To gain insight into the mobility of the KpnI 1.8 kb family, nucleotide sequences of the terminal regions of the inserted KpnI 1.8 kb DNA together with that of the flanking mtDNA-like regions were determined (Fig. 3). Comparison of these mtDNA-like sequences with that of the corresponding regions of human mtDNA⁸ revealed that they share 73% homology and that the region homologous to the URF2 DNA is interrupted by a DNA segment of 1.8 kb in length (Fig. 3). The insert DNA extends six base pairs (bp) upstream from the BamHI site to 190 bp downstream from the MluI site (Fig. 3). Contrary to the results shown in Fig. 2A (panel c), the MluI/EcoRI fragment apparently contains the 190 bp 3'-end region of the insert. Accordingly, we looked to see whether the 3'-end region of the insert is homologous to the KpnI 1.8 kb family. The subcloned #212 DNA was digested with MluI and hybridized with the ³²P-labelled pBK(1.8)11 DNA (Fig. 4). This experiment confirmed that the MluI/EcoRI fragment is homologous to the KpnI 1.8 kb DNA. Recently, DiGiovanni *et al*²⁰ reported the isolation of the three different cDNA clones carrying DNA sequences complementary to the human KpnI families. The 5'-terminal region of the KpnI 1.8 kb DNA described here shows considerable sequence homology to one of their cDNA clones, pcD-KpnI-4 which carries 1.3 kb length of the cDNA insert (Fig. 3). This region is also homologous both to the sequence of one of the African green monkey KpnI families (KpnI-LS1)²³ (positions 565-761, 1669-1756) and to that of a human KpnI family present in a rearranged KpnI cluster²² (positions 3486-3566, 2355-2441). Interestingly, this conserved region terminates in an A-rich sequence (Fig. 3). However, we detected no significant sequence homology between the remaining 470 bp length DNA of the 3' terminal region of our KpnI 1.8 kb

FokI

CATCCTGAAT ATAAGCACTA TGACCCTATC ACTATTCTCA CACCTGAAC AAATGACAT G^{*}taactgga
 A T A C TCC C C CT - G GC A

tccttcctt^{*} at^{*}atcttata caaaaattaa ttcaatatg^{*}g attaaagact taaatgttac^{*} acctaaaacc
 ataaaaacc^{*} cagaagaaaa ctaggcaat accattcagg acataggcat gggcaag^{*}t^{*} ttcatgact^{*}a
 aaacacaaa^{*} agcaatgg^{*}ca acagaag^{*}c^{*} aaattgaca^{*} atgagatct.....(1.0 kb)....

BglIII

.....gcacacca^{*} acatggcaca
 tgtatacata tgtaacaaac nnnacggtg tgcacatgta ccctagaact taaaga^{***} ata ataaaaaaa
 ttagtactca gaattatttcg aaacaaacca cccagtagaa aattaggcaa aatacatgag cagccaacta
 tgaaaaatga aatcaaatga taagaaagat gatcagcctc actagtatag tcatgtgctc cataacaatg
 ttttggtcaa tgagggaacca catatatgac agtgggtctca tagattataa tggagctcat aagattataa
 tggagccgaa aaatcccat tgcctagtga ctttgtagtt atcctgacat agccatggaa atgtagcaca
 acgcggtgggt gtggtgatac tgggtgaaac acacctactg tgctgccagt cctatcaaag tatagcaaat
MluI
 acaatgatgt acagtacata atacttgata atgataataa atgaccatat tactgggtata cgtatttatg
 atgctttgga atattatttt acagtgtact tcttctactt attaaaa^{AA A-AAGCTAAC-T} GTTAATACC
 C A AC C
 TATAATTCCA TTAATTCTAT TATCTCTAGG AGGTTTACCC CCATTAACAG GATTCTTACC TAAAT....
 CT CC CC CC C C CC G GC C C T G C

Fig. 3. Nucleotide sequence of the flanking regions of the KpnI 1.8 kb DNA (lower case letters) present within the nuclear mtDNA-like sequence (upper case letters). The line below the mtDNA-like sequences indicates a part of the URF2 sequence of human mtDNA (position 5126-5259)⁸. Only the bases that differ from the mtDNA-like sequence are shown in this line. Each hyphen (-) indicates a base that is missing in the mtDNA-like sequence or mtDNA sequence. The letter 'n' denotes an unidentified nucleotide. Homologous regions between KpnI 1.8 kb DNA and pcD-KpnI-4 (positions 50-198, 1159-1246)²⁰ are underlined and the asterisks above the KpnI 1.8 kb DNA show the nucleotides differing from the cDNA sequence. Direct repeats flanking the KpnI 1.8 kb DNA are boxed. The A-A-T-A-A-A polyadenylation signal sequence is indicated by a line above the sequence.

