Volume 11 Number 19 1983 Nucleic Acids Research

Some KpnI family members are associated with the Alu family in the human genome

Toshio Miyake*, Kaoru Migita and Yoshiyuki Sakaki+

Research Laboratory for Genetic Information, Kyushu University School of Medicine, Fukuoka 812, Japan

Received 16 August 1983; Accepted 9 September 1983

ABSTRACT

The structures of the termini and their flanking regions of two human KpnI family members were investigated. The two differed in length, but the starting sequence at one terminal (defined as the 5' terminal) was found to be common to both members. The Alu family sequence was found in the 5' flanking regions. The KpnI family sequence started several base-pairs downstream from the 3' end of the Alu family sequence. In both cases, the Alu family sequence was not flanked by the direct repeat sequence common to the Alu family. These two members showed no sequence homology in 3' terminal regions. Interestingly, the Alu family plus the KpnI family unit was found to be flanked by a direct repeat sequence of several base-pair length. Based on these findings, relationship between the Alu family and KpnI family is discussed.

INTRODUCTION

The mammalian genome contains a number of interspersed repetitive sequences(see ref. 1 for review). The Alu family is repeated about 3×10^5 times(2-4) in the human genome and has a length of about 300 base pairs, consisting of two direct repeats of a monomeric unit sequence and a characteristic A-rich sequence at the 3' end of each monomeric unit(2, 5-7). The sequence (the dimeric form) is flanked by direct repeats of several base pairs length(6-11). Since transposable DNA elements are generally flanked by direct repeat sequences, the Alu sequence is thought to be a kind of mobile DNA element(1, 7, 10-12). Since the Alu sequence is transcribed by polymerase III(at least in vitro) (7, 8, 12-14) and transcripts containing the Alu sequence have been found in cytoplasmic RNA and hnRNA(4, 12-15), a model for transposition of the Alu sequence has been proposed, in which cDNA of the Alu sequence transcript is an intermediate in transposition(12).

In the last three years, another class of interspersed repetitive sequences, designated "LINE" by Singer(1), has been identified in the mammalian genome. LINE has a length of 5-6kb and is repeated about 10^4 times. In the primate genome, a repetitive sequence family belonging to this class has been identified (16-24) and designated the "KpnI family"(or "HindIII family") (17, 18). Since the KpnI family is interspersed throughout the genome and is repeated so many times(10^4)(16, 17, 19-22), it has been thought that the KpnI family is(or was) a mobile DNA element(21). This was strengthen by the findings that a short KpnI family member in a monkey genome is flanked by direct repeat sequence(25). However, the mechanism of amplification and dispersion of the KpnI family is not yet well understood. Since transposable DNA elements generally have characteristic structures in their terminal regions, we investigated the nucleotide sequences of the terminal regions of two human KpnI family members. Our results show that the KpnI family is closely associated with the Alu family(at least in these two cases) and can be transposed together with the the Alu family.

MATERIALS AND METHODS

The human genomic DNA clones HH10 and TβG41 have been described by Sakaki et al.(22) and Adams et al.(16), respectively. The 2.2- and 2.3-kb fragments of a HindIII digest of HH10 were inserted into the HindIII site of pBR322 and designated as pH12 and pH21, respectively. Subclones of TβG41(pRK12 and pRK20) have been described previously(16).

Blotting experiments were carried out essentially as described by Southern(26). DNA sequencing was done by chemical degradation methods(27). Sequence homology was analyzed by the program developed by Staden(28).

Restriction enzymes and other enzymes were purchased from Bethesda Research Laboratories and Takara Shuzo Co., Ltd.(Kyoto, Japan). α-^{32}PATP and α-^{32}PdCTP were obtained from New England Nuclear and Amersham.

RESULTS

Structures of two members of the KpnI family

We previously isolated a human genomic DNA clone which contains a repetitive sequence belonging to the KpnI family(22). Adams et al.(16) found a long repetitive sequence about 3kb downstream from the human β-globin gene(the 5' direction is defined as "upstream"), which has been shown to be a member of the KpnI family(29). We employed these two sequences(designated as HH10 and TβG41, respectively) to investigate the structure of terminal regions of the KpnI family.

