MOLECULAR AND CELLULAR BIOLOGY, Sept. 1992, p. 3910–3918
0270-7306/92/093910-09$02.00/0

Vol. 12, No. 9

HeT-A, a Transposable Element Specifically Involved in "Healing" Broken Chromosome Ends in *Drosophila melanogaster*

HARALD BIESSMANN,[1] KATRIN VALGEIRSDOTTIR,[2] ARIEL LOFSKY,[2] CASEY CHIN,[1] BRET GINTHER,[1] ROBERT W. LEVIS,[3] AND MARY-LOU PARDUE[2]*

Developmental Biology Center, University of California, Irvine, California 92717[1]; Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139[2]; and Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98104[3]

Received 21 May 1992/Accepted 8 June 1992

Eight terminally deleted *Drosophila melanogaster* chromosomes have now been found to be "healed." In each case, the healed chromosome end had acquired sequence from the HeT DNA family, a complex family of repeated sequences found only in telomeric and pericentric heterochromatin. The sequences were apparently added by transposition events involving no sequence homology. We now report that the sequences transposed in healing these chromosomes identify a novel transposable element, HeT-A, which makes up a subset of the HeT DNA family. Addition of HeT-A elements to broken chromosome ends appears to be polar. The proximal junction between each element and the broken chromosome end is an oligo(A) tract beginning 54 nucleotides downstream from a conserved AATAAA sequence on the strand running 5′ to 3′ from the chromosome end. The distal (telomeric) ends of HeT-A elements are variably truncated; however, we have not yet been able to determine the extreme distal sequence of a complete element. Our analysis covers approximately 2,600 nucleotides of the HeT-A element, beginning with the oligo(A) tract at one end. Sequence homology is strong (>75% between all elements studied). Sequence may be conserved for DNA structure rather than for protein coding; even the most recently transposed HeT-A elements lack significant open reading frames in the region studied. Instead, the elements exhibit conserved short-range sequence repeats and periodic long-range variation in base composition. These conserved features suggest that HeT-A elements, although transposable elements, may have a structural role in telomere organization or maintenance.

The ends of eukaryotic chromosomes analyzed to date contain multiple repeats of very short, simple G-rich sequences (e.g., TTAGGG). These repeats are now considered the telomere sequences. There is evidence that the G-rich repeats can be added by telomerase, an enzyme that uses an RNA template to add copies of the telomere repeat to the chromosome end (5). Just internal to the telomere repeats, eukaryotic chromosomes have more complex sets of repeats (3, 4, 32) which are called telomere-associated repeats; however, their constant association with chromosome ends raises the possibility that telomere-associated sequences may be responsible for some of the functions that cytologists have suggested for the telomere. These possible functions include mediation of telomere-telomere and telomere–nuclear lamina interactions (14, 24). Thus, there is a possibility that telomere-associated sequences play a role in chromosome pairing at meiosis and/or in maintenance of the three-dimensional organization of chromosomes within the nucleus.

No simple telomerase-generated sequences have been found on *Drosophila* chromosomes. The failure to find simple repeats is not proof that they do not exist; however, it is possible that *Drosophila* species have lost this part of the ancestral telomere. On the other hand, *Drosophila* chromosomes do have larger, more complex telomere-associated sequences which may be evolutionarily related to the telomere-associated repeats of other organisms. It is possible that in *Drosophila* species, the telomere-associated sequences have assumed the entire function of the telomere.

D. melanogaster offers several advantages for the study of telomere-associated sequences. The many mutants and rearranged chromosomes make genetic analyses possible, while the polytene chromosomes offer cytological resolution well beyond that seen with other kinds of chromosomes. A family of telomere-associated sequences (the HeT DNA family) has been identified in *D. melanogaster*. This family was initially identified by a DNA fragment cloned in the lambda phage λT-A (31). The 9.4-kb cloned fragment hybridized exclusively with the last band at the telomere of each of the polytene chromosomes and with the pericentric beta heterochromatin. The morphology of the most terminal polytene bands cannot be unambiguously classified as either heterochromatin or euchromatin, but these bands show several properties of heterochromatin (27); therefore, the DNA family that these regions share with pericentric heterochromatin has been named HeT DNA (28).

