Isolation of cDNA Including a Reverse Transcriptase-like Sequence Transcribed from the Long Interspersed Repetitive DNA Sequence of Rat

Tadahiro Edamura, Yoshiharu Maruyama, Eiichi Soeda,* Genki Kimura** and Kazukiyo Onodera†

Department of Agricultural Chemistry, The University of Tokyo, Yayoi, Bunkyo-ku, Tokyo 113, Japan
*Tsukuba Research Center of Life Science, The Institute of Physical and Chemical Research, Koyadai, Tsukuba 305, Japan
**Department of Virology, Medical Institute of Bioregulation, Kyushu University, Maidashi, Higashi-ku, Fukuoka 812, Japan
Received February 12, 1991

The mammalian genome contains long interspersed repetitive sequences, but the role of these repetitive sequences is not clear. A cDNA clone has been isolated that contains part of the LI sequences from a cDNA library of rat liver. The DNA sequence analysis showed the homology of cDNA to several reverse transcriptases. The homology between the amino acid sequences predicted from LI consensus sequences and reverse transcriptases has been reported previously. However, this is the first isolation of a cDNA clone containing a reverse transcriptase-like sequence.

The 3Y1 cell line was established from a fibroblast of a Fischer rat embryo.1) BamHI-digested high molecular weight DNA of 3Y1 cells showed a discrete band at the size of 5.5 kb on agarose gel electrophoresis stained with ethidium bromide. We cloned the 5.5 kb DNA into pBR322 and designated it pXFR-4. Then BamHI digested 3Y1 DNA was Southern transferred and hybridized to 32P labeled pXFR-4 and a strong signal appeared at 5.5 kb among heterogeneous sizes of DNA. As a result, the 5.5 kb BamHI fragment was taken to be a fragment from the long interspersed repetitive sequences reported by Rogers.2)

LI is a long interspersed repetitive element that is present throughout marsupial and placental mammalian orders.3) The LI sequence has been estimated to be as large as 7 kb in length.4) A majority of members of this family are not full-length copies of the consensus sequence. Most members are truncated at apparently random distances from a common 3' end.5) Therefore, the extreme 3' sequences of LI are represented less frequently than extreme 3' sequences of LI. The 3' end of the repeat, the most repetitive portion of the family, is present in about 85,000 copies per haploid mouse genome.6)

It was found that an amino acid sequence predicted from an open reading frame in mouse and primate genomic LI DNA have homology to highly conserved regions in the reverse transcriptases encoded by retroviruses and retrotransposons.6,7) So it is possible that one or more of the approximately 85,000 units found in mammalian genomes might encode a functional reverse transcriptase. Probably a small percent of L1Md (mouse LI) are known to be transcribed to yield cytoplasmic polyadenylated RNA in mouse cells,8) but no cDNA clone has been identified.

In our study, one cDNA clone was isolated from a cDNA library of rat liver. Here we report the structure of one EcoRI fragment of...
L1 transcript that corresponds to the reverse transcriptase domain of some retroviruses and retrotransposons.

**Materials and Methods**

*Molecular clone.* 3Y1 high molecular weight DNA was prepared by the method of Dilella.9) The 5.5 kb DNA band appeared when BamHI-digested 3Y1 DNA was electrophoresed on an agarose gel and stained with ethidium bromide. This 5.5 kb DNA was extracted from the gel and purified. Then the DNA was cloned into the BamHI site of pBR322. pxFR-4 is one of these recombinants. It was used as a probe DNA of hybridization reactions in this study. 32P labeled pxFR-4 was prepared with multiprime DNA labeling system (Amersham).

*RNA extraction and northern blotting.* Cytoplasmic RNA of 3Y1 was extracted by the method of Perbal10) and polyadenylated RNA was selected with oligo (dT)30 latex.11) Glyoxalated RNA was transferred to nylon membranes after electrophoresis from the agarose gels.12) Hybridization reactions of RNA immobilized on membranes to 32P labeled pxFR-4 were done by the method of Meinkoth and Wahl.13)

*Screening of cDNA library.* The lambda gt11 cDNA library of rat liver (Clonetech) was screened by plaque hybridization. Hybridization reactions to nylon membranes containing replicas of bacteriophage plaques were done as mentioned above.13)

*Sequence analysis of RLcl.* DNA sequencing of RLcl, a pxFR-4 positive EcoRI fragment, was done on Exonuclease III or restriction enzyme deletions using the dideoxy chain termination method.14) The RLcl nucleotide sequence was compared with GENBANK for the DNA data search and its predicted amino acid sequence was compared with the NBRF protein data base. These data searches were done with the Integrated Database and Extended Analysis System (Kyoto University).