DNA and the corresponding regions of the previously sequenced DNAs of the KpnI families.

The inserted KpnI 1.8 kb DNA is flanked with an imperfect 15 bp duplication of the mtDNA-like sequence (Fig. 3). One of the direct repeats differs from the corresponding mtDNA sequence in having three base-substitutions and another, one-base deletions at two different positions. The KpnI 1.8 kb DNA has an A-rich sequence at the 3'-end, and contains an A-A-T-A-A-A polyadenylation signal sequence within the A-rich sequence (Fig. 3).

Fig. 4. Hybridization of #212 fragment with labelled pBK(1.8)11. A subcloned fragment #212 was digested with MluI, electrophoresed on a 0.8% agarose gel (lane a), transferred to a filter and the filter was hybridized with a ^{32}P -labelled pBK(1.8)11⁹ (lane b).

DISCUSSION

The DNA sequence duplication at the point of insertion is one of the distinctive features of the prokaryotic and eukaryotic transposable elements and our finding that this KpnI 1.8 kb DNA is immediately flanked with direct repeats indicates the mobility of this DNA family. However, the KpnI 1.8 kb DNA present within the mtDNA-like sequence shares no common structures with those of transposable elements, *i.e.*, it has neither long direct repeats nor short inverted repeats, at both termini²⁴ (Fig. 3). On the other hand, the KpnI 1.8 kb DNA has an A-rich sequence at one end, and contains an A-A-T-A-A polyadenylation signal sequence within the A-rich region (Fig. 3). Some of the structural characteristics of the KpnI 1.8 kb DNA region resemble that of the Alu family²⁵. The Alu family consist of short, dispersed repetitive sequences and exist in the nuclear DNAs in humans and other mammals. They also contain an A-rich sequence at one end and are flanked with short direct repeats of variable length (6-19 bp). The dispersion of both the Alu and the KpnI 1.8 kb families throughout the human genome and the structural similarities suggest that they possess common mechanisms for transposition. The model proposed for the transposi-

tion of Alu family explains the presence of an A-rich sequence at one end of the Alu family and requires that the repetitive DNA element be transcribed to RNA after which it is reverse-transcribed to cDNA before it integrates into nuclear DNA^{26,27}. Some KpnI families, including the KpnI 1.8 kb family, are reportedly transcribed in HeLa cells by RNA polymerase II²⁸. Actually, we found that about 0.08% of the cDNA clones in a human liver cDNA library carry sequences complementary to the KpnI 1.8 kb DNA (unpublished results). These results suggest that an RNA intermediate may be involved in transposition of the KpnI 1.8 kb family.

As shown in Fig. 2A, the 5'-end of the KpnI 1.8 kb DNA region is highly repeated in the human nuclear genome whereas the 3'-end region is not highly repeated. We found that the 1.3 kb length DNA of the 5'-end region of the KpnI 1.8 kb DNA has a considerable homology with one of the 1.3 kb KpnI cDNAs and that the homologous region terminates in an A-rich region (Fig. 3). Although the 470 bp length DNA of the 3'-end region of our KpnI 1.8 kb DNA shares no homology with the corresponding regions of the already sequenced KpnI families, this DNA region hybridized with one of the cloned KpnI 1.8 kb family (Fig. 4). From these results, we speculate that the 1.3 kb length DNA of the 5' portion of the KpnI 1.8 DNA is a prototype of a highly abundant KpnI family. This prototype 1.3 kb KpnI family transposed initially right next to the 5'-end of a DNA sequence which corresponds to the 470 bp length of the 3' portion of the KpnI 1.8 kb DNA. Subsequently, the prototype KpnI 1.3 kb and the 470 bp DNAs are transcribed as a single mRNA, reverse-transcribed into cDNA, and then inserted into the nuclear mtDNA-like sequence.

