Nucleic Acids Research

Insertion of an Alu SINE in the human homologue of the *Mlvi*-2 locus

Agathe Economou-Pachnis and Philip N.Tsichlis

Fox Chase Cancer Center, Philadelphia, PA 19111, USA

ABSTRACT

    Fifty-nine human DNA samples derived from either normal tissues or hematopoietic neoplasias were examined for rearrangements in the Mlvi-2 locus, a putative oncogene. The rearranged Mlvi-2 sequences in one of them, a B cell lymphoma, were shown to result from the insertion of an approximately 300 bp DNA fragment that hybridized to a human Alu probe. DNA sequence analysis of both the rearranged and the nonrearranged allele around the site of the insertion revealed the following: a) the insert was 88.4% homologous to the consensus sequence of the Alu family of repeats and 75% homologous to the Alu related sequence in the human 7SL RNA; b) similar to other sequenced SINES, a poly(d•A) tract was present at the 3' end of this element; c) an 8 bp direct repeat was present at both ends of the inserted element; d) this repeat was present as a single copy in the unrearranged allele.

    We conclude from these findings that: Alu sequences can transpose and that the direct repeats flanking certain Alu SINES may be generated by the duplication of single copy cellular sequences at the site of the insertion. Furthermore the recent nature of the Alu insertion in the Mlvi-2 locus coupled to the low degree of homology of the inserted Alu to the Alu related sequence in the 7SL RNA suggest that this event did not occur via reverse transcription and reintegration of the 7SL RNA.

INTRODUCTION

    Approximately one third of the mammalian genome is composed of highly

repeated DNA sequences (1,2). These sequences, with the exception of the

satellite DNA (2), are interspersed among single copy genomic elements (2,3).

The interspersed repeated elements are divided into two major families: Long

(LINES) and Short (SINES) (2). Representatives of SINES in various mammalian

species include the Alu family in humans and other primates (4,5,6), the B_1

and B_2 families in mice (7,8,9) and the Rde-1 family in rats (10,11). It is

believed that both the LINES and the SINES are transposable and as such they

may play a significant role in genome function and evolution (2,12). There

has been circumstantial evidence that suggests the transposable nature of

Alu-like elements: DNA sequence analysis of certain Alu-like elements has

revealed flanking direct repeats 5-19 nucleotides long. These repeats are

thought to be formed by duplication of a single copy sequence at the site of insertion (13,14), and are characteristic of sequences that have been integrated in the genome (15,16).

Human Alu DNA is a head to tail dimer of two similar sequences ~130 bp long. The right monomer contains an insert (~30 bp long) which is not present in the left half (5). Alu-like sequences contain significant sequence homology to the small cytoplasmic 7SL (17) and 4.5S RNAs (18). In the human 7SL RNA, about 100 nucleotides at the 5' end and 45 nucleotides at the 3' end are homologous with the human Alu consensus sequence (19). The central portion of 155 nucleotides is unique to 7SL RNA and does not show homology with the Alu DNA (19). Since there is circumstantial evidence that RNA information can flow back into the genome probably through reverse transcription (12,20-22), it has been speculated that the prototypical Alu sequence is a processed 7SL RNA gene and that a possible mechanism for transposition of the Alu sequences may be reverse transcription and reintegration of the 7SL RNA (23,24).

The purpose of the work presented in this report was to identify and characterize DNA rearrangements occurring in the human homologue of the Mlvi-2 locus in a variety of lymphoid cell neoplasias. The Mlvi-2 locus, a putative oncogene, was originally defined as one of at least five common regions for proviral integration in Moloney murine leukemia virus (MoMuLV) induced rat thymomas (25-29). In this report we present evidence that one allele of the human homologue of the Mlvi-2 locus was rearranged in a B cell lymphoma and that the rearrangement was due to the insertion of an Alu SINE. These data support the transposable nature of the Alu SINES and allow that such transposition events may be involved in oncogenesis.

