Volume 13 Number 21 1985 Nucleic Acids Research

Sequence analysis of a KpnI family member near the 3' end of human β-globin gene

Masahira Hattori, Saburo Hidaka[1] and Yoshiyuki Sakaki

Research Laboratory for Genetic Information, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812, and [1]Department of Oral Biochemistry, Fukuoka Dental College, 700 Oaza Ta Sawara-ku, Fukuoka 812, Japan

Received 19 September 1985; Accepted 12 October 1985

ABSTRACT

We determined the complete nucleotide sequence (6125 bp) of a full-length member of human KpnI family, designated TβG41, which is located about 3 kb downstream from the β-globin gene. Comparison of the sequence with the KpnI family sequence compiled by Singer revealed that a new 131 bp sequence is present in the TβG41. Hybridization analyses showed that a few thousand of human KpnI family members are carrying this additional sequence. Computer search of DNA databases for TβG41-homologous sequence showed that some TβG41-homologous sequences were closely associated with pseudogenes. The TβG41 sequence also showed significant sequence homology with ChBlym-1, a transferrin-like transforming gene of chicken. Furthermore, an amino acid sequence deduced from the TβG41 nucleotide sequence revealed a relatively-high homology to those of human transferrin and lactotransferrin.

INTRODUCTION

The KpnI family is a long interspersed repetitive sequence family of primate genomes. The sequence is about 6 kb in full-length and repeated about 10^4 times per haploid genome (see ref. 1 and 2 for reviews). Repetitive sequences evolutionally related to the KpnI family have been identified in other mammalians (see ref. 2). These sequences including the KpnI family were designated LINEs or L1 families (1, 3). Many members of KpnI family have been isolated and their structural characteristics were analyzed (4-20). In most cases, the 3' portions of each members were conserved and a long consecutive A-rich sequence was found at their 3' ends, but truncations were found at different positions of their 5' portions, which caused a length-variation of the family members. In some cases, internal rearrangements were found including deletions, insertions and inversions (8, 10, 13). Some open reading

frames (ORFs) were identified in L1 family sequences (6, 14, 21) and the possibility was pointed out that the L1 family was derived from a sequence(s) encoding a protein(s) (21). Some LINEs appeared to move recently in evolutionary time (see ref. 2) and at least in one case a transposition of the LINE affected the cellular regulation (22). Thus, the data have been accumulated, but it is still obscure how the sequences dispersed on the genome during evolution and whether the LINEs have some function. It is important to know the overall structure of the LINEs at sequence level to get more information on the "history" of the LINEs and, if any, the function of them. In KpnI family, the sequence data are available only from truncated members or parts of "full-length" members. These data have been compiled by Maxine Singer (personal communication), but it should be finally confirmed by the sequence of a member having the "full-length". Although there is no definite criterion for the "full-length" of KpnI family member, a member (TβG41) isolated by Adams et al. (4, 5) appears to be a "full-length" one : 1) the TβG41 has a A-rich sequence at the 3' end just as the 3' ends of many other members, and also has the 5' terminal sequence which is common to the 5' ends of some large-sized members (12, 13), 2) it showed a restriction map similar to a consensus map of KpnI family which was derived from Southern blotting analysis of total human DNA (19, 20), and 3) it has a length of ~ 6 kb which is a reasonable size as a "full-length" member (5, 13).

In this paper we present the complete nucleotide sequence and structural characteristics of the TβG41. Implication of the results will be discussed in terms of evolution of LINEs.

MATERIALS AND METHODS

Materials. The phage clone TβG41 has been described by Adams et al. (4, 5). The plasmid pUC series have been described by Messing and his colleagues (23, 24). Sequencing primer I (5'-CAGGAAACAGCTATGAC-3'), deoxyribonucleoside triphosphates (dNTPs) and dideoxyribonucleoside triphosphates (ddNTPs) were purchased from Takara Shuzo (Kyoto, Japan), Pharmacia P-L Biochemicals and Amersham. Sequencing primer II (5'-CCAGTCACGACGTTGTA-3') was

prepared by ourselves using Applied Biological System DNA synthesizer Model 380A. Restriction enzymes and other enzymes were obtained from Takara Shuzo and Nippon gene (Toyama, Japan). [α-^{32}P]dCTP (specific radioactivity: 400Ci/mmol, 10mCi/ml) was purchased from Amersham.

Plasmid constructions. Appropriate restriction fragments of the TβG41 DNA (Fig. 1) were subcloned into pUC 13 or pUC18. When the inserted DNA was too long to be sequenced at a time, the inserted DNA was shortened from both ends by BAL 31 exonuclease and then ligated with SmaI-digested pUC13. The ligated DNAs were introduced into Escherichia coli JM83 (23). Transformants were selected on agar plate containing ampicillin and X-gal.