HeT DNA is a family related by its heterochromatic chromosomal locations. However, the DNA in λT-A appears to be a mosaic of several kinds of sequences (28), and therefore the HeT DNA family may contain several sequence subfamilies. We now report sequence analyses that identify one subfamily as a set of transposable elements which we have named the HeT-A element. This element has transposed onto the ends of the eight known "healed" terminally deleted chromosomes in *D. melanogaster* (1a, 4, 26). The term "healed" is initially placed in quotation marks to emphasize that the addition of HeT-A may not completely reestablish the telomere. It seems likely that this is only a

* Corresponding author.

first step in the process. Acquisition of the HeT-A element gives the chromosome end homology to the telomere regions of other chromosomes. This could set the stage for recombination events that add other sequences.

In this article, we report sequences from two of the healed telomeres and compare them with sequences from the ends of unbroken chromosomes. Figure 1 presents a schematic diagram of the sequences analyzed. This analysis allows us to identify and partially define a novel transposable element, HeT-A. This analysis compares five HeT-A elements, including two that transposed onto the broken ends of X chromosomes (in the RT473 and RT394 stocks) less than 3 years before they were sequenced. These newly transposed elements are compared with elements from the ends of unbroken chromosomes: one, A4-4, found in cloned DNA from the subterminal region of chromosome 3R (16a) and two, A1 and A2, from the genomic clone λT-A, which appears to have come from the tip of an X chromosome (28).

The HeT-A element has a number of unusual features. First, transposition appears to be strictly limited to terminal regions of the chromosome. Second, the healing events show an invariant orientation of the element with respect to the chromosome end. Third, the element contains >2 kb of well-conserved sequence with an 80-bp repeat motif which could affect the structure of chromatin containing HeT-A. These special features of HeT-A elements suggest that the elements may represent a novel class of transposable elements, a class which may play a role in the telomere.

MATERIALS AND METHODS

Isolation of HeT-A elements from the healed terminal deficiency chromosomes RT394 and RT473 and from λT-A has been described recently (4, 28). The A4-4 element was isolated from genomic DNA of the A4-4 strain (16) cloned in λDASH (Stratagene). Details of the construction and screening of the A4-4 library are being prepared for publication (16a). The A4-4 stock has a P[*w*,*ry*] transposon near the telomere on chromosome 3R. ^{32}P-nick-translated subclones of HeT-A elements 394 and 473 were used as probes for HeT-A-homologous sequences on Southern blots of cloned DNA. Homology was detected in a 4.0- to 4.6-kb region located 5.0 to 9.6 kb distal to the P[*w*,*ry*] insertion site. A 5.7-kb *Bam*HI fragment containing the region of homology was subcloned from phage λD17. A 0.35-kb *Bgl*II fragment, the most proximal fragment cross-hybridizing to the HeT-A probes, was then subcloned for sequencing.

DNA fragments were subcloned into Bluescript (Stratagene) or M13mp18/mp19 phage and sequenced by the dideoxynucleotide termination technique. For double-stranded sequencing, T4 DNA polymerase (Pharmacia) was used. Klenow fragment of DNA polymerase I was used for single-stranded sequencing of M13mp18/mp19. All regions of sequences from RT394, RT473, and λT-A were sequenced on both strands. The short HeT-A element on A4-4 was sequenced on one strand only, and the sequence was extended 170 bp past the point at which the sequence diverged from the consensus HeT-A sequence.

Computerized searches for similarity between HeT-A elements and other known DNA sequences included translated searches with Genepro software (Riverside Scientific, Seattle, Wash.). The conceptual translation products of all six open reading frames (ORFs) of HeT-A DNA sequences were compared with all of the sequences in the GenBank and EMBL DNA data bases, each translated in all six possible ORFs. In most cases, a window of 30 amino acids was used.