**Results and Discussion**

Cytoplasmic and polyadenylated RNA of 3Y1 were analyzed by Northern blotting (Fig. 1). Cytoplasmic RNA contained a population from greater than 10 kb to less than 1 kb hybridized to pxFR-4. On the other hand, polyadenylated RNA contained a large RNA of approximately 20 kb.

In mouse cells, the L1 repeat appeared to be transcribed into a heterogeneous population of RNAs.15) While discrete RNA species have been reported in a few cells such as murine lymphoid cells.8) The LI family is thought to be transposed by a mechanism involving reverse transcription by an RNA intermediate. If the LI family is still active in mammalian cells as a mobile gene, a large discrete RNA might be transcribed from a L1 sequence. The 20 kb RNA indicated in Fig. 1 is larger than a full-length L1 unit, so it might be a template for L1 reverse transcription.

As a result of plaque hybridization, one L1 positive clone, lambda RLc19, was selected out from 1.5 × 105 bacteriophage plaques of the rat liver cDNA library. The cDNA was approximately 3.4 kb long and consisted of three EcoRI fragments. They were 0.9, 1.0, and 1.5 kb in length and were analyzed by Southern blotting. The 1.0 kb EcoRI fragment that showed the strongest homology to pxFR-4 was isolated and cloned into the EcoRI site of...
Isolation of cDNA Including LI Sequence

In this study, we characterized the EcoRI fragment, designated RLcl. From the nucleotide sequencing, it appeared the RLcl was 1024 bp in length (Fig. 2). Then the GenBank Database was searched for sequences related to the RLcl nucleotide sequence. RLcl was the most similar to the DNA sequence of LINE3, a member of the LI family isolated from a rat genome (Fig. 3). It showed a homology of 98.6% to the LINE3 DNA sequence. The data suggests that one reading frame of the RLcl sequence is an ORF comprising 340 amino acids (Fig. 2). This ORF corresponds to a part of the 3900 bp ORF of the consensus sequences of L1Md, L1Hs and L1Ne (Fig. 3).

The NBRF protein database was searched for sequences related to the RLcl deduced amino acid sequence. The amino acid sequence of this ORF showed comparatively high homology with that of some reverse transcriptases and retrotransposons.

A common set of amino acids is conserved in the amino-terminal region of all retrovirus reverse transcriptases. Their regions were numbered I to III from the amino-terminus. Our alignment covers an area that contains boxes II and III (Fig. 4). The sequence called the YXDD box (as described in the single-letter code) flanked by three hydrophobic residues in box III (Fig. 4) is thought to be essential for reverse transcriptase activity. Although RLcl uses F (phenylalanine) instead of Y (tyrosine), for example, the reverse transcriptase of the yeast Ty element also uses F instead of Y. Therefore it is interesting to study whether the protein translated from this cDNA has reverse transcriptase activity.

The existence of homology between primate or rodent LI consensus sequences and several reverse transcriptases has been previously reported. However, there were little evidence showing that the reverse transcriptase-like sequence was transcribed. In this study, we isolated a cDNA clone transcribed from a LI sequence of the rat DNA and one fragment of it showed similarity to sequences of reverse

**Fig. 2.** Nucleotide and Amino Acid Sequences of RLcl.

The amino acid sequence was deduced from the nucleotide sequence of RLcl. Regions 1 to 5 indicated by horizontal bars are conserved in the amino-terminal region of retrovirus reverse transcriptases (Fig. 4).

**Fig. 3.** Consensus Restriction Enzyme Map of LINE3. Location of pXFR-4 and RLcl. The open reading frames of the LI consensus sequence are indicated by the boxes. Restriction sites are abbreviated: B, BamHI; H, HindIII. The nucleotide sequence of RLcl, indicated by the hatched box, showed a homology of 98.6% to LINE3.
Fig. 4. Comparison of Amino Acid Sequences among RLcl, LiMd, Mo-MuLV, and Ty. An identity or a favored amino acid substitution between a sequence and RLcl is indicated as a + or -, directly beneath that residue. Favored substitutions are grouped as follows: A, S, T, P, and G; N, D, E, and Q; H, R, and K; M, L, I, and V; F, Y and W. Regions 1 to 5 are indicated in Fig. 2. Boxes II and III are conserved in all retrovirus transcriptase genes.12) The YXDD box, FADDIV, is inside of box III. Positions that are invariant among retroviruses12) and the L1 family; ○, positions that are invariant only among retroviruses.

References