Volume 14 Number 12 1986 Nucleic Acids Research

Structural variations in the *Drosophila* retrotransposon, *17.6*

Satoshi Inouye, Kazue Hattori, Shunji Yuki and Kaoru Saigo*

Department of Biochemistry, Kyushu University 60 School of Medicine, Fukuoka 812, Japan

Received 23 April 1986; Revised and Accepted 23 May 1986

ABSTRACT

More than 21 members of 17.6, a Drosophila retrotransposon, were isolated and their possible structural changes were examined by restriction mapping, blot hybridization, heteroduplex analysis and nucleotide sequence determination of long terminal repeats (LTRs). At least 7 members were found to suffer with terminal or internal long deletions. No pair of LTRs having an identical nucleotide sequence was found either within an element or between elements. Although an initiation site for the presumable genome-sized transcript of 17.6, a potential substrate for reverse transcription on translocation, was identified within the left-hand LTR, our results as a whole support the notion that the majority of 17.6s have continued to reside for a long period of time at their present chromosomal loci and hence the rate of translocation of 17.6 is very low.

INTRODUCTION

17.6 is a member of copia-like transposable genetic elements (retrotransposons) in Drosophila melanogaster and has been suggested to have about 40 copies per haploid chromosome (1). In a previous experiment (2), we determined the complete nucleotide sequence of a 17.6 element inserted in a histone gene and found the 17.6 genome to include three long open reading frames (ORFs) comparable in size and location with retroviral gag, pol and env. We also found the second ORF of 17.6 not only to contain the nucleotide sequence coding for reverse transcriptase-like enzyme but also to be very similar in overall structure to the pol gene of Moloney murine leukaemia virus (2,3). Since subsequent nucleotide sequence analyses (4-10) have revealed the presence of similar pol-like genes in several other copia-like elements in Drosophila and yeast, it is now quite reasonable to presume that the translocation of copia-like

elements is generally carried out via RNA intermediates by reverse transcription. In fact, Boeke et al. (11) have recently presented more direct evidence for this notion at least in the case of a yeast copia-like element, Ty.

Here, we examined possible structural changes among more than 21 members of a 17.6 family, representing about half the total per haploid chromosome, to see how fast these elements change their chromosomal positions. Although an initiation site for the presumable, genome-sized transcript of 17.6 , a potential substrate for reverse transcription on translocation, was identified within the left-hand LTR, our results as a whole support the notion that the majority of 17.6s have continued to reside for a long period of time at their present chromosomal loci and hence the rate of translocation of 17.6, if it actually occurs, is very low.

MATERIAL AND METHODS

Materials. The restriction enzymes, DNase, DNA ligase and reverse transcriptase were obtained from commercial sources. ^{32}P-labelled nucleotides were purchased from New England Nuclear.An oligonucleotide primer(3'GGAACTACGCCATTGGTCTAAACGG5') for primer extention was synthesized using a Model 380A synthesizer (Applied Biosystems).

General Methods. Restriction enzyme digestions, agarose gel electrophoresis, subcloning of restriction segments and blot hybridization were carried out essentially as described by Maniatis et al. (12). As a cloning vehicle for sub-cloning, only pUC9 (13) was used. The DNA probe was labelled with ^{32}P by kination. DNA sequencing was conducted by the Maxam-Gilbert method (14). Heteroduplex analysis was carried out according to Modolell et al.(15).

Phages and Plasmids. The origin of λhist17.6 has been described previously (16). Other 17.6 containing phages were screened from an EcoRI partial library of D. melanogaster (Oregon R) embryonic DNA essentially as described elsewhere (16).

Preparation of RNA and a primer extension experiment. GM2 cells were cultured in 400ml D20 medium to late log phase as described previously (17) and collected by low-speed centrifugation. RNA was prepared from GM2 cells essentially as described by Deeley et al.(18). Using 1mg/ml of GM2 total RNA, a primer extention experiment was carried out as described elsewhere (4).