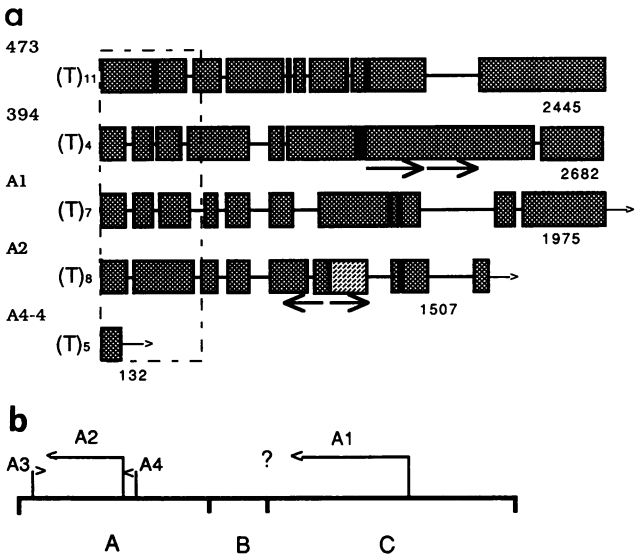

FIG. 1. (a) Diagram of HeT-A elements showing regions in which sequences have been aligned (blocks) and gaps have been inserted (thin horizontal lines) to yield optimal alignment. Elements 473 and 394 are cloned from the recently healed chromosome ends in the RT473 and RT394 stocks (4). A1 and A2 are the two large elements from λT-A (see panel b). A4-4 is cloned from telomere-associated sequences on chromosome 3R. All gaps of >3 bp are indicated. Because homology of the elements always begins with an oligo(A) segment at the 3′ end of one strand and extends 5′ for various distances, we have begun alignment with the first nucleotide 5′ of the oligo(A) segment. In order to present sequences 5′ to 3′, the strands depicted are the reverse complement of the oligo(A)-containing strand. The oligo(T) complement of the oligo(A) segment is indicated by $(T)_n$, where n is the number of residues of T for each element. The dashed box indicates the sequence in the sample alignment shown in Fig. 2. Numbers on the right are the actual numbers of nucleotides in each sequence. Arrows at the ends indicate that the sequence continues past the end of homology to the 394/473 sequence (indicated by boxes). The large tandem duplication in element 394 is underlined by boldface arrows under the 394 diagram. The large inverted repeat in element A2 is underlined by boldface arrows under the A2 diagram. The sequence of the repeat on the right diverges sharply from the consensus sequences of the other elements, and the box has been filled differently to indicate this and to suggest that this is a replacement rather than an insertion. (This sequence could have also been presented over gaps in the other sequences if a reciprocal gap had been inserted in the A2 sequence to accommodate the sequences of the other elements.) Several of the gaps in this alignment are identical for three sequences and presumably indicate an insertion in the fourth sequence. In several regions, gaps in one sequence include, wholly or partially, gaps in one or more of the other sequences. Such nested deletions may indicate regions particularly prone to rearrangement. (b) Diagram of the HeT-A elements found in λT-A. The 9.4 kb of cloned sequence is divided into segments A, B, and C by digestion with *Hin*dIII (28). The arrows indicate the sizes and locations of elements identified by sequence homology with the 394/473 sequence. Arrowheads indicate the end opposite the oligo(A) segment. Elements A2, A3, and A4 each diverge from the 394/473 sequence when they meet with another HeT-A element. The question mark at the end of element A1 indicates that its true end cannot be determined because the sequence that we have from the 394/473 sequence ends at the site where we have placed the arrowhead. It is possible that A1 continues to the end of A4.

Nucleotide sequence accession numbers. The GenBank accession numbers for the sequences are as follows: HeT-A2, M84197; HeT-A1, M84198; HeT-A4-4, M84199; HeT-A-RT473, M84200; and HeT-A-RT394, M84201.

RESULTS

HeT-A elements are defined by the sequences transposed to healed telomeres. The identification of HeT-A as a transposable element is based on demonstrated transposition onto the broken ends of chromosomes. At this point, the longest DNA sequences that we have obtained of this transposed DNA are from the new ends of the X chromosomes in the RT394 and RT473 stocks. As discussed below, we do not yet have sequences from the more distal part of any HeT-A element. However, clones from the healed RT telomeres have permitted us to define >2 kb of each of the elements 394 and 473. Their sequences have been compared with DNA sequences cloned from stocks without broken chromosomes. Regions with homology to the 394/473 consensus sequence have identified several preexisting HeT-A elements, three of which, A1, A2, and A4-4, are analyzed here. In these sequence comparisons, we have not detected significant differences between recently transposed elements and elements present in normal telomere-associated sequences.