MATERIALS AND METHODS
Human DNA samples and probes
    The human DNA samples were kind gifts from F. Ruscetti, D. Watson, A. J. Bakshi and E. Jaffe (National Institutes of Health). The rat Mlvi-2 probes have been described previously (26,27).
Southern blot analysis of genomic and cloned DNA
    The DNA was digested with restriction endonucleases and after agarose electrophoresis and transfer to nitrocellulose it was hybridized to the appropriate nick translated DNA probes. Hybridizations were performed in 50% Formamide and 5X SSC at 42°C for approximately 15 hours. The filters were washed in 0.1X SSC and 0.1% sodium dodecyl sulfate at 65°C for 30-60 min.

When rat DNA probes were used to detect sequences in human DNA the hybridizations were performed in 40% Formamide and 6X SSC at 37°C and the filters were washed in 1X SSC and 0.1% SDS also at 65°C.

Genomic DNA cloning and DNA sequencing

The genomic DNA (~0.25 mg) was digested with SacI and it was fractionated by preparative agarose gel electrophoresis. The fractions containing the desired DNA fragments were ligated into the SacI arms of the bacteriophage lambda vector, λgtWES.λB' (30). Following in vitro packaging the resulting phage was plated on the Escherichia coli K12 strain LE 392. DNA from the recombinant phage plaques was transferred onto nitrocellulose filters and it was hybridized to the appropriate DNA probes. DNA sequencing was done by using the dideoxy chain termination method (31). HpaII and HaeIII fragments (Fig. 2B) were cloned in M13mp10 and mp11 and they were sequenced in both directions.
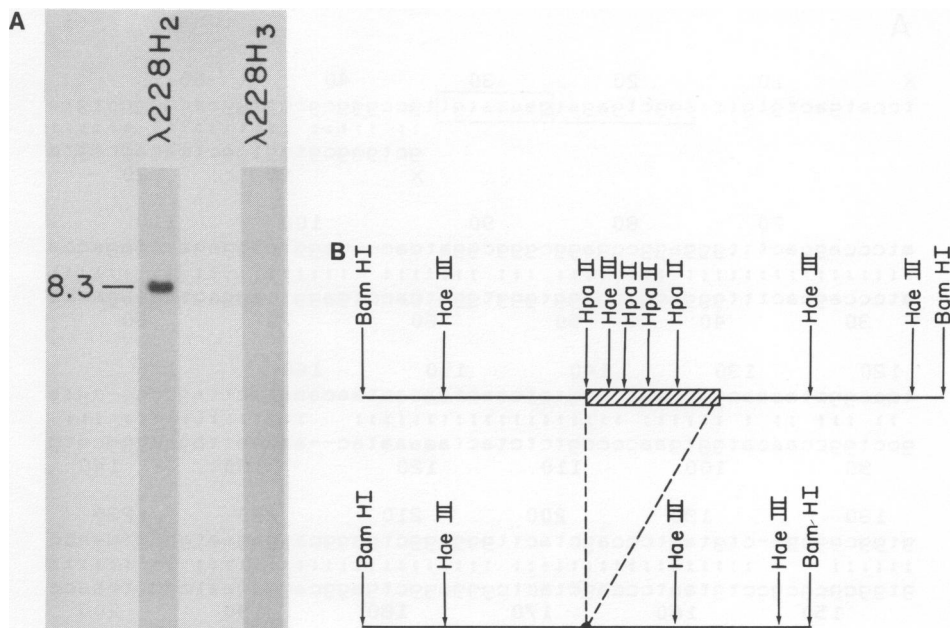