Heteroduplex and hybridization studies showed that the KpnI family sequences in HH10 and TβG41 had lenghts of 3.5−4.0kb and 6.4kb, respectively(16, 22). The restriction maps of the two KpnI family members are presented in Fig. 1. Since regions B and C of HH10 hybridized with regions II and III of TβG41, respectively(data not shown), these two KpnI members were aligned

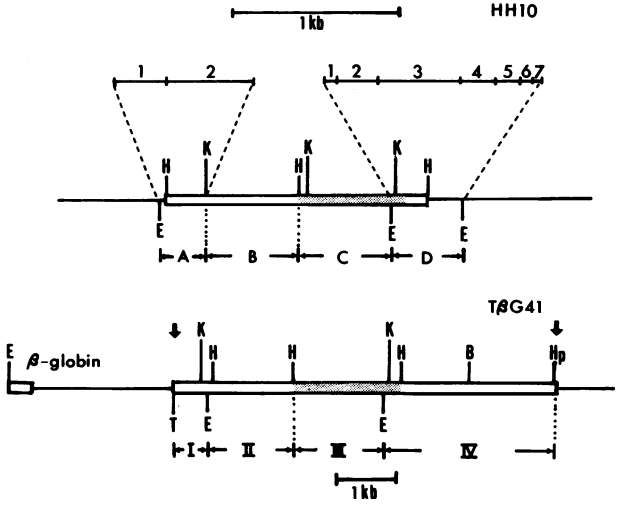

Fig. 1. Physical maps of the KpnI family in HH10 and TβG41. The restriction sites for EcoRI(E), BamHI(B), HindIII(H), KpnI(K), BglII(Bg) and HpaI (Hp) have been reported previously (16, 22). The HinfI sites in regions A and D(shown in the upper lane of HH10 map) and the TaqI site(T) in region I were determined in this work. Arrows shows terminal points of the repetitive sequence determined in a heteroduplex study(16). Regions estimated to contain the KpnI family are shown by open boxes. Regions containing the HindIII 1.9kb sequence(18, 19, 22) are shaded.

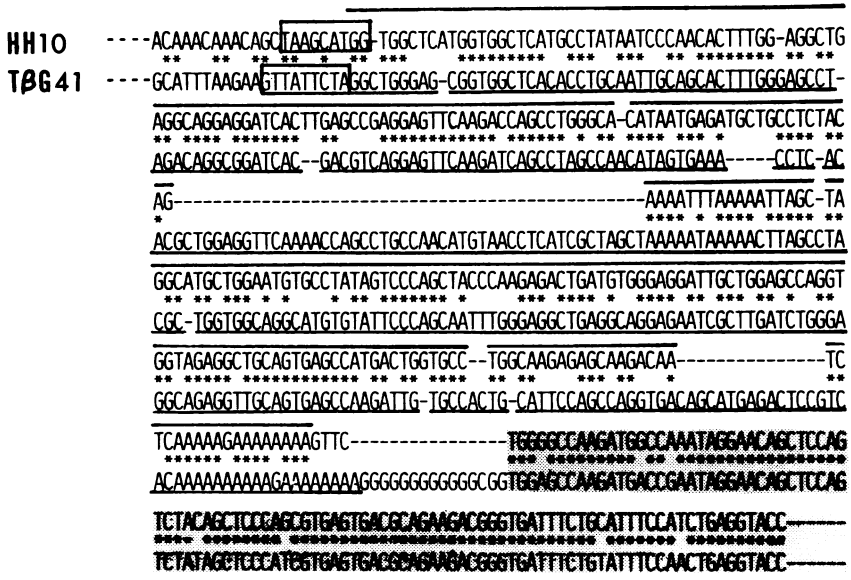

Fig. 2 Nucleotide sequences of the 5' terminal regions of the KpnI family in HH10 and TβG41. DNA fragments corresponding to region A and I were prepared from plasmid clones pH12(see MATERIALS AND METHODS) and pRK12(16). DNA sequencing was carried out as described by Maxam and Gilbert(26). Homologous sequences in HH10 and TβG41 are shown by asterisks. The region homologous to the Alu family sequence is underlined(or overlined). The 5' starting point of the Alu family was determined by comparing our data with those of others(5-11). Highly conserved sequences (presumably KpnI family sequences) are shaded. Sequences in boxes are those which form direct repeats with the 3' flanking sequences of the KpnI family(see Fig. 3).