One of the defining characteristics of HeT-A is an oligo(A) segment of variable length (3 to 28 residues in the elements studied thus far [4]) marking one end of each element. Immediately 3′ of this segment, we find no sequence homology between any of the elements studied (4). Immediately 5′ of the oligo(A) segment, all of the elements show very strong homology. On the healed telomeres, this oligo(A) tract forms the junction between the transposed DNA and the broken chromosome end on the strand running 5′ to 3′ from the new chromosome end. Elements 394 and 473 were cloned by using a restriction site within the yellow gene (the gene in which chromosome was broken in both stocks) and a second site a few hundred base pairs from the distal end of the HeT-A addition. Thus, we do not know the most distal sequences of either of the added elements. Mapping by Southern blot hybridization indicated that the distal *Hin*dIII site used for cloning was ca. 500 bp from the end of element 394 and 600 bp from the end of element 473 at the time of cloning (4). In those experiments, the DNA was mapped by using several restriction sites in the adjacent yellow gene, and the end of the chromosome was defined by the apparent clustering of different restriction sites. However, the distal ends of these additions change with time (see Discussion).

HeT-A elements in cloned DNA from telomere-associated sequences of unbroken chromosomes were identified by comparison with the consensus 394/473 sequence. The sequence in λT-A (Fig. 1b) has four such elements: two very short elements, A3 and A4, which have been presented elsewhere (28), and two longer ones, A1 and A2, which are discussed here. The A4-4 element was isolated from the right end of chromosome 3 by a chromosome walk (16a) starting from a P-element insert in transformant A4-4 (16). This HeT-A element is located 5 kb distal to the P-element insert. For all of these elements, homology to 394/473 begins abruptly with an oligo(A) segment and continues 5′ on that strand for various distances. The sequence of A1 can be matched to all of the 394/473 sequence and continues beyond it. Therefore, it is very possible that A1 contains a more complete HeT-A element than 394 or 473. However, our identification of HeT-A sequences depends on evidence of transposition, and at this point, we lack evidence that the additional sequence associated with A1 has ever transposed. The longest HeT-A sequence considered here is defined by the sequence homology between A1 and element 473 and contains >2 kb.

Elements A2 and A4-4 have both been sequenced past the points where their sequences diverge from the consensus sequence. The A2 sequence diverges after 1,507 nucleotides, only 3 nucleotides before the boundary of HeT-A element A3 (28). The A4-4 element extends only 132 nucleotides 5′ of its oligo(A) segment before diverging from the consensus sequence. However, since HeT-A probes hybridized to cloned DNA that is distal to the A4-4 element described here, the A4-4 element may be interrupted by an insertion of >170 bp (the amount of additional sequence obtained from the clone containing A4-4). Alternatively, the cloned DNA may contain a second element rather than an insertion, since in situ hybridization suggests that most wild-type chromosomes have multiple copies of HeT-A (4).

HeT-A elements have highly conserved sequences interrupted by small insertions, deletions, and duplications. The nucleotide sequences of the five elements analyzed here are generally colinear and show strong sequence conservation. No two are absolutely identical, but the major differences lie in small insertions or deletions. The deletions and insertions are indicated schematically in Fig. 1, and a sample of the aligned sequences is shown in Fig. 2. Because the shared starting point for the five elements is the oligo(A) segment on the 3′ end of one strand of DNA, we have chosen to present the alignments as representations of the opposite strand of DNA so that the common point of each element, the oligo(T) complement of the oligo(A) segment, is at the 5′ end. In Fig. 1, the oligo(T) complement is indicated just before the boxes representing the rest of the sequence and the number of T residues for each element is given.

The sequence analyses show no obvious difference between the recently transposed elements 394 and 473 and the other elements. Table 1 gives pairwise comparisons of sequence identities between all elements (% identity = number of identical bases/number of bases compared, omitting bases over gaps). Table 1 also lists the numbers of gaps introduced to yield the alignment. If allowance is made for differences in the lengths of the elements, the two measures of similarity, nucleotide identity and number of gaps, tend to be consistent. The two most different sequences are those of elements 394 and 473, with 77% nucleotide identity and 49 gaps. It should be noted that these are the two recent transpositions. The two most similar of the long elements are A1 and A2, which are adjacent on the λT-A clone (87% identity and 20 gaps). All sequences are generally more conserved near the oligo(A) segment, as can be seen by the comparisons with the very short A4-4 sequence. Within this highly conserved region, there is a polyadenylation signal, AATAAA, lying 54 nucleotides 5′ of the oligo(A) segment.

A more detailed examination of the sequence alignment shows that a large portion of the nucleotide substitutions are shared by more than one element. Surprisingly, these do not give a clear indication of the relationships between the elements, since substitutions are shared sometimes between one set of elements and sometimes between other sets.