RESULTS

A SacI fragment from human DNA, homologous to Mlvi-2, was cloned in the SacI site of the bacteriophage λgtWES•λB' (30). A restriction endonuclease map of the cloned sequences is shown in Figure 1C. To determine the possible presence of Mlvi-2 rearrangements in human tumors, a ^{32}P labeled probe, derived from the plasmid pA$_4$ (Figure 1C) was hybridized to SacI digested DNA from 59 independent human samples (21 normal tissues, 4 T cell acute lymphoblastic leukemias, 7 T cell lymphomas, 9 T cell lymphocytic leukemias, 1 thymoma, 5 nodular poorly differentiated lymphocytic lymphomas, 2 nodular mixed cell type lymphomas, 1 leukemic reticuloendotheliosis, 1 diffuse histiocytic lymphoma, 1 Burkitts lymphoma, 5 diffuse poorly differentiated lymphocytic lymphomas and 2 B cell lymphomas). Only one sample, obtained from a cell line derived from a human B cell lymphoma (kindly provided by Dr. F. Ruscetti, Frederick Cancer Research Facility), had a rearranged Mlvi-2 locus (Figure 1A). The Mlvi-2 heterozygocity detected in this sample could be due to the insertion of an approximately 0.3 kb DNA element lacking a SacI site. Digestion of the same tumor DNA with BamHI and hybridization to a probe derived from plasmid pA$_4$ demonstrated that the insertion occurred within the DNA sequences represented by this plasmid (Figure 1B). To confirm this finding the two Mlvi-2 alleles were cloned into the SacI site of the lambda vector λgtWES•λB'. Restriction endonuclease mapping of the cloned sequences, representing the two alleles, confirmed the presence of the insert. Because of its size, it was suspected that this element may represent a member of the
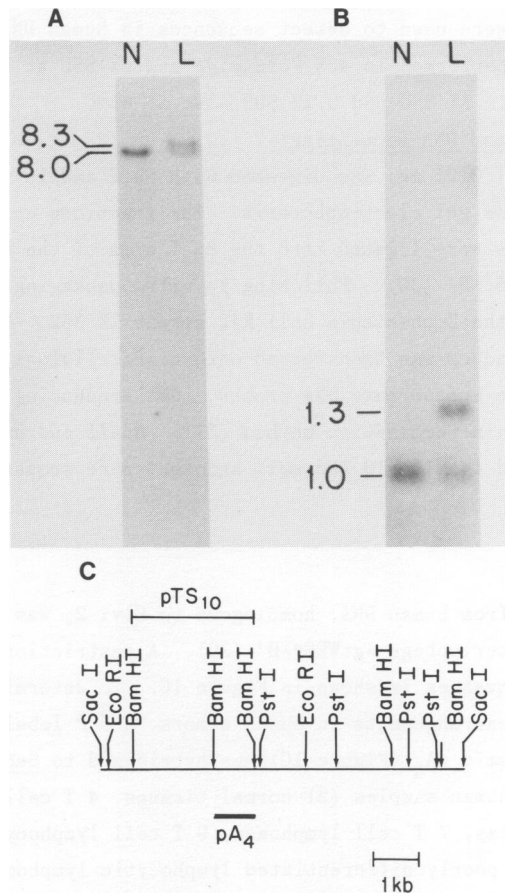
Fig. 1. Restriction fragment length polymorphism in the human homologue of
the Mlvi-2 locus.
A. Southern blot analysis of SacI digested normal human DNA (lane N) and DNA
derived from a human B cell lymphoma (lane L) hybridized to a ^{32}P labeled
probe derived from plasmid pA$_4$ (panel C).
B. Southern blot analysis of BamHI digested DNA from the same normal (lane N)
and lymphoma (lane L) tissues shown in panel A, following hybridization to a
^{32}P labeled probe derived from plasmid pA$_4$ (panel C).
C. Restriction endonuclease map of the cloned human Mlvi-2 sequences. The
line above the map, designated pTS10, indicates the DNA fragment that hybrid-
izes to the rat Mlvi-2 probe utilized for the cloning of the human homologue.
The bar designated pA$_4$ below the map indicates the origin of the plasmid that
was used as a probe in the experiments in panels A and B.