as shown in Fig. 1. For convenience, we defined the left-side ends of the sequences in Fig. 1 as the 5' ends. For sequence analysis, the boundary between repetitive and non-repetitive sequences in HH10 was estimated with Southern blotting experiments: fragments A and D were isolated, digested with HinfI and subjected to polyacrylamide gel electrophoresis. Then, DNA was transferred to a nitrocellulose filter and hybridized with ^{32}P-labeled total human DNA as a probe. Regions which hybridized efficiently with total human DNA were defined as repetitive sequence regions. Results(not shown) showed that the boundaries(that is, the 5' and 3' termini of the repetitive sequence) were on the A2 and D2 fragments. The 5' and 3' termini

of the repetitive sequence in TβG41 were previously studied by Adams et al.(16) and found to be in region I and near the HpaI site(indicated by arrows in Fig. 1), respectively. These results were summarized in Fig. 1. It is obvious that the 5' terminal regions of the repetitive sequences are conserved in these two clones, but that the repetitive sequences differ very much in length on their 3' side.

Nucleotide sequences of terminal regions

We determined the nucleotide sequences of regions A and D of HH10, and region I and that including the HpaI site of TβG41, which were thought to contain the termini of the repetitive sequences.

Fig. 2 shows the sequences of regions A and I including the 5' termini of HH10 and TβG41, respectively. The sequence homology between these two sequences was analyzed by computer. Two homologous regions were found in these two clones : a highly homologous region(shaded in Fig. 2) and a moderately homologous region(underlined in Fig. 2). The highly homologous region seemed to be the 5' terminal region of the KpnI family because the sequence was so highly conserved in these two clones. The moderately homologous sequences included an A cluster at their 3' end and were found to have high homology to Alu family sequences(5). These regions hybridized well with a typical Alu family sequence, BLUR8(5). Therefore, we concluded that the moderately homologous sequences belonged to the Alu family. It should be noted that the Alu family sequences in these clones were not flanked by the direct repeat characteristic of the Alu family. The Alu family sequence and the KpnI family sequence are close to each other, but are interrupted by unique sequences of several bp length.

The sequences of region D of HH10 and the region near the HpaI site of TβG41 are presented in Fig. 3. These sequences were concluded to contain the 3' termini of the repetitive sequences. Since the two sequences showed no significant homology, it was hard to define the 3' termini of the repetitive sequences with accuracy. However, it should be pointed out that the sequence underlined in Fig. 3a showed high homology(>90%) with the HindIII 1.9kb repetitive sequence determined by Manuelidis (18). The 3'

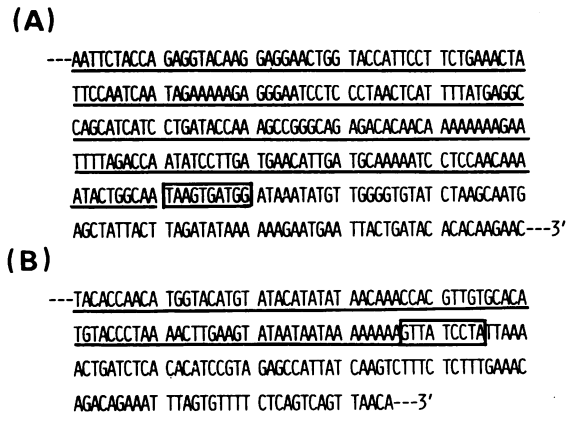

Fig. 3. Nucleotide sequences of the regions estimated to contain the 3' ends of KpnI family in HH10(A) and TβG41(B). DNA fragments containing region D of HH10 and the HpaI site of TβG41 were isolated from plasmid clones pH21(see text) and pRK20(16). DNA sequences were determined by the method of Maxam and Gilbert(27). The underlined sequence in (A) is homologous with that of HindIII 1.9kb family determined by Manuelidis(18). The underlined sequence in (B) is homologous with that of a short member of the KpnI family in a monkey genome(25). Sequences in boxes are those which form direct repeats with the 5' flanking sequence of the Alu family(see Fig. 2).