DNA sequencing. We developed a rapid and simple dideoxy sequencing method in which denatured plasmid DNA is used as a template. Since this method appears very useful for people in this field, the details are described below.

Plasmid DNA was extracted from 1.5ml of overnight culture of E. coli JM83 harboring a recombinant plasmid by alkaline lysis method (25). After treatment with ribonuclease A (final concentration; 10mg/ml) at 37^{0}C for 30 min, the DNA solution (50μl) was mixed well with 30 μl of 20% polyethylene glycol 6000 - 2.5M NaCl solution, and kept on ice for 1 hr. The precipitates were collected by centrifugation at 12,000 rpm for 5 min and rinsed once with 70% ethanol. The pellet was dried and dissolved in 50 μl of TE buffer (10mM Tris-HCl, pH 8.0 and 1mM EDTA). The purity of plasmid DNA is critical for the subsequent steps. Contamination of RNA and open circular plasmid should be minimized.

The above DNA solution (18 μl) was mixed with 2 μl of 2N NaOH and kept at room temperature for 5 min. Then, 8 μl of filter-sterilized 5M ammonium acetate (pH 7.4) was added and denatured DNA was precipitated by addition of 100 μl of ethanol at -70^{0}C for 5 min. The precipitates were harvested by centrifugation at 12,000 rpm for 5 min, rinsed once with 70% ethanol and dried under vacuum. This denatured DNA pellet can be kept in this form for a few weeks and dissolved in water (10 μl) just before use in the subsequent step.

The sequencing reaction was performed essentially by the

procedure commonly used for M13 phage vectors (26). The mixture of denatured plasmid (usually 0.75~1.0pmol or more, at least 0.5pmol in 5 μl), primer I or II (0.5pmol in 1 μl), 1.5 μl of 10 X Klenow buffer (70mM Tris-HCl, pH 7.5, 200mM NaCl, 70mM $MgCl_2$ and 1mM EDTA) and water (4.5 μl) was heated at 60°C for 15 min in a 1.5-ml microfuge tube. The sample (12 μl) was then kept at room temperature for 15 min. [α-^{32}P]dCTP (2 μl, 20 μCi) and Klenow fragment of polymerase I (1 μl, 2 units) were added and mixed well. The sample was immediately divided into four parts and each part was mixed with 2 μl of G-, A-, T- and C- specific dideoxy-deoxynucleotide mixture. The reactions were carried out at 37°C or more for 15~20 min. Reaction at 37°C or higher temperature is important to avoid the formation of extra bands. Then, chase solution (1 μl of 0.5~1mM dNTPs solution) was mixed and the samples were further kept at the same temperature for 15~20 min. Six microliter of formamide-dye solution (95% formamide, 0.1% bromophenol blue, 0.1% xylene cyanol) was added and the samples were kept on ice. Aliquots of four reaction mixtures (2 μl) were loaded on the sequencing gel immediately following heating at 95 °C for 3 min.

We employed 6% acrylamide-7M urea wedge gel (27) and 5% acrylamide-7M urea stretch gel for electrophoresis. The electrophoreses were done at 2000V in contact with an aluminium plate. After electrophoresis, the gel on one glass plate was immersed in 10% methanol-10% acetic acid solution for 15 min., transferred to a paper (Whatmann 3MM) and dried on the paper at 80°C under vacuum. The autoradiography was done for 12~20 hr at room temperature without intensifying screen. We can usually obtain sequence data more than 500 nucleotide at a time.

Computer analysis of DNA sequence Nucleotide sequence was analyzed by the GENAS system at Kyushu University Computer Center (28) which enables us to retrieve any sequence data from DNA and protein databases and readily to analyze them by various application programs. Search for KpnI family-homologous sequences was carried out by the program of Wilbur and Lipman (29, 30) in the GENAS.