Alu family of repeats. This was confirmed by hybridization of the BLUR-8 Alu

probe (5) to both human Mlvi-2 clones which revealed homology only to the

clone containing the insert (Figure 2A). A restriction endonuclease map of

A

λ228H₂   λ228H₃

8.3 —

B

Bam HI   Hae III   Hpa II Hae III Hpa II Hae III Hpa II   Hae III   Hae III Bam HI

Bam HI   Hae III   Hae III   Hae III Bam HI

Fig. 2. Insertion of an Alu SINE in the human homologue of the Mlvi-2 locus.
A. Hybridization of the BLUR-8 Alu probe to filter immobilized DNA of the
clones λ228H₂ and λ228H₃ representing the two alleles of the Mlvi-2 locus in
the human B cell lymphoma presented in this report.
B. Restriction endonuclease map of the plasmid pA₄ with (upper panel) and
without (lower panel) the Alu insert. The Alu insert in the upper panel is
shown as a hatched bar. The site of the Alu insertion is indicated by the
convergent dotted lines in the lower panel.

the cellular DNA sequences in plasmid pA₄ with and without the inserted

element is shown in Figure 2B.

    Nucleotide sequence analysis of the cloned DNA derived from the two

allelic forms of Mlvi-2 (Figure 3) revealed the following: a. The insert

indeed belongs to the Alu family of repeats since it was shown to be 88.4%

homologous to the consensus sequence (5) of these repeats. However, compari-

son to the Alu homologous sequences of the 7SL RNA (19) revealed only a 75%

homology. The comparison was done by aligning the Alu-homologous sequences of

the 7SL RNA (omitting the 155 bp long unique sequence) with the right monomer

of the Mlvi-2 Alu (75% homology). The comparison of the 7SL RNA to the left

monomer of the Mlvi-2 Alu revealed only a 63% homology, due to the absence of

the 30 bp insert from the left monomer of the Alu repeat. b. A poly(dA) tract

was detected at the 3' end of this element. c. A short, 8 bp direct repeat

A

```
X          10             20             30             40             50
tccatgactgtgtctaggctgagatgaaaatgttgccgggcgcggtggctcacgcctgta
                                   ::  :::::  ::::::::::::  ::::::
                                   gctgggcgtggtggctcacacctgta
                                   X          10             20

           70             80             90            100            110
atcccagcactttgggaggccgaggcgggcggatcac--gaggtcaggagatcgagacca
::::::::::::::::::::::::::::::::::  :::  :::::::    ::::::::::::  ::  ::::::
atcccagcactttgggaggccgaggtgggtggatcacctgaggtcaggagttcaagacca
   30             40             50             60             70             80

120          130            140            150            160            170
tcccggctaaaacggtgaaacccgtctctactaaaaataacaaaaaattagccgg-cgta
  ::  :::  ::  :  ::::::  ::::::::::::::::::::::       ::::::::::::::  :::
gcctggccaacatggtgaaccccgtctctactaaaaatac--aaaaattagccgggcgtg
   90            100            110            120            130            140

180          190            200            210            220            230
gtggcggcg--ctgtagtcccagctacttgggaggctgaggcaggagaatggcgtgaacc
::::::::       :::::  ::::::::::::  :::::::::::::::::::::::::  ::  ::::::
gtggcgcgcgcctgtaatcccagctactcgggaggctgaggcaggagaatcgcttgaacc
   150            160            170            180            190            200

240          250            260            270            280            290
cgggaggcggc--ttgcagtgagccgagatcccgccactgcactccagcctgggcgacag
:  :::::  ::       :::::::::::::::::::::  ::::::::::::::::::::::::::::::::  ::::
caggaggtggaggttgcagtgagccgagatcgcgccactgcactccagcctgggcaacag
   210            220            230            240            250            260

300          310            320            330            340            350
agcgagactccgtctcaaaaaaaaaaaaaaaaaaaaaaaaaagaaaatgtcatctgtcct
::::::::::::::    ::::::::::::::::::::::::::::::::::::
agcgagactccatctcaaaaaaaaaaaaaaaaaaaaaaaa
   270            280            290             X
```