Insertion and deletion events seem to be as frequent in the newly transposed elements as in the others (Fig. 1). There is also a tendency for different events to occur in the same region of more than one element, producing apparent nested deletions and insertions in the aligned sequences. Figure 2 shows two examples of such nested events. One set is seen

FIG. 2. Aligned sequences of five HeT-A elements, presented as the strand diagrammed in Fig. 1. All sequences begin with the first nucleotide 3′ of the last residue in the oligo(T) segment and proceed 3′ on that DNA strand. Only the first 600 nucleotides of the aligned sequences are shown. The dashed box in Fig. 1 indicates the portion of the sequence for which alignment is shown. Spaces indicate gaps introduced for alignment. Asterisks indicate nucleotides that are identical to the nucleotide at the top of that particular column. Where the sequence from element 473 is present, all sequences are compared with 473. Where element 473 is gapped, sequences are compared with whichever of the other sequences is at the top of that column. Overlining indicates the TTTATT complementary to the possible polyadenylation signal 54 nucleotides from the start of the element. The sequences were aligned by eye with the aid of the EyeBall Sequence Editor (7), by using pairwise dot matrix comparisons of the sequences as a guide. This alignment was chosen to provide the best fit for the entire set of elements.

at position 331 in the element 473 sequence, where a gap inserted in element 473 is also seen in element 394 and is overlapped by a gap in the A1 sequence. A second set is seen at position 542 in element 473, where the gap is overlapped at different points by gaps in both A1 and A2. Other examples can be seen in Fig. 1. These nested events may indicate regions where the sequence is particularly prone to rearrangement or sites where such rearrangement is most easily tolerated. On the other hand, those cases where all elements but one show identical gaps probably indicate an insertion in the ungapped element rather than multiple events.

Another notable feature of these sequences is the variable number of imprecise 80-bp tandem repeats. Each carries a highly conserved nonamer (CAGGTACAT) at one end, preceded by AT-rich runs of varying length. The most conserved of these repeats begin at about position 780 in element 473. Element 473 has five repeats, 394 has three, A2 has three, and A1 has two. When the sequences are analyzed by dot plot comparisons with reduced stringency (Fig. 3), a more extensive pattern of these repeats becomes apparent. In addition to the strongly conserved repeats, other copies with lower degrees of similarity can be detected in each of the elements. The regularity of the pattern of those residual repeats suggests that the HeT-A sequence may have arisen from a more perfect repeat and later undergone some decay. Although we do not know that these repeats are confined to HeT-A elements, in situ hybridization experiments do not detect them outside of the telomeric region of polytene chromosomes (28). Thus, although these repeats may not be specific for HeT-A, they are limited to heterochromatin.

TABLE 1. Pairwise sequence comparisons[a] of HeT-A elements

Sequences compared	No. of identical bases/ no. of bases compared	% Identity[b]	No. of gaps[c]
473 and 394	1,657/2,165	77	49
473 and A1	1,633/1,909	86	44
473 and A2	1,036/1,357	76	36
394 and A1	1,533/1,859	82	40
394 and A2	1,137/1,463	78	28
A1 and A2	1,095/1,260	87	20
A4-4 and 473	121/132	92	0
A4-4 and 394	123/132	93	0
A4-4 and A1	124/132	94	0
A4-4 and A2	124/132	94	0

[a] Sequences in the alignment were compared by counting only those bases that were aligned with a base in the other sequence being compared.

[b] Number of identical bases/number of bases compared, omitting bases over gaps.

[c] Number of gaps of one or more bases introduced to give the alignment.

Two of the elements have undergone other events. Element 394 is unique in containing a 359-bp duplication with 76% identity running from positions 1248 to 1965. A2 has an almost precise inverted repeat of the sequences between positions 859 and 1054.

Base composition oscillates sharply along the length of these elements (Fig. 4). The region around position 1000 in the longest elements has particularly notable regions of high AT composition. This is the area in which the most conserved 80-bp repeats are concentrated.