end of the underlined sequence is the 1865th nucleotide in the HindIII 1.9kb repetitive sequence which is in all 1894 bp long. Therefore, it is reasonable to consider that the repetitive sequence in HH10 has an incomplete structure as a KpnI family and that the 3' end of the underlined sequence in Fig. 3a is the actual 3' end of the KpnI family in HH10. It is interesting that the sequence(TAAGTGATGG) flanking the 3' end of the underlined sequence has high homology with the 5' flanking sequence(TAAGCATGG) of the Alu family shown in Fig. 2a. The sequence including the HpaI site in TβG41 had high sequence homology with the 3' end of the short member of the KpnI family in the monkey genome(25). The homologous region including an A cluster is underlined in Fig. 3b. A-rich sequences have been also found at the 3' ends of some putative mobile DNA elements, such as the Alu family(5) and a *Drosophila* insertion element(30). Therefore, it was concluded that the A cluster is the 3' end of the KpnI family in TβG41. It is of interest that the

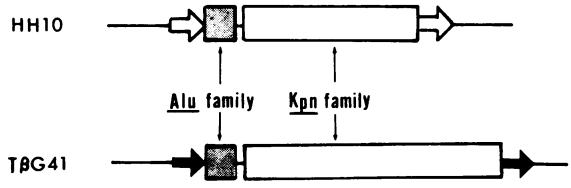

Fig. 4. Schematic presentation of structures of two KpnI family members. Direct repeats are shown by arrows.

sequence(GTTATCCTA) flanking the 3' end of the A cluster has high homology with the 5' flanking sequence(GTTATTCTA) of the Alu family shown in Fig. 2b as in the case of the KpnI family in HH10.

These results are summarized schematically in Fig. 4. It should be emphasized that the Alu family plus KpnI family unit is flanked by 8-10 bp direct repeat sequences just as in the case of mobile DNA elements.

DISCUSSION

The structures of two KpnI family members were investigated. Comparison of the overall structures of these two members showed that the 5' ends of both were fixed but that the positions of their 3' ends differed greatly. The length-heterogeneity at the 3' end appears common to the KpnI family, as shown in the heteroduplex study by Adams et al.[see Fig. 5 in Adams et al.(16)].

Sequence analysis showed that the Alu family is present close to the KpnI family in both cases. The characteristics of these two the Alu, KpnI-associated structures are: 1) the Alu sequences are present at quite similar positions and in the same direction relative to the KpnI family, 2) the Alu sequences are not flanked by a direct repeat sequence, which is common to mobile DNA elements, and 3) the Alu family plus KpnI family sequence is flanked by a direct repeat sequence as a unit. From these findings, it is hard to think that the Alu sequence is accidentally near the KpnI family. Rather, it is reasonable to think that the Alu family plus KpnI family structure behaves as a mobile unit because the Alu—KpnI structure is flanked by a

direct repeat in both cases. However, it must also be remembered that the sequence conservations of the Alu and KpnI regions in the Alu-KpnI units are quite different. This implies that association of the KpnI family with the Alu family took place recently during evolution and that at least more than two independent Alu - KpnI units were formed. All these findings could be interpreted as follows: the Alu family and the KpnI family were originally independent of each other and were amplified independently. But once the two family members came near each other in some way, they formed a new mobile unit. Since the Alu family has so many members, various Alu - KpnI mobile units must have been formed and have amplified during evolution. Two of these were investigated in this work. Our hypothesis is compatible with the findings by Shafit-Zagardo et al.(17) that a human KpnI 1.9kb segment belonging to the KpnI family contained an Alu-like sequence. The Alu family may also be associated with the Kpn family in the African green monkey genome(21), in which some KpnI family members were found near the Alu family, although their precise locations are unknown.

The possiblility that the Alu and KpnI family could form a new unit is strengthen by the findings that the KpnI sequence near the human β-globin was transcribed(at least in vitro) using the adjacent Alu sequence as a promoter(31). Assuming that the Alu-promoted transcripts are intermediates in transposition, this possibility provides a reasonable explanation not only for the formation of the Alu-KpnI mobile unit, but also for our observations that KpnI sequences were found only at the 3' end of the Alu sequence and that the length of KpnI family members varied very much only at their 3' side. Recently, transcripts containing the KpnI family sequence were found in vivo(32, 33). It is conceivable that some of these transcripts are intermediates for transposition of the KpnI family.

ACKNOWLEDGEMENTS

We thank Prof. Yasuyuki Takagi for encouragement throughout this work. We are also grateful to Dr. Arthur W. Nienhuis for kindly supplying the TβG41 phage and its related clones, and Dr. Prescott L. Deininger for kindly sending the BLUR8 clone. Miss Honoho Hamada prepared the manuscript with great care.