RESULTS AND DISCUSSION

Variation in overall structure of 17.6 elements

Using a 2.1 kb (kb=1000 base pairs) HindIII fragment of

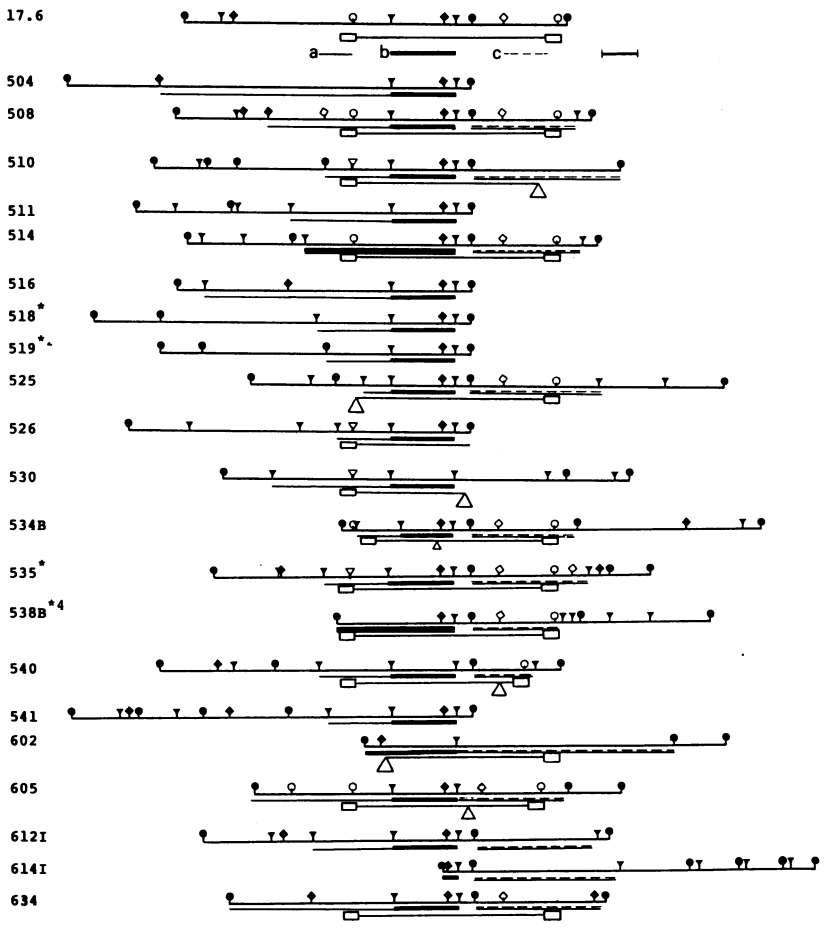

Fig. 1. Restriction maps of cloned Drosophila fragments containing 17.6-related sequences along with the locations and deduced structures of 17.6s. The top line labelled 17.6 shows the restriction map of the Drosophila insert in λhist17.6. The fragments used for probes for hybridization are marked by three underlines a, b and c. ▼, HindIII; ●, EcoRI; ◆, BamHI; ○, ClaI; ▽ SinI. In the case of ClaI and SinI, only the sites within some LTRs are shown. ———, homology with probe a for detection of LTR sequences; ———, homology with probe b; _ _ _ _, homology with probe c. Asterisks show the presence of plural clones exhibiting identical restriction maps. The locations and the deduced structures of 17.6s are schematically illustrated just beneath the restriction map of each clone. The locations of deletions are shown by triangles, while those of LTRs, by boxes. The estimated lengths of the deletions are as follows; λ510, 0.6 kb; λ525, 0.5kb;λ530, 3.3kb;λ540, 1.0kb;λ602, 2.2kb;λ605, 1.2kb; λ534B,0.4kb.

17.6 in λhist17.6 (see an underline b in Fig.1) as a probe, 50 recombinant clones were isolated from about 5x 10^4 phages in a cloned library of Drosophila melanogaster DNA. Mapping of restriction enzyme sites followed by Southern hybridization showed that these recombinant clones as a whole represent at least 21 chromosomal loci corresponding to half the total 17.6-insertion sites per haploid chromosome (Fig.1). Since 17.6 genomes in several clones were found to be artificially truncated at internal EcoRI sites, their structures were not analyzed furthermore except for that in λ526. The precise locations of the ends of 17.6s together with those of possible internal deletions were determined in the majority of the remaining clones by mapping the cutting sites of restriction enzymes, heteroduplex analyses and/or partial nucleotide sequencing. The structures of thirteen 17.6 elements thus determined are schematically shown in Fig.1.