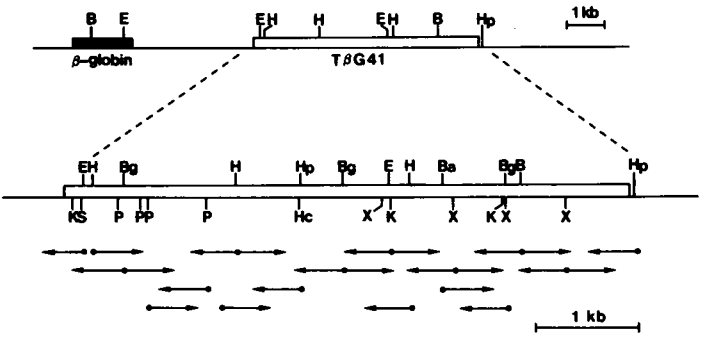


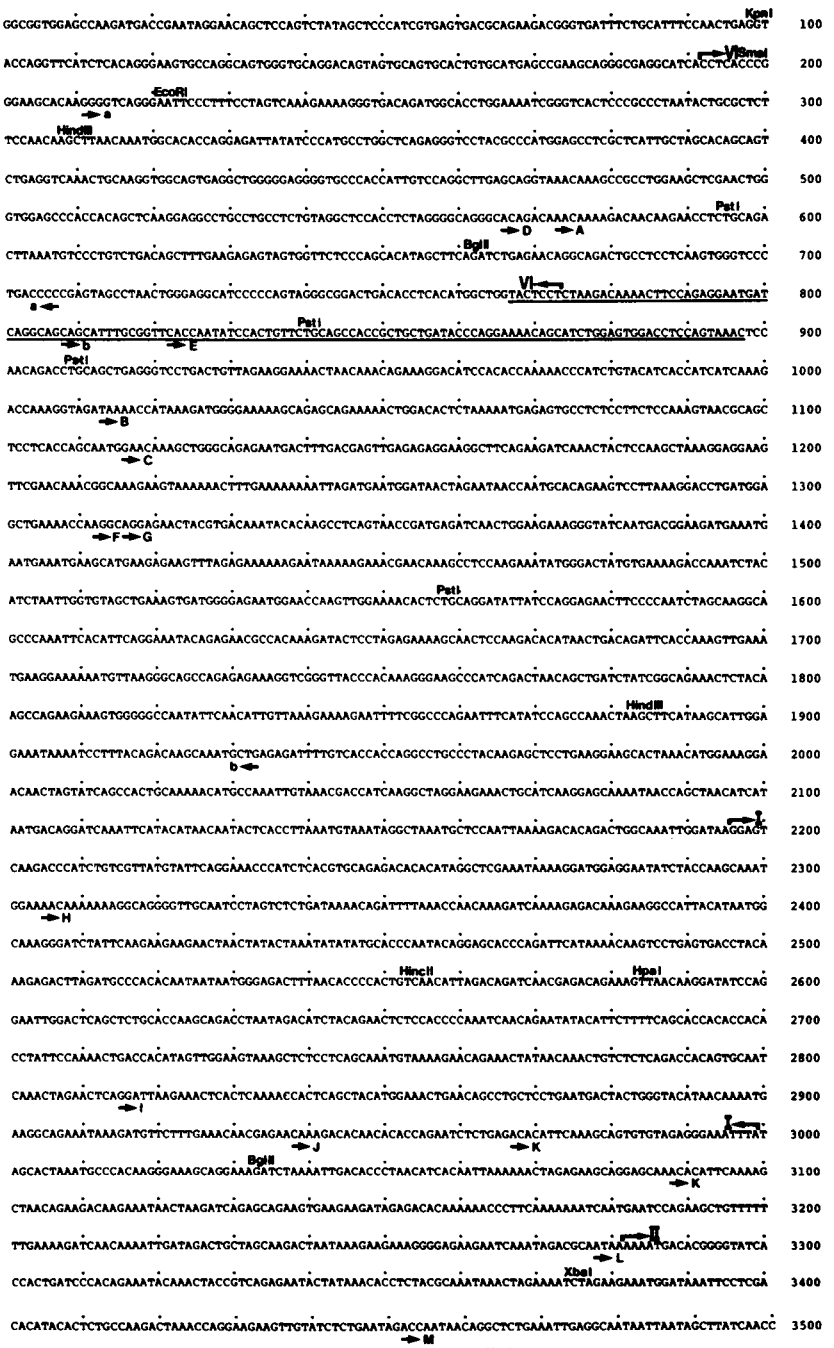
Fig. 1. Restriction map and sequence strategy of the TβG41 located downstream from the human β-globin gene. TβG41 is indicated by an open bar. The β-globin gene is indicated by a heavy bar. The sequencing strategy is indicated by the arrows below the restriction map. The sequencing was done by the modified dideoxy method described in Materials and Methods. Restriction sites are : EcoRI (E), HindIII (H), BamHI (B), KpnI (K), PstI (P), HpaI (Hp), HincII (Hc), BglII (Bg), XbaI (X), BalI (Ba), SmaI (S).

RESULTS

Nucleotide sequence of the TβG41

A KpnI family member (designated TβG41) has been found about 3 kb downstream from human β-globin gene (4, 5). Hybridization and partial sequence analysis suggested that it is a "full-length" member of the family (5, 12, 13). Clones carrying the restriction fragments of the TβG41 DNA were isolated and subjected to sequencing. To determine the long DNA sequence rapidly, we developed a modified dideoxy sequencing method using a denatured plasmid DNA as a template. The details of the method are described in Materials and Methods. This method allowed us to determine about 1 kb sequence from a single plasmid template. Using the method, the sequence of the TβG41 was determined according to the strategy shown in Fig. 1, and the results are summarized in Fig. 2.