HeT-A elements have no significant ORFs within the region sequenced. The evidence that HeT-A elements are capable of transposition raises the question of whether these elements have ORFs, as do many transposable elements. All of the elements we have studied have the oligo(A) segment at the 3′ end of one strand of DNA. This aspect of the structure

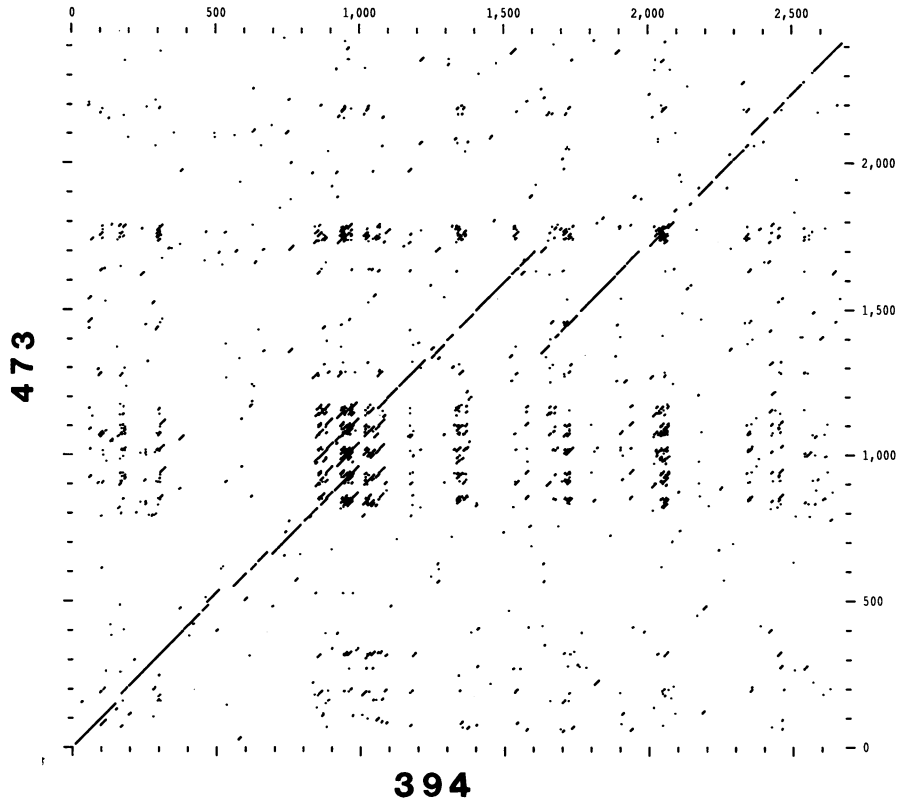

FIG. 3. Dot plot comparison of elements 473 and 394, showing the overall similarity and residual repeat structure that characterizes HeT-A elements. The comparison was made with a window of 21 bp and a stringency of 14. The best-conserved repeats center near position 1000 in each sequence. Element 473 has five repeats in this region, and element 394 has three repeats. In addition, there is evidence of more degenerated repeats at regular intervals outside the more conserved region.

resembles that of non-long terminal repeat retroposons such as mammalian LINES and the *Drosophila* elements jockey, I, F, and G (1, 6, 29). Those elements encode proteins thought to be involved in their transposition. In contrast, we have been unable to detect significant ORFs in the HeT-A elements sequenced to date. Studies of other transposable elements have shown that many elements can be defective and are no longer able to encode functional proteins (22, 25). Our search for ORFs has been carried out with this in mind. The search has centered on elements 473 and 394 since those elements transposed less than 3 years before they were sequenced. Therefore, they have recently been capable of transposition, even if they were not capable of providing the transposition proteins.

None of our searches have detected similarity between HeT-A and known transposable elements. A FASTA search of nucleotide data banks (21) yielded two sequences with significant identity to HeT-A. One was an unidentified *Drosophila* sequence that had been pulled out of a clone library accidentally by a heterologous probe. The sequence (DMO in reference 4) includes the oligo(A) segment of HeT-A and has 630 bp 5′ of the oligo(A) segment, all of which can be matched to the HeT-A sequence. The second sequence with HeT-A homology was from the *Drosophila* clone Dm665 (8), which is derived from an internal site on the heterochromatic Y chromosome. The Dm655 sequence has homology to the middle part of HeT-A, which has been joined to a part of the Stellate gene (18) to form a unit which is then tandemly repeated.

We have also failed to find similarities to any of the retroposon ORFs by searching with the fragments of amino acid sequence that can be translated from the HeT-A sequence. Several algorithms have been devised to look for residual ORFs in sequences which have been interrupted either by mutation or by errors in sequencing. One of these looks for regions rich in the organism's preferred codons (15). We have used the Genetics Computer Group (GCG) CodonPreference program with codon preference tables both for *Drosophila* sequences and for sequences from transposable elements of *Drosophila* spp. (tables provided by M. Ashburner, Cambridge, United Kingdom). A second program, GCG TestCode, is based on a statistic derived from the compositional bias of triplet position (13). Neither analysis gave evidence of large but degenerate ORFs in the 2.6 kb of HeT-A sequence that we have analyzed.