About half these thirteen elements (those in λ510, λ525,λ530, λ534B,λ540, λ602 and λ605) were each found to have a large deletion, internal or terminal, with respect to the authentic 17.6 in λhist17.6. The distribution of these deletions appears random. Some examples of electron micrographs of heteroduplex molecules between λhist17.6 and other clones are shown in Fig.2. In the case of 17.6 in λ605, which contains a 1.2kb-long internal deletion, the duplex region with the right-hand LTR (see Fig.1) was occasionally observed to be single-stranded (see the regions with large arrowheads in Figs.2D and E). Though the reason for this is unclear, it is unlikely that this structural change arises from possible variation in nucleotide sequence in that particular region of 17.6 in λ605. In fact, except for a 50bp(base pair)-long, highly reshuffled region located immediately next to the deletion site (Fig.3), only one nucleotide change could be found in a total of 1100 nucleotides flanking the deletion site in 17.6 in λ605 (data not shown). 17.6 in λ534B contained a very short (about 200bp-long), internal deletion which could not be clearly visualized by electronmicroscope (Fig.2B): the location of this short deletion was determined by fine restriction mapping not shown here.

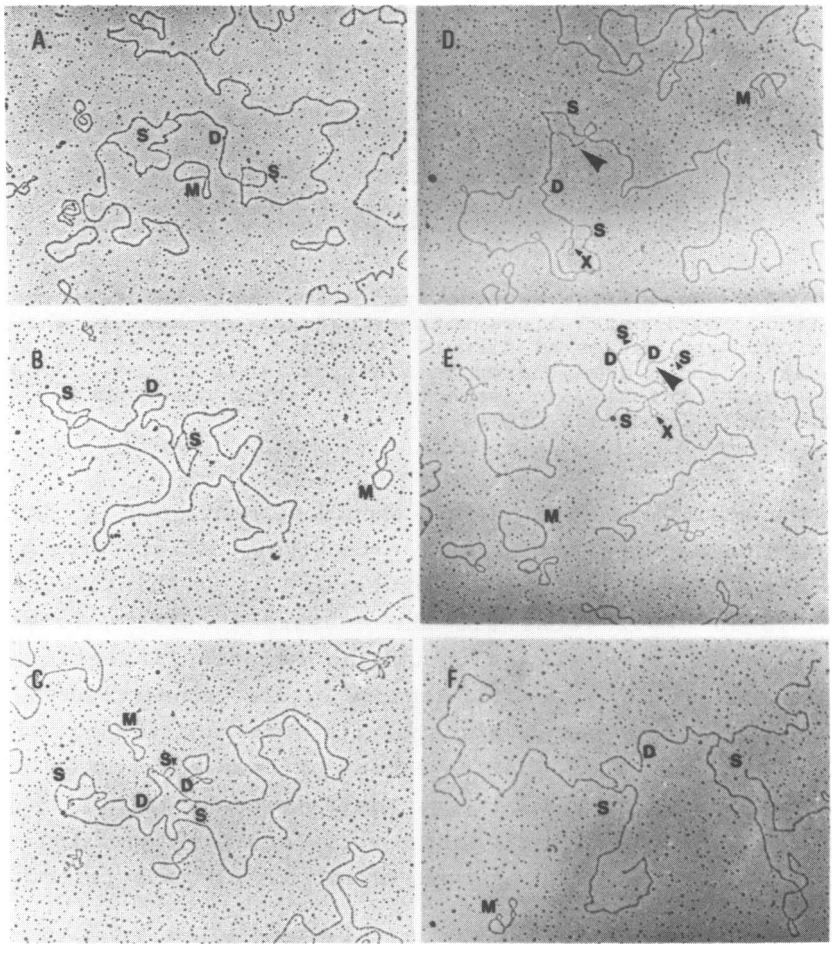

Fig. 2. Electron micrographs of heteroduplex molecules between the DNA from a cloned phage having 17.6-related sequence and λhist17.6 DNA. A, λ530; B, λ534B; C, λ540; D and E, two different examples of λ605; F,λ634. M, marker DNA (pBR322 with a molecular weight of 4.4kb); D, double-stranded region; S, single-stranded region. Large arrows in D and E show two variations in the heteroduplex molecules of λhist17.6 and λ 605. Arrows labelled X indicate stem and loop structures found in the DNA from λ605. The lengths of the duplex regions expected to include 17.6 are estimated as follows. A, (4.1 ± 0.2)kb. B, (7.0 ± 0.3)kb. C; (5.3 ± 0.3 and 1.1 ± 1.0)kb; D, (3.9 ± 0.2)kb; E, (3.8 ±0.2 and 2.5 ± 0.2)kb; F, (7.1 ± 0.3)kb. (A slightly reduced value in F with respect to the authentic 17.6 is at least partly due to a high degree of variation in nucleotide sequence at the 3'terminal regions of LTRs (see Fig.3).)

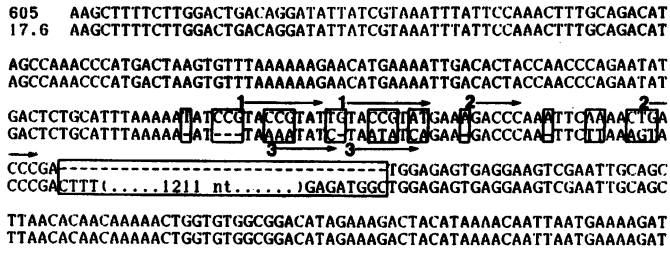

Fig. 3. Nucleotide sequence variation near the deletion site in λ605. Horizontal lines with arrowheads and numbers show the locations and sizes of direct repeats. Note that, although repeats 1 and 3 are located at identical loci in the 17.6s of λ605 and λhist17.6, respectively, their sequences differ considerably from each other.