The sequence was compared with the KpnI family sequence compiled by Singer from published and unpublished data. The compiled sequence is 5994 bp long, whereas the TβG41 is 6125 bp long (A-rich sequence at the 3' end was not included). The difference of the length is mainly due to a newly identified 131 bp sequence (nucleotides 767 to 897) in the TβG41 (Fig. 2).



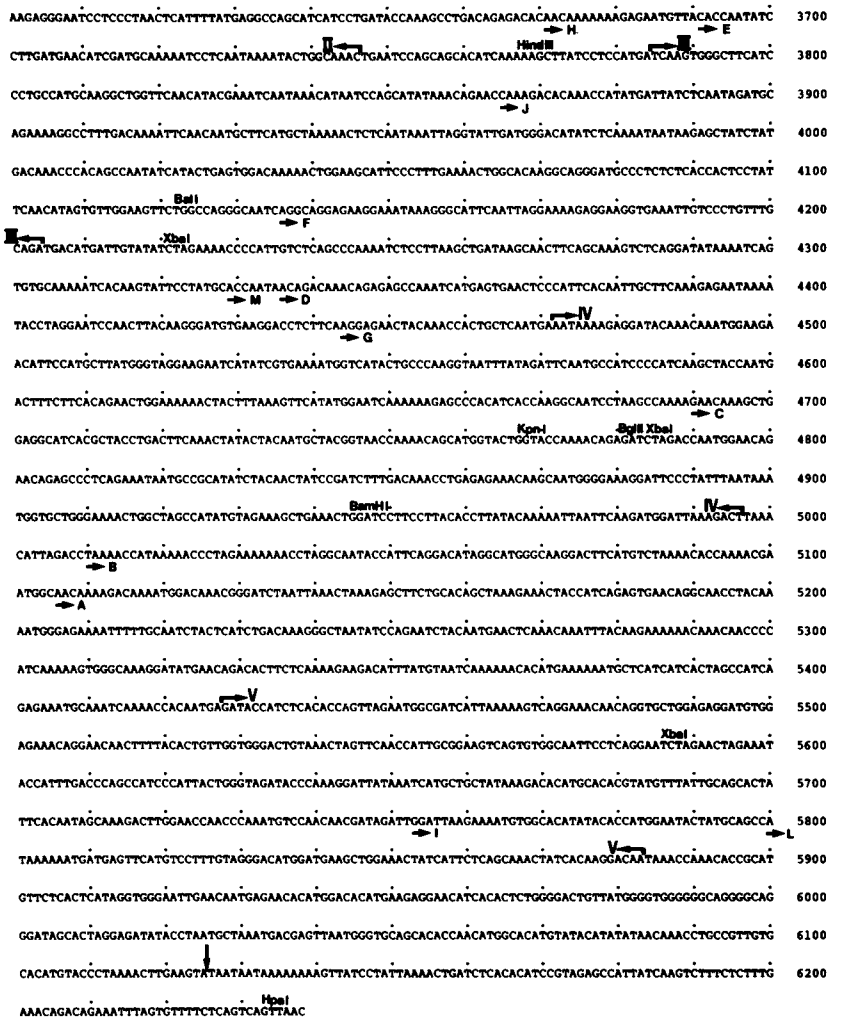


Fig. 2. Nucleotide sequence of TβG41. The starting and ending positions of the sequence have been described previously (12, 13). The ending position is indicated by a vertical arrow. The 131 bp additional sequence is underlined. Six possible ORFs are shown by arrows above the sequence. Direct (A-M) and inverted repeats (a and b) are indicated by short arrows under the sequence. Restriction sites are indicated above the sequence.

This sequence has not been found in the KpnI family members sequenced to date. However, a human KpnI family clone, Hspc6, which has been isolated from extrachromosomal circular DNA in the HeLa cell (31) had a similar sequence at the same position