B

```
gtgtctaggctgagatgaaaatgtcatctgtcctttatgc
```

Fig. 3. Nucleotide sequence of the Alu insert, its flanking sequences, and
the target site before the insertion.
A.   Comparison to the concensus Alu sequence: The first line shows the
nucleotide sequence of the Alu insert and its flanking sequences. The con-
sensus sequence for the human Alu family is shown below. The direct repeats
flanking the Alu sequence are boxed.
B.   Sequence of the Alu target site before the insertion: The octanucleotide
at the target site before the insertion is boxed. The sequences flanking the
octanucleotide of the target site before the insertion (panel B) and after the
insertion (panel A) are underlined.

was detected at the ends of the Alu sequence. d. This 8 bp sequence was
present in single copy in normal human DNA.

DISCUSSION
    The observed polymorphism in the human Mlvi-2 locus and the direct repeats
detected in the boundaries of the inserted Alu sequence indicate that the Alu
repeats can transpose. The described transposition of the Alu repeat in the
Mlvi-2 locus may be either a somatic event related to tumor induction, or a
germ line event that occurred earlier in evolution and continues segregating
in the human population. Since normal tissue from the individual whose B cell
lymphoma gave rise to the cell line with the polymorphic Mlvi-2 locus was not
available, nor was there any access to his relatives, this question could not
be answered. However, since the Alu SINE insertion was detected in a lymphoid
cell tumor and since the insertion had occurred in the domain of the Mlvi-2
locus, a DNA region associated with lymphoid cell neoplasias (25-27), we
suggest that this insertion may be a somatic event related to tumor induction.
    Independently of whether the described Alu insertion represents a somatic
or a germ line event, the Mlvi-2 polymorphism presented in this report
supports the hypothesis that the Alu SINES in humans can transpose. This
hypothesis is further supported by the finding that the Mlvi-2 Alu SINE is
flanked by direct repeats that appear to have been generated by duplication of
a single copy cellular sequence. This interesting structural feature of SINES
had been inferred previously from the sequence of SINES inserted within
pseudogenes (13) and repeated DNA sequences such as satelite DNA (14).
    The low frequency of detection of the Mlvi-2 polymorphism among human DNA
samples suggests that, if the Alu insertion we described was a germline event,
it occurred more recently than the insertion of most Alu sequences that are
fixed in the human genome. It has been proposed that transposition of at
least some of the Alu sequences may occur via reverse transcription and rein-
tegration of the Alu homologous sequences in the 7SL RNA (23,24). Therefore,
we examined the degree of homology between the Mlvi-2 Alu sequence and the
published sequence of the human 7SL RNA (19). We reasoned that if this was a
recent insertion and if it occurred via reintegration of 7SL RNA sequences,
the degree of homology between the Mlvi-2 Alu and the 7SL Alu sequences would
be higher than that detected between the Mlvi-2 Alu and the consensus Alu
sequence. This analysis revealed that the homology between the Mlvi-2 Alu and
the 7SL RNA Alu-homologous sequence was 75% while the homology between the
Mlvi-2 Alu and the consensus Alu sequence was 88.4%. These data therefore

indicate that the insertion of the <u>Mlvi</u>-2 Alu did not occur via reverse
transcription and reintegration of the 7SL RNA.

    Germ line polymorphisms of single copy loci due to the insertion of
members of the families of the interspersed repeated sequences have been shown
before on two different occasions, in rats. In one case we showed that the
<u>Mlvi</u>-2 and the immunoglobulin heavy chain (<u>Igh</u>) locus in the rat are poly-
morphic due to the insertion of members of the family of the long interspersed
repeated elements (32). In another case Schuller et al. showed that sequences
in the 5' flanking region of the rat prolactin gene is polymorphic due to the
insertion of a short interspersed repeated element (33). Although the active
transposition of repeated DNA sequences in mammalian DNA has not been shown
directly yet, these findings present strong evidence that such transpositions
indeed occur.