One additional finding that should be mentioned is that the restriction map of 17.6 in λ602 extensively differs from those of other 17.6s including the authentic 17.6 in λhist17.6. That is, 17.6 in λ602 lacks all the conserved cutting sites of EcoRI, BamHI and SalI, although no deletion/insertion was observed in the corresponding regions by electronmicroscopy (see Fig.1). Because of the deletion of its left-hand LTR-containing region, 17.6 in λ602 must be a kind of dead element at least with respect to translocation. There is the possibility that this 17.6 became a dead element a very long time ago, followed by introduction of the observed variations in the restriction sites.

Variations in nucleotide sequences of LTRs.

To ascertain the degree of variation at the level of nucleotide sequence, we determined the complete nucleotide sequences of the right-hand LTRs of nine 17.6 elements. As shown in Fig.4, these nucleotide sequences are similar but not identical to one another. Since we determined the LTR sequences of about one quarter of 17.6s in a haploid chromosome, the above finding that there is no pair of 17.6s with LTRs having identical sequences may suggest that the last duplication and insertion of the majority of these 17.6s occurred long ago so that even newly inserted 17.6s have become the targets for additional mutations with respect to parental ones.

In addition to many 1-3 bp long nucleotide changes, several long (6-34bp) insertions and deletions were detected, most explainable by duplication of neighbouring sequences or deletion of already duplicated sequences, as indicated by the horizontal lines with arrowheads in Fig.4. In particular, the 3'terminal one third of the LTR of 17.6 in λ634 was found to have an exceptionally varied nucleotide sequence with only 45 % homology with the counterpart of the authentic 17.6 in λhist17.6. As described below, this highly varied region corresponds to R + U5 in the LTR of the authentic 17.6. To our knowledge, no other copia-like elements have been shown to include a member having LTRs with such a high degree of variation (19-23). However, this unusual variation in LTR of 17.6 in λ634 appears not to influence the sequence-specific insertion of 17.6 (24), since, as shown previously (24), the insertion of this 17.6 resulted in a duplication of ATAT as also in the case of all other 17.6 elements.

In addition to these, 9 changes in nucleotide sequences were found to be shared by plural elements (Fig.4), indicating 17.6 LTR to possibly contain many polymorphic nucleotide sequences having equal biological activity. Since combinations of polymorphic changes among 17.6s appear to be rather random, the present distribution of combinations of polymorphic changes in 17.6s may possibly result in part from either cumulative recombinations between LTRs of different 17.6s or plural independent mutational events or both.