(Yoshioka et al., unpublished). We tested whether the 131 bp sequence is a part of the KpnI family by dot and plaque hybridizations. About 10^5 plaques of human genomic DNA library were screened with two specific probes: a specific fragment containing the 131 bp sequence and the EcoRI-BglII fragment (nucleotides 220 to 661) of the TβG41 which is located 106 bp upstream from the 131 bp sequence. Most of positive plaques by the 131 bp sequence-specific probe hybridized with the EcoRI-BglII fragment. Approximately 30% of plaques hybridized to the EcoRI-BglII fragment did not hybridize with the 131 bp sequence-specific probe. Dot hybridization test indicated that the 131 bp sequence is repeated approximately 2000 times in the human haploid genome. On the other hand, the sequence of the EcoRI-BglII fragment was estimated to be repeated approximately 4000 times. The latter value is in good agreement with that obtained previously in the monkey DNA (13). These data indicated that the 131 bp sequence is a part of the KpnI family and present in 50～70 % of the KpnI family members containing 5' portions. In addition, the 131 bp sequence was found to hybridized with the monkey DNAs with similar frequency, suggesting that the 131 bp sequence is also present in monkey KpnI family. Another sequence difference between TβG41 and the compiled data was found at the nucleotides 1637 to 1653, where the 17 bp sequence had opposite directions each other in these two sequences. In another human clone (Hspc 6), the 17 bp sequence had the same orientation with that of the TβG41.

Characteristics of the TβG41 sequence

Total GC content in the TβG41 sequence is about 42%. GC contents of each 500 nucleotides from the 5' to the 3' ends were calculated to be 56, 59, 40, 43, 38, 40, 36, 38, 41, 39, 38 and 42% respectively. Thus, the first 1000 bp region showed significantly higher GC contents than other regions. This may mean that the 5' portion of the KpnI family sequence is originally different from other parts of the sequence.

Analysis by a computer showed that the TβG41 contains 177 pairs of perfect direct repeats and 28 pairs of perfect inverted repeats having the length of more than 9 bp. Among them, 13 pairs of direct repeats (A-M) having 11 bp or more in length and

two pairs of inverted repeats (a and b) having 10 bp or more are shown in Fig. 2. These inverted or direct repeats were found at random, suggesting that the TβG41 sequence had not been organized by duplication of large segment or repetition of small repeating units. High frequency of direct or inverted repeats may facilitate the intramolecular recombination which causes rearrangement of the sequence. There was no significant repeated structure at both terminal regions, suggesting that the TβG41 sequence shares no common structural features with typical transposable elements.

Six possible ORFs (I-VI) were found in both strands of the TβG41, and these are indicated by arrows in Fig. 2. The longest ORF-I consists of 268 amino acid residues. The ORF-I~V are mostly overlapped with ORFs found in the compiled sequence. This is consistent with the idea that the KpnI family was derived from a sequence(s) which encodes protein(s) (21).

It is of interest to ask whether KpnI family sequence contains some signals for biological activity. We searched for TATA box ($TATA\frac{T}{A}A$), polyadenylation signal (AATAAA-----CA), enhancer core sequence ($GTGG\frac{AAA}{TTT}G$) (32), nuclear factor I (NF-1) binding site ($TGGCN_5GCCA$) (33), possible methylation sites (CG) and possible Z-DNA sequence ((PyPu)n). Twenty two TATA sequences were identified and among them, those at positions 2441, 4290, 5647 and 6078 showed perfect match with the consensus sequence $TATA\frac{T}{A}A$. Sixteen AATAAA sequence were found and, among them, those at position 1903, 2909, 3277, 3950, 4394, 4472, 5686 and 5882 had CA sequence at appropriate distance (34). Sequences similar to that of enhancer core were identified at 44 positions and, among them, only one sequence CAAACCAC at position 4454-4461 showed perfect homology with the consensus sequence, although the orientation of the sequence was opposite. No possible NF-1 binding site sequence was found. Number of CG sequence was 60. The low content of CG suggested that the KpnI family will not be highly methylated. Clusters of $(\text{pyrimidine-purine})_n$ sequence were found at positions 3400-3409 (10 bp), 5670-5685 (16 bp), 5769-5780 (12 bp), 6066-6083 (18 bp) and 6097-6108 (12 bp). Although we have no evidence that these signals are actually active, combination of them with other

sequences through DNA rearrangement might form biologically-active signals.

Computer search for TβG41-homologous sequences in various genomes.