We propose that one of the following two events had occurred within the human cells; a) mtDNA fragments were taken up at various sites of the nuclear DNA, after which one member of the nuclear KpnI 1.8 kb family was translocated within one of the nuclear mtDNA sequences, or b) the KpnI 1.8 kb family moved into the mitochondria, transposed into mtDNA and subsequently various mtDNA fragments including a fragment containing the KpnI 1.8 kb DNA were integrated into the nuclear DNA. The divergence of the nuclear mtDNA sequence from that of mtDNA is 27% (Fig. 3). If

we assume that mutations in the nuclear genome are fixed in germ lines of mammals at the rate of 0.7% per million years²⁹, this mtDNA region would thus be transferred from the mitochondrion into the nuclear genome about 40 millions years ago. The observation that the sequence homology between the 5'-end region of the inserted KpnI 1.8 kb DNA and pcD-KpnI-4 is 92% (Fig. 3), indicates that the event a) is preferable.

ACKNOWLEDGEMENTS

We thank Dr. J. J. Maio for the gift of plasmids containing KpnI families and Mariko Ohara of Kyushu University for reading the manuscript. This work was supported by grants from the Ministry of Education, Science and Culture, Japan.

REFERENCES

1. Tsuzuki, T., Nomiya, H., Setoyama, C., Maeda, S. and Shimada, K. (1983) *Gene* 25, 223-229.
2. Farrelly, F. and Butow, R. A. (1983) *Nature* 301, 296-301.
3. Gellissen, G., Bradfield, J. Y., White, B. N. and Wyatt, G. R. (1983) *Nature* 301, 631-634.
4. Wright, R. M. and Cummings, D. J. (1983) *Nature* 302, 86-88.
5. Jacobs, H. T., Posakony, J. W., Grula, J. W., Roberts, J. W., Xin, J-H., Britten, R. J. and Davidson, E. H. (1983) *J. Mol. Biol.* 165, 609-632.
6. Kemble, R. J., Mans, R. J., Gabay-Laughnan, S. and Laughnan, J. R. (1983) *Nature* 304, 744-747.
7. Hadler, H. I., Dimitrijevic, B. and Mahalingam, R. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 6495-6499.
8. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. and Young, I. G. (1981) *Nature* 290, 457-465.
9. Shafit-Zagardo, B., Brown, F. L., Maio, J. J. and Adams, J. W. (1982) *Gene* 20, 397-407.
10. Blin, N. and Stafford, D. W. (1976) *Nucleic Acids Res.* 3, 2303-2308.
11. Drouin, J. (1980) *J. Mol. Biol.* 140, 15-34.
12. Smith, G. E., and Summers, M. D. (1980) *Anal. Biochem.* 109, 123-129.
13. Rigby, P. W. J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
14. Maxam, A. M. and Gilbert, W. (1980) *Meth. Enzym.* 65, 499-560.
15. Kuhara, S., Matsuo, F., Futamura, S., Fujita, A., Shinohara, T., Takagi, T. and Sakaki, Y. (1984) *Nucleic Acids Res.* 12, 89-99.
16. Tsuzuki, T., Nomiya, H., Setoyama, C., Maeda, S., Shimada, K. and Pestka, S. (1983) *Biochem. Biophys. Res. Commun.* 114, 670-676.

17. Shafit-Zagardo, B., Maio, J. J. and Brown, F. L. (1982) *Nucleic Acids Res.* 10, 3175-3193.
18. Sakaki, Y., Kurata, Y., Miyake, T. and Saigo, K. (1983) *Gene* 24, 179-190.
19. Miyake, T., Migita, K. and Sakaki, Y. (1983) *Nucleic Acids Res.* 11, 6837-6846.
20. DiGiovanni, L., Haynes, S. R., Misra, R. and Jelinek, W. R. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 6533-6537.
21. Schmeckpeper, B. J., Scott, A. F. and Smith, K. D. (1984) *J. Biol. Chem.* 259, 1218-1225.
22. Potter, S. S. (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81, 1012-1016.
23. Lerman, M. I., Thayer, R. E. and Singer, M. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 3966-3970.
24. "Mobile Genetic Elements", Shapiro, J. A., ed., Academic Press, New York, 1983
25. Jelinek, W. R. and Schmid, C. W. (1982) *A. Rev. Biochem.* 51, 813-844.
26. Van Arsdel, S. W., Denison, R. A., Bernstein, L. B., Weiner, A. M., Manser, T. and Gesteland, R. F. (1981) *Cell* 26, 11-17.
27. Jagadeeswaran, P., Forget, B. G. and Weissman, S. M. (1981) *Cell* 26, 141-142.
28. Shafit-Zagardo, B., Brown, F. L., Zavodney, P. J. and Maio, J. J. (1983) *Nature* 304, 277-280.
29. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolondner, R. and Dodgson, J. (1980) *Cell* 20, 555-566.