The nucleotide sequences of both LTRs of a copia-like element must be perfectly corrected at the time of insertion, if its translocation is carried out via an RNA intermediate by a mechanism similar to that in retrovirus (25). In fact, in the case of copia, such a phenomenon has been reported (19). To examine this possibility in 17.6, all or some of the nucleotide sequences of the left-hand LTRs of five 17.6 elements were determined and compared with those of their right-hand counterparts.The results shown in Fig.4 are consistent to some degree with our expectation, since all the 17.6 elments examined here as well as the authentic 17.6 in λhist17.6 have at least a few minor base differences in both LTRs. In particular, the LTRs

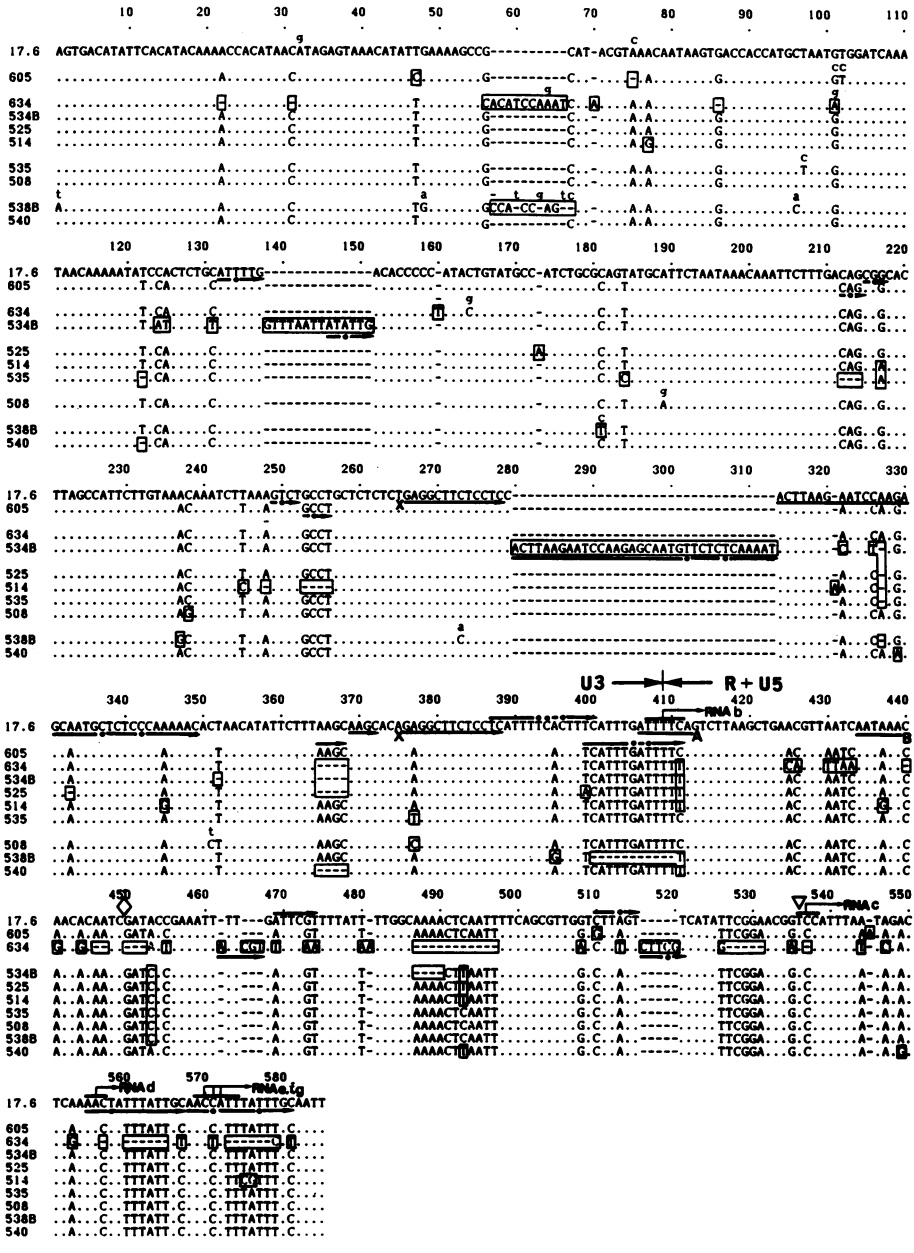

Fig. 4. The nucleotide sequences of LTRs of 17.6s from various clones. Capital letters on the line labelled by 17.6 show the nucleotide sequence of the left-hand LTR of the authentic 17.6 in λhist17.6, whereas those on other lines