We searched the GenBank DNA sequence database by computer for sequences homologous to the TβG41 and found a variety of TβG41-homologous sequences. Most of them were sequences which have been already identified as KpnI or L1 family. But some of them were newly identified in this study. Those are summarized in Table 1. Interestingly, four KpnI (L1) family members were found to be closely associated with pseudogenes such as those of snRNA (35, 44) and immunoglobulin (36). The relation between these pseudogenes and L1 family sequences is schematically shown in Fig. 3. Recently, association of L1 family sequence with some pseudogenes or processed pseudogenes has been described in mouse interferon pseudogene (37), mouse γ-actin processed gene (38) and rat cytochrome c processed gene (39). These results are consistent with the idea of Scarpulla who has pointed out that L1 family and processed gene appear to have a tendency to associate with one another (39). L1 family sequence found in the first intron of mouse kallikrein gene (mGK-1) (40) had unusually rearranged structure in which truncated R element is associated with an inverted truncated Bam5 sequence. The composite sequence are flanked by 10 bp direct repeats. This structure resembles to the KpnI-RET in the monkey satellite DNA (9). Interestingly, a certain homology is found in the chicken Blym-1 (ChBlym-1), a transforming gene which encodes a protein partly homologous to transferrins (41). The sequence homology is shown in Fig. 4a. Considering about 65% sequence homology between primate and mouse L1 family (42), sequence homology of 60% over 300 bp between the TβG41 and the ChBlym-1 may have striking significance. The flanking region of ChBlym-1 is known to be highly repetitive (41). This repetitive sequence may belong to chicken L1 family. The protein sequence database by National Biomedical Research Foundation was also searched for amino acid sequences homologous to those deduced from the T G41 nucleotide sequence. A relatively-high homology was found between TβG41 and transferrin (47, 48) and

Table 1. Genes or cloned DNA containing sequences homologous to the TβG41

Gene or clone	Position	Corresponding region in the TβG41*	Homology	ref.
Human proenkephalin	3'flanking	-6125 — -6062	89%	43
Human U3 snRNA pseudogene (U3.2)	3'flanking	3109 — 3154	89%	35
Human U3 snRNA pseudogene (U3.6)	Both flanking	5337 — 5360 5351 — 5398	79% 88%	35
Human immunoglobulin processed gene (E3)	3'flanking	5026 — 5636	88%	36
Mouse kallikrein (mGK-1)	1st intron	-5565 — -5364	78%	40
Mouse U1 snRNA pseudogene ($U1_1$)	5'flanking	-3964 — -3908	77%	44
Mouse R3 repeat	Both flankings	2853 — 3094 3096 — 3964	63% 68%	45
Rat prolactin	4th intron	-5999 — -5155	70%	46
Chicken Blym-1	-	1637 — 2001	60%	41

*Minus numbers show the positions in the complementary strand of the TβG41 sequence.

lactotransferrin (49)(Fig. 4b). Similar amino acid sequence homology was also found between the ORF of mouse L1Md4 (21) and transferrins (Fig. 4b).

DISCUSSION

A complete nucleotide sequence of a "full-length" member of KpnI family was determined. The data revealed that the sequence contains a newly-identified 131 bp sequence which is a part of the KpnI family. Hybridization analyses showed that the primate KpnI family consists of at least two subfamilies, which can be distinguished by the presence or absence of this 131 bp sequence. Analysis of the 131 bp sequence and its flanking region showed that the sequence ACTCC was repeated at position 768 to 772 and at position 896 to 900. A recombination between these homologous short sequences may have generated a subfamily lacking the 131 bp sequence. Alternatively, the additional sequence may have integrated into a progenitor of KpnI family to form a new subfamily carrying the 131bp sequence. A number of direct and inverted repeats was found in the TβG41 sequence.

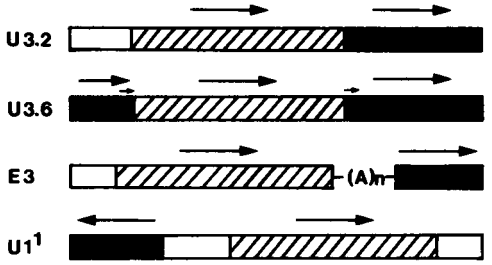


Fig. 3. Schematic illustration of the association of L1 family sequences with pseudogenes. Hatched bar show the pseudogenes of human U3 snRNA (U3.2 and U3.6 in ref. 35), mouse U1 snRNA ($U1^1$ in ref. 44) and human immunoglobulin (E3 in ref. 36). The L1 family sequences are indicated by heavy bars. The arrows indicate directions of the L1 sequence and the pseudogene. Short arrows indicate the direct repeats flanked by the pseudogene.

These internal repeats might facilitate internal rearrangements which were often found in the KpnI family sequences. In addition to internal rearrangements, our computer analysis showed that KpnI (L1) family sequence was associated with pseudogenes of snRNAs and processed pseudogene of immunoglobulin. L1 family sequences have been already known to be associated with mouse γ-actin processed pseudogene (38), rat cytochrome c processed pseudogene (39) and mouse interferon pseudogene (37). Association of L1 family sequence with pseudogenes appeared not to be the results of random events. As pointed out by Scarpulla (39), L1 family and processed pseudogene may have a tendency to associate with one another in general through unknown mechanism.