*Present address: Sagami Chemical Research Center, Sagamihara, Kanagawa, Japan

+To whom correspondence should be addressed

REFERENCES

1. Singer, M. F. (1982) Cell 28, 433-434
2. Houck, C. M., Rinehart, F. P., and Schmid, C. W. (1979) J. Mol. Biol. 132, 289-306
3. Rubin, C. M., Houck, C. M., Deininger, P. L., Friedman, T. and Schmid, C. W. (1980) Nature 284, 372-374
4. Tashima, M., Calabretta, B., Torelli, G., Scofield, M., Maizel, A. and Saunders, G. F. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 1508-1512
5. Deininger, P. L., Jolly, D. J., Rubin, C. M., Friedmann, T. and Schmid, C. W. (1981) J. Mol. Biol. 151, 17-33
6. Bell, G. I., Pictet, R. and Rutter, W. J. (1980) Nucl. Acids Res. 8, 4091-4109
7. Pan, J., Elder, J. T., Duncan, C. H. and Weissman, S. M. (1981) Nucl. Acids Res. 9, 1151-1170
8. Duncan, C. H., Jagadeeswaran, P., Wang, R. R. C. and Weissman, S. M. (1981) Gene 13, 185-196
9. Baralle, F. E., Shoulders, C. C., Goodbourn, S., Jeffreys, A. and Proudfoot, N. J.. (1980) Nucl. Acids Res. 8, 4393-4404
10. Grimaldi, G., Queen, C. and Singer, M. F. (1981) Nucl. Acids Res. 9, 5553-5567
11. Grimaldi, G. and Singer, M. F. (1982) Proc. Natl. Acad. Sci. U.S.A. 79,1497-1500
12. Jagadeeswaran, P., Forget, B. G. and Weissman, S. M. (1981) Cell 26,141-142
13. Elder, J. T., Pan, J., Duncan, C. H. and Weissman, S. M. (1981) Nucl. Acids Res. 9,1171-1189
14. Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C. H., Biro, P. A., Choudary, P. V., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L. and Schmid, C. W. (1980) Proc. Natl. Acad. Sci. U.S.A. 77,1398-1402
15. Ullu, E., Murphy, S. and Melli, M. (1982) Cell 29, 195-202
16. Adams, J. M., Kaufman, R. E., Kretschmer, P. J., Harrison, M. and Nienhuis, A. W. (1980) Nucl. Acids Res. 8, 6113-6128
17. Shafit-Zagardo, B., Maio, J. J. and Brown, F. L. (1982) Nucl. Acids. Res.10, 3175-3193
18. Manuelidis, L. (1982) Nucl. Acids Res.10, 3211-3219
19. Manuelidis, L. and Biro, P. A. (1982) Nucl. Acids Res.10, 3221-3239
20. Schmeckpeper, B. J., Willard, H. F. and Smith, K. D. (1981) Nucl. Acids Res. 9, 1853-1872
21. Grimaldi, G. and Singer, M. F. (1983) Nucl. Acids Res. 11, 321-338
22. Sakaki, Y., Kurata, Y., Miyake, T. and Saigo, K. (1983) Gene in press
23. Maio, J. J., Brown, F. L., McKenna, W. G. and Musich, P. R. (1981) Chromosoma 83, 127-144
24. Gillespie, D., Adams, J. W., Costanzi, C. and Caranfa M. J. (1982) Gene 20, 409-414
25. Thayer, R. and Singer, M. F. (1983) Mol. Cell. Biol. 3,

967-973
26. Southern, E. M. (1975) J. Mol. Biol. 98. 503-517
27. Maxam, A. and Gilbert, W. (1979) Methods Enzymol. 65, 504-561
28. Staden, R. (1977) Nucl. Acids Res. 4, 4037-4051
29. Shafit-Zagardo, B., Brown, F. L., Maio, J. J. and Adams, J. W. (1982) Gene 20, 397-407
30. Dawid, I. B., Long, E. O., DiNocera, P. P. and Pardue M. L. (1981) Cell 25, 399-408
31. Manley, J. L. and Colozzo, M. T. (1982) Nature 300, 376-379
32. Kole, L. B., Hayness, S. R. and Jelinek, W. R. (1983) J. Mol. Biol. 165, 257-286
33. Shafit-Zagardo, B., Brown, F. L., Zavodny, P. J. and Maio, J. J. (1983) Nature 304, 277-280