exhibit the sequences of the right-hand LTRs of the 17.6s from various clones examined in the present research. Small letters appearing above these sequences indicate nucleotide sequence changes observed in the opposite LTRs of the same elements. Note that the complete nucleotide sequences of the left-hand LTRs were determined in 17.6s in λ605, λ634 and λ508, while only the 5'terminal one third and two thirds of the left-hand LTRs were sequenced for 17.6s in λ535 and λ538B, respectively. Bars indicate deletions, and dotted lines the absence of nucleotide change with respect to the left-hand LTR of the authentic 17.6. Boxes indicate variation in the nucleotide sequences from the left-hand LTR of the authentic 17.6. Direct repeats and locations of RNA starts are indicated by horizontal and L-shaped arrowheads, respectively. Dots in some horizontal arrowed lines denote mismatches and X identifies a particular repeat discussed in the text. A large lozenge and triangle, located respectively at 450 and 536 nucleotides from the 5' end, indicate the locations of ClaI and SinI characteristic of 17.6 LTR.

of 17.6 in λ538B was found to have many right-left differences. However, at the same time, that the degree of difference in LTR sequence within an element is much less than that among different elements in most cases should be mentioned. Note the right-left difference in LTR sequence in 17.6 in λ634 having an unusually high degree of variation, is also very small. We interpret these results as indicating that 17.6 possibly has, basically, the ability to correct LTR sequences on translocation, but since they may have resided for a long period of time at the present chromosomal loci, the LTRs of the majority of 17.6s have incurred mutations, the number and extent of which depending on the duration of this period.

Variation in RNA start sites from within left-hand LTR.

Our previous nucleotide sequence analysis of the entire 17.6 element (2) showed the direction of transcription of all three major open reading frames of 17.6 to be from left to right in the case of orientation of 17.6s as in Fig.1. The genome-sized RNA of copia was also shown to be transcribed from within LTR (4,27). To see if transcription of the 17.6 genome starts from within left-hand LTR, all the RNA extracted from Drosophila tissue cultured cells (GM2 cells) was examined by primer extension, using a 25 nucleotide-long synthetic oligonucleotide complementary to the region P in Fig.5 as a primer. In Fig.6, seven intensive bands along with many faint ones and smear in

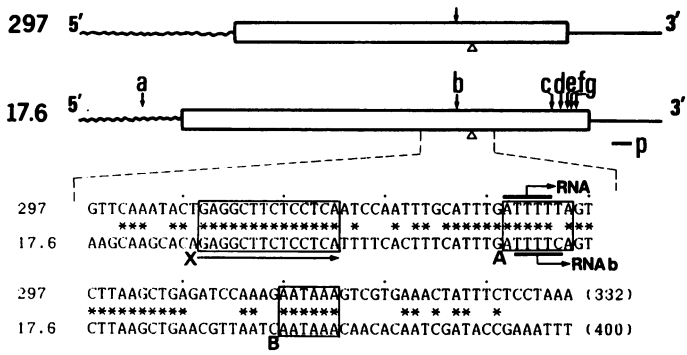

Fig. 5. Locations of RNA starts and nucleotide sequence conservation between 17.6 and 297. The locations of RNA starts a-g of 17.6 (see Fig.6) and that of the major RNA of 297 (unpublished observation) are shown by vertical arrowheads. Wavy lines indicate host DNA, while the boxes show the size and locations of LTRs. A bar labelled P indicates the location of the primer used for primer extention (region P; nucleotides 542-547 from the 5'end of the authentic 17.6 (see Fig.1 of ref. 2).). Triangles show the locations of the putative polyadenylation signals. The nucleotide sequences of 17.6 and 297, delimited by broken lines, are shown in the lower half. Note that the location of RNA start b of 17.6 and that of the major 297 RNA are identical to each other within a limit of error. Boxes A and B indicate the nucleotide sequences assumed in a previous experiment to correspond to the Hogness box and polyadenylation signal of avian leukaemia-saroma virus (1), while box X shows the sequence conserved between 17.6 and 297. As shown by arrowheads X in Fig.4, this conserved sequence is repeated within the 17.6 LTR. Asterisks show the homology between 17.6 and 297 sequences.

the high molecular-weight region can be seen. Since this banding pattern did not change with the further addition of reverse transcriptase (data not shown), 17.6s may possibly have heterogeneous RNA start sites in or near left-hand LTRs.