We found a significant sequence homology in the chicken DNA clone ChBlym-1 which has been characterized as a transforming gene encoding a protein partly homologous to transferrins (41). The sequence had a homology of 60% over 300 bp (Fig. 4a). Interestingly, the region upstream from the EcoRI site in the ChBlym-1 sequence was shown to be highly repetitive in chicken and mouse genomes (41). This repetitive region covers the most part of sequences homologous to the TβG41 sequence, suggesting that the repetitive sequence in the ChBlym-1 is a chicken LINE. Recently an example was shown that a LINE may affect the regulation of c-myc oncogene in dog (22). So, it

Fig.4(a) Sequence homology between the TβG41 and the ChBlym-1. The numbers below the ChBlym-1 show the same positions as those in ref. 41. Identical nucleotides are indicated by colons. Hyphens indicated missing bases. (b) Alignment of amino acid sequences from TβG41 and L1Md4 with those of human lactotransferrin (LTf) and transferrin (Tf). The numbers show the positions of the amino acid residues in the references 47 and 49. Amino acid sequences of TβG41 and L1Md4 were deduced from nucleotides 5265-5396 in the TβG41 and nucleotides 726-860 in the L1Md4 (21), respectively. Asterisks above the amino acid sequences indicate identical amino acid between transferrins and sequences from either TβG41 or L1Md4.

may be of interest to test whether this chicken LINE sequence is also involved in the transforming activity of the ChBlym-1. We also detected significant amino acid sequence homology between human transferrins and amino acid sequences in TβG41 and mouse L1Md4 (Fig. 4b). The ORF in L1Md4 shown in Fig. 4b is the region that are highly conserved among Mus species (21). Transferrins are members of a protein family including melanoma antigen p97 (50), ChBlym-1 and probably HuBlym-1 (51). One might consider that the LINEs or a part of LINE sequence are closely related to transferrin gene family.

ACKNOWLEDGEMENTS

We thank Dr. A. W. Nienhuis for supplying the phage clone TβG41 and its related clones, Drs. M. F. Singer, G. Grimaldi, R. E. Thayer, J. Skowronski and S. Contente for kindly providing the compiled sequence of the KpnI family including unpublished data. Drs. K. Yoshioka and H. Yamagishi also kindly provided their unpublished data of the clone Hspc6. We are grateful to Dr. S. Kuhara for his advice on the computer search and to Miss H. Hamada for assistance in preparation of the manuscript.

REFERENCES

1. Singer, M. F. (1982) Cell, 28, 433-434.
2. Singer, M. F. and Skowronski, J. (1985) TIBS 10, 119-122.
3. Voliva, C. F., Jahn, C. L., Comer, M. B., Hutchison, C. A. III and Edgell, M. H. (1983) Nucl. Acids Res. 11, 8847-8859.
4. Adams, J. W., Kaufman, R. E., Kretschmer, P. J., Harrison, M. and Nienhuis, A. W. (1980) Nucl. Acids Res. 8, 6113-6128.
5. Shafit-Zagardo, B., Brown, F. L., Maio, J. J. and Adams, J. M. (1982) Gene 20, 397-407.
6. Manuelidis, L. (1982) Nucl. Acids Res. 10, 3211-3219.
7. Ullrich, A., Gray, A., Goeddel, D. V. and Dull, T. J. (1982) J. Mol. Biol. 156, 467-486.
8. Grimaldi, G. and Singer, M. F. (1983) Nucl. Acids Res. 11, 321-338.
9. Thayer, R. E. and Singer, M. F. (1983) Moll. Cell. Biol. 3, 967-973.
10. Lerman, M. I., Thayer, R. E. and Singer, M. F. (1983) Proc. Natl. Acad. Sci. USA 80, 3966-3970.
11. DiGiovanni, L., Haynes, S. R., Misra, R. and Jelinek, W. R. (1983) Proc. Natl. Acad. Sci. USA 80, 6533-6537.
12. Miyake, T., Migita, K. and Sakaki, Y. (1983) Nucl. Acids Res. 11, 6837-6846.
13. Grimaldi, G., Skowronski, J. and Singer, M. F. (1984) EMBO. J. 3, 1753-1759.
14. Potter, S. S. (1984) Proc. Natl. Acad. Sci. USA 81, 1012-1016.
15. Sun, L. S., Paulson, K. E., Schmid, C. W., Kadyk, L. and Leinwand, L. (1984) Nucl. Acids Res. 12, 2669-2690.
16. Nomiyama, H., Tsuzuki, T., Wakasugi, S., Fukuda, M. and Shimada, K. (1984) Nucl. Acids Res. 12, 5225-5234.
17. Mager, D. L. and Henthorn, P. S. (1984) Proc. Natl. Acad. Sci. USA 81, 7510-7514.
18. Shafit-Zagardo, B., Maio, J. J. and Brown, F. L. (1982) Nucl. Acids Res. 10, 3175-3193.
19. Manuelidis, L. and Biro, P. A. (1982) Nucl. Acids Res. 10, 3221-3239.
20. Sakaki, Y., Kurata, N., Miyake, T. and Saigo, K. (1983) Gene 24, 179-190.
21. Martin, S. L., Voliva, C. F., Burton, F. H., Edgell, M. H. and Hutchison III, C. A. (1984) Proc. Natl. Acad. Sci. USA 81, 2308-2312.
22. Katzir, N., Rechavi, G., Cohen, J. B., Unger, T., Simoni, F., Segal, S., Cohen, D. and Givol, D. (1985) Proc. Natl. Acad. Sci. USA 82, 1054-1058.
23. Vieira, J. and Messing, J. (1982) Gene 19, 259-268.
24. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene

33, 103-119.
25. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning, A laboratory manual, Cold Spring Harbour Laboratory, Cold Spring Harbour, NY, pp. 368-369.
26. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. and Roe, B. A. (1980) J. Mol. Biol. 143, 161-178.
27. Carlson, J. and Messing, J. (1984) J. Biotech. 1, 253-264.
28. Kuhara, S., Matsuo, F., Futamura, S., Fujita, A., Shinohara, T., Takagi, T. and Sakaki, Y. (1984) Nucl. Acids Res. 12, 89-99.
29. Wilbur, W. J. and Lipman, D. J. (1983) Proc. Natl. Acad. Sci. USA 80, 726-730.
30. Kanehisa, M. (1984) User manual NIH, Bethesda.
31. Kunisada, T. and Yamagishi, H. (1984) Gene 31, 213-223.
32. Weiher, H., Konig, M., and Gruss, P. (1983) Science 219, 626-631.
33. Rawlins, D. R., Rosenfeld, P. J., Wides, R. J., Challberg, M. D. and Kelly, T. Jr. (1984) Cell 37, 309-319.
34. Berget, S. M. (1984) Nature 309, 179-182.
35. Bernstein, L. B., Mount, S. M. and Wiener, A. M. (1983) Cell 32, 461-472.
36. Ueda, S. Nakai, S., Nishida, Y., Hisajima, H. and Honjo, T. (1982) EMBO. J. 1, 1539-1544.
37. Roscovet, D. LE., Vodjdani, G., Lemaigre-Dubreuil, Y., Tovey, M. G., Latta, M. and Doly, J. (1985) Moll. Cell. Biol. 5, 1343-1348.
38. Tokunaga, K., Yoda, K. and Sakiyama, S. (1985) Nucl. Acids Res. 13, 3031-3042.
39. Scarpulla, R. C. (1985) Nucl. Acids Res. 13, 763-775.
40. Mason, A. J., Evans, B. A., Cox, D. R., Shine, J. and Richards, R. I. (1983) Nature 303, 300-307.
41. Goubin, G., Goldman, D. S., Luce, J., Neiman, P. E. and Cooper, G. M. (1983) Nature 302, 114-119.
42. Singer, M. F., Thayer, R. E., Grimaldi, G., Lerman, M. I. and Fanning, T. G. (1983) Nucl. Acids Res. 11, 5739-5745.
43. Comb, M., Rosen, H., Seeberg, P., Adelman, J. and Herbert, E. (1983) DNA 2, 213-229.
44. Nojima, H. and Kornberg, R. D. (1983) J. Biol. Chem. 258, 8151-8155.
45. Gebhard, W., Meitinger, T., Hochtl, J. and Zachau, H. G. (1982) J. Mol. Biol. 157, 453-471.
46. Cooke, N. E. and Baxter, J. D. (1982) Nature 297, 603-606.
47. Yang, F., Lum, J. B., McGill, J. R., Moore, C. M., Naylor, S. L., vanBragt, P. H., Baldwin, W. D. and Bowman, B. H. (1984) Proc. Natl. Acad. Sci. USA 81, 2752-2756.
48. MacGillivray, R. T. A., Mendez, E., Shewale, J. G., Sinha, S. K., Lineback-Zins, J. and Brew, K. (1983) J. Biol. Chem. 258, 3543-3553.
49. Metz-Boutigue, M.-H., Jolles, J., Mazurier, J., Spik, G., Montreuil, J. and Jolles, P. (1982) FEBS Letters 142, 107-110.
50. Brown, J. P., Hewick, R. M., Hellstrom, I., Hellstrom, K. E., Doolittle, R. F. and Dreyer, W. J. (1982) Nature 296, 171-173.
51. Diamond, A., Cooper, G. M., Ritz, J. and Lane, M-A. (1983) Nature 305, 112-116.