REFERENCES
1. Britten, R.J. and Kohne, D.E. (1968) Science 161, 529-540.
2. Singer, M.F. (1982) Int. Review of Cytol. 76, 67-112.
3. Davidson, E.H., Hough, B.R., Amenson C.S. and Britten, R.J. (1973)
    J. Mol. Biol. 77, 1-23.
4. Houck, C.M., Rinehart, F.P., Schmid, C.W. (1979) J. Mol. Biol. 132, 289-
    306.
5. Deininger, P.L., Jolly, D.J., Rubin, C.M., Friedmann T. and Schmid, C.W.
    (1981) J. Mol. Biol. 151, 17-33.
6. Grimaldi, G., Queen C. and Singer, M.F. (1981) Nucleic Acids Res. 9,
    5553-5568.
7. Kramerov, D.A., Grigoryan, A.A., Ryskov A.P. and Georgiev, G.P. (1980)
    Nucleic Acids Res. 6, 697-713.
8. Krayev, A.S., Kramerov, D.A., Skryabin, K.G., Ryskov, A.P., Bayev, A.A.
    and Georgiev, G.P. (1980) Nucleic Acids Res. 8, 1201-1215.
9. Krayev, A.S., Markusheva, T.V., Kramerov, D.A., Ryskov, A.P., Skryabin,

K.G., Bayev, A.A. and Georgiev, G.P. (1982) Nucleic Acids Res. 10, 7461-7475.

10.   Barta, A., Richards, R.I., Baxter J.D. and Shine, J. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 4867-4871.

11.   Sutcliffe, J.C., Milner, R.J., Bloom F.E. and Lerner, R.A. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 4942-4946.

12.   Sharp, P.A. (1983) Nature 301, 471-472.

13.   Lemischka, I. and Sharp, P.A. (1982) Nature 300, 330-335.

14.   Grimaldi G. and Singer, M.F. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 1497-1500.

15.   Shapiro, J.A. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 1933-1937.

16.   Temin, H.M. (1980) Cell 21, 599-600.

17.   Weiner, A.M. (1980) Cell 22, 209-218.

18.   Haynes, S.R., Toomey, T.P., Leinwand, L. and Jelinek, W.R. (1981) Mol. Cell. Biol. 1, 573-583.

19.   Ullu, E., Murphy, S. and Melli, M. (1982) Cell 29, 195-202.

20.   Jagadeeswaran, P., Forget, B.G. and Weissman, S.M. (1981) Cell 26, 141-142.

21.   Hollis, G.F., Hieter, P.A., McBride, O.W., Swan, D. and Leder, P. (1982) Nature 296, 321-325.

22.   Baltimore, D. (1985) Cell 40, 481-482.

23.   Ullu, E. and Tschudi, C. (1984) Nature 312, 171-172.

24.   Chen, P.-J., Cywinski, A. and Taylor, J.M. (1985) J. Virol. 54, 278-284.

25.   Tsichlis, P.N., Strauss, P.G., Hu, L-F. (1983) Nature 302, 445-449.

26.   Tsichlis, P.N., Hu, L-F. and Strauss, P.G. (1983) UCLA Symp. Mol. Cell. Biol. 9, 399-415.

27.   Tsichlis, P.N., Strauss, P.G. and Lohse, M.A. (1985) J. Virol., in press.

28.   Lemay, G. and Jolicoeur, P. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 38-42.

29.   Steffen, D. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 2097-2101.

30.   Enquist, L., Tiemeier, D., Leder, P., Weisberg, R. and Sternberg, N. (1976) Nature 259, 596-598.

31.   Sanger, F., Nicklen. S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.

32.   Economou-Pachnis, A., Lohse, M.A., Furano, A. V. and Tsichlis, P.N. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 2857-2861.

33.   Schuler, L.A., Weber, J.L. and Gorski, J. (1983) Nature 305, 159-160.