As summarized in Fig.7, RNA starts, c-g, are located near or within a 9 nucleotide-long canonical sequence, CCATTTATT, whereas RNA start b is localized in the very vicinity of TCAGT, a sequence proposed as being near the major RNA starts of various Drosophila copia-like elements including mdg1 (26). Since a putative polyadenylation signal AATAAA is located upstream of RNA starts c-g for all LTR sequences presented in Fig.4, these sites do not appear to participate in the transcription of the genome-sized RNA of 17.6 expected to serve

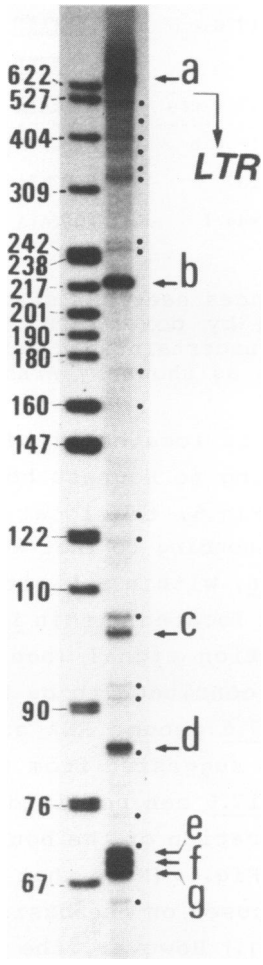

Fig. 6. A primer extention pattern of 17.6 RNA. Horizontal arrowheads labelled by alphabetical letters show the major RNA starts of 17.6 in or near the left-hand LTR. Minor RNA starts are indicated by dots. The range of left-hand LTR is shown by an L-shaped arrowhead. The markers used were HpaII digests of pBR322 (4).

as a substrate for reverse transcription on translocation. RNA start a ,located upstream of the 5'end of the left-hand LTR of the authentic 17.6, must be the strongest initiation site for the transcription of 17.6 from an outside promoter. Similar transcription from outside of LTR has been found in the case of

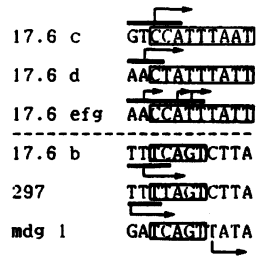

Fig.7. Consensus sequences near RNA starts. Deduced consensus sequences are enclosed by boxes. Thick bars with L-shaped arrowheads indicate the uncertainty in determining RNA starts. The RNA start of mdg1 is as shown by Arkhipova et al.(26).

copia (27). RNA start b is located within a region previously proposed as corresponding to Hogness box of avian leukaemia-sarcoma virus (1). In Fig.5, the location of RNA start b is within the region corresponding to that with the major RNA start of 297, a sibling of 17.6, within a limit of error. Since it is the only major RNA start located within 17.6 LTR and upstream of a putative polyadenylation signal (see box B in Fig.5), RNA start b is tentatively concluded to be the initiation site of transcription of the 17.6 genome RNA expected to be reverse transcribed. It is also suggested from these results that, as with copia (4), LTR of 17.6 can be divided into three regions U3, R, and U5 and the location of the boundary between U3 and R + U5 is as indicated in Fig.4. (Note that this location differs somewhat from that proposed on the basis of the homology with the avian retrovirus (1).) However, the above findings do not necessarily mean that the genome RNA of 17.6 is always reverse transcribable for its translocation. In fact, no change was found in the blotting hybridization pattern of 17.6 among DNAs extracted from GM2 cells cultured separately for more than two years (unpublished observation).

In this connection, it should perhaps be mentioned that a 14-nucleotide-long sequence, GAGGCTTCTCCTCA, is present 20-30 nucleotides upstream of both 17.6 RNA b and the major RNA of 297 (see box labelled by arrowed underline X in Fig.5). Because of its location, this sequence is likely related to the transcriptional regulation of these RNAs. Interestingly, the

same sequence is repeated in the 5'-upstream region in the case of 17.6 and is the longest in the repeated sequences found in the LTR of the authentic 17.6 (see arrowed underlines X in Fig.4).