DISCUSSION

Transposition of HeT-A elements appears to be strictly limited to telomere-associated regions and newly broken ends. Previous studies of HeT DNA have strongly suggested that members of this family have some degree of mobility. HeT DNA is found on all telomeres, as well as in pericentric regions, in *D. melanogaster* (27, 28, 31). Multiple sites of hybridization are seen even under very stringent hybridization conditions. Thus, elements at different sites must have sequence similarity that could be maintained by reciprocal recombination, by gene conversion, or by transposition. It is

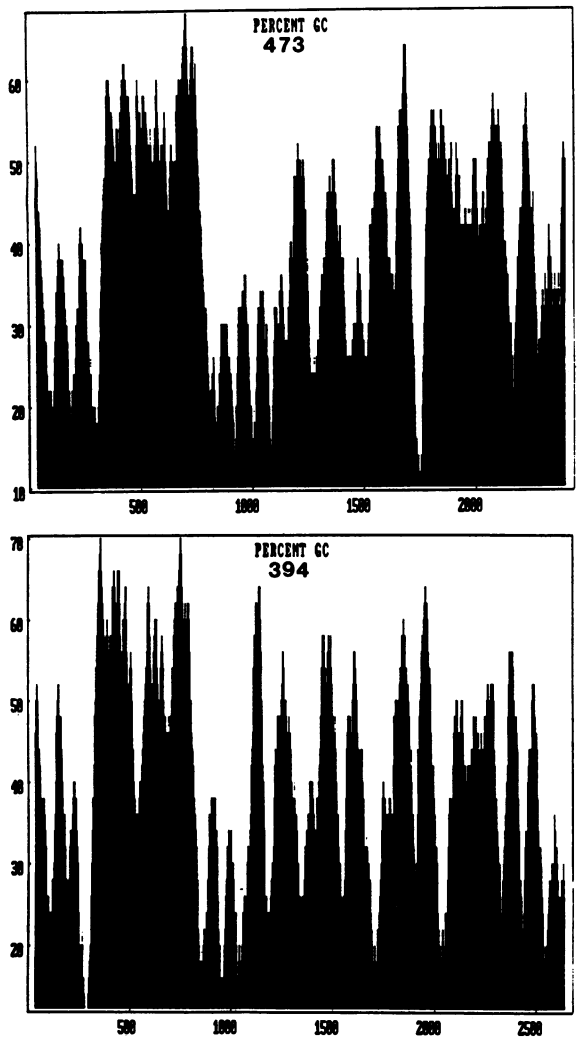

FIG. 4. Distribution of GC content in elements 473 and 394. Data were calculated with a 50-bp window by using the Genepro program. The orientation is the same as in Fig. 1 through 3, with the oligo(T) tail at position 1.

likely that all of these mechanisms play some role; however, at this point the only mechanism for which we have evidence is transposition.

All eight of the healed chromosome ends known in *Drosophila* spp. have acquired HeT DNA (1a, 4, 26). In the six cases in which the DNA sequence has been analyzed, the transposition has been accomplished by transposition of the element HeT-A. Although most known *Drosophila* transposable elements are found in euchromatic as well as heterochromatic regions, HeT-A elements are highly restricted. HeT DNA sequences have been used for in situ hybridization to >20 *D. melanogaster* stocks under conditions in which 40 bp of single copy sequence is easily identified in the euchromatic regions of polytene chromosomes (25a). In every stock, hybridization is seen only to telomeric and pericentric regions, and never within euchromatic regions. When the HeT-A sequence alone is used as a probe, hybridization only to telomeres is detected (28). Thus, in contrast to other known elements, HeT-A elements appear to be restricted to telomere regions. The only nontelomeric sequences with homology to HeT-A that have been detected are on the Y chromosome (8), which is not detectable in polytene nuclei. As discussed above, the tandemly repeated sequences on the Y chromosome (8) are quite different in structure from the HeT-A described here, suggesting that they did not move to the Y chromosome by the transposition mechanism that appears to be involved in healing chromosome ends.

Although most of the transposable elements in *Drosophila* spp. and other species are distributed throughout both euchromatic and heterochromatic regions, there are a few exceptions. The *Drosophila* ribosomal DNA type I and type II insertions are each limited primarily to a single site in ribosomal DNA, and G elements have been detected only in pericentric heterochromatin. It has been suggested that these positional restrictions are determined by sequence-specific endonucleases encoded by the elements (9, 30). We have no evidence that the positional restriction of HeT-A is determined by such an endonuclease. The 3′ flanking sequences for 11 HeT-A elements have been determined and each is different, giving no evidence for sequence-specific insertion (1a, 4, 28).

The limitation of HeT-A transposition to chromosome ends might be caused either by the mechanism of transposition or by an activity of the HeT-A element after integration. For example, transposition could be limited if the element recognizes only double-strand breaks as attachment sites. Under normal conditions, terminal double-strand breaks might remain available longer than internal breaks, thus allowing time for HeT-A attachment. Alternatively, HeT-A transposition to internal sites might prove to be a lethal event for the chromosome (12). For example, if addition of HeT-A leads to a cascade of recombination with telomere-associated sequences on other chromosomes, internal insertion of HeT-A could lead to loss of genes distal to the insertion. In most cases, this could cause aneuploidy that would be cell lethal.

Transposition of HeT-A sequences may not completely reestablish the telomere. Our initial study (4) indicated that the healed ends of elements 394 and 473 differed somewhat from established telomeres. By in situ hybridization, the new ends showed less homology to HeT-A than other ends and no homology to a second fragment of HeT DNA (fragment B) (28). In addition, Southern hybridization analysis indicated that after 2 years, the added DNA fragment was shorter on some chromosomes and longer on others. It is not known whether such length changes are found on chromosomes with established telomeres; however, terminally broken chromosomes do become progressively shorter if there has been no healing. The rate of such shortening is that expected if there is a failure to replace the 5′-most RNA primer in each round of replication (2, 17). Recently it has been found that the HeT-A sequences on healed ends are subject to similar decreases. These studies also detected transposition of a second HeT-A element onto the first, thus explaining how healed ends might grow longer. The second transposition also involves an oligo(A) junction rather than homologous recombination and thus links the newly added element to the earlier, possibly truncated, element by an oligo(A) end (1a, 18a).

Established telomeres may also have chains of partial HeT-A elements, linked by their oligo(A) ends. This suggestion is supported by the sequence of λT-A, which contains the oligo(A) ends of two long (A1 and A2) and two short (A3 and A4) HeT-A elements (4, 28). Elements A2, A3, and A4 are truncated at the end opposite the oligo(A) segment, apparently by the addition of the next HeT-A element (Fig.

1b). The sequence of A4 diverges from the consensus sequence only six nucleotides before the start of the oligo(A) segment of A2. The A2 sequence, in turn, diverges from the consensus sequence three nucleotides before the end of A3. The origins of the six nucleotides before A2 and the three nucleotides between A2 and A3 are unclear. In neither case are the nucleotides the ones that would be expected for a target site duplication.

As mentioned earlier, we cannot define the end of A1 because it appears to be longer than the consensus sequence used to identify HeT-A. A1 is separated from the oligo(A) segment of the shortest λT-A element by *Hin*dIII fragment B (28). As discussed below, we suspect that fragment B will eventually prove to be part of the complete HeT-A sequence. If so, element A1 may extend to the oligo(A) segment of HeT-A element A-4.

These observations suggest that the sequence of λT-A may have been formed originally by sequential addition of HeT-A elements onto a chromosome end. Three of the elements in the fragment have the same polarity, and each may extend to the beginning of the next. Only the short A3 element has the opposite polarity. A3 is also unusual in that it is flanked by 8-bp repeats (28). Most of the repeated bases lie within the element and overlap its 3-bp oligo(A) segment [i.e., GTTAAATT, where AAA in one of the repeats is the oligo(A) segment of this element]. The structure of the flanking repeats suggests that A3 may have undergone a rearrangement after addition to the chromosome end. The rearrangement could have inverted its sequence on the chromosome.

Do linked chains of partial HeT-A elements make up the functional *Drosophila* telomere? It should be noted that there are additional telomere-associated sequences on established telomeres. Their origin and function are not clear at this point. HeT-A elements added in the initial healing should have the potential to stimulate recombination reactions with other telomeres, thereby expanding the