ACKNOWLEDGEMENTS
The authors express their appreciation to Prof. M. Sekiguchi for his critical comments. This work was supported in part by grants from the Ministry of Education, Science and Culture, Japan to K.S.

*To whom correspondence should be addressed

REFERENCES

1. Kugimiya, W., Ikenaga, H. and Saigo, K. (1983) Proc. Natl. Acad. Sci. USA 80, 3193-3197.
2. Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K. and Yuki, S. (1984) Nature 312, 659-661.
3. Toh, H., Kikuno, R., Hayashida, H., Miyata, T.,Kugimiya, W., Inouye, S., Yuki, S. and Saigo, K. (1985) EMBO J. 4, 1267-1272.
4. Emori, Y., Shiba, T., Kanaya, S., Inouye, S., Yuki, S. and Saigo, K. (1985) Nature 315, 773-776.
5. Mount, S.M. and Rubin, G.M. (1985) Mol. Cell. Biol. 5, 1630-1638.
6. Clare, J. and Farabaugh, P. (1985) Proc. Natl. Acad. Sci. USA 82, 2829-2833.
7. Hauber, J., Hochstetter, P.-N. and Feldman, H. (1985) Nucl. Acids Res. 13, 2745-2758.
8. Warmington, J.R., Waring, R.B., Newlon, C.S., Indge, K. and Oliver, S.G. (1985) Nucl. Acids Res. 13, 6679-6693.
9. Inouye, S., Yuki, S. and Saigo, K. (1986). Eur. J. Biochem. 154, 417-425.
10. Yuki, S., Ishimaru, S., Inouye, S. and Saigo, K.(1986) Nucl. Acids Res. 14, 3017-3030.
11. Boeke, J.D., Garfinkel, D.J., Styles, C.A. and Fink, G.R. (1985) Cell 40, 491-500.
12. Maniatis, T., Fritch, E.F. and Sambrook, J.(1982) Molecular cloning, A laboratory manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
13. Vieira, J. and Messing, J. (1982) Gene 19, 259-268.
14. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65. 499-560.
15. Modolell, J., Bendor, W. and Meselson, M. (1983) Proc. Natl. Acad. Sci. USA 80, 1678-1682.
16. Saigo, K., Millstein, L. and Thomas, C.A., Jr. (1981) Cold Spring Harbor Symp. Quant. Biol. 45, 815-827.
17. Shiba, T. and Saigo, K. (1983) Nature 302, 119-124.
18. Deeley, R.G., Gordon, J.I., Burns, A.T.H., Mullinix, K.P., Bina-Stein, M. and Goldberger, R.F. (1977) J. Biol. Chem. 252, 8310-8319.

19. Levis, R., Dunsmuir, P. and Rubin, G.M. (1980) Cell 21, 581-588.
20. Will, B.M., Bayev, A.A. and Finnegan, D.J.(1981) J.Mol.Biol. 153, 897-915.
21. Shepherd, B.M. and Finnegan, D.J. (1984) J. Mol. Biol. 180, 21-40.
22. Freund, R. and Meselson, M. (1984) Proc. Natl. Acad. Sci. USA 81, 4462-4464.
23. Bayev, A.A., Lyubomirskaya, N.V., Dzhumagliev,E.B.,Ananiev, E.V., Amiantova, I.G. and Ilyin, Y.V. (1984) Nucl. Acids Res. 12, 3707-3723.
24. Inouye,S.,Yuki, S. and Saigo,K. (1984) Nature 310, 332-333.
25. Varmus, H.E. (1983) in Mobile Genetic Elements, Shapiro, J.A. (Ed.), pp411-503, Academic Press, New York.
26. Arkhipova, I.R., Mazo, A.M., Cherkasova, V.A., Gorelova, T.V., Schuppe, N.G. and Ilyin, Y. (1986) Cell 44, 555-563.
27. Flavell, A.J., Levis, R., Simon, M.A. and Rubin, G.M. (1981) Nucl. Acids Res. 9, 6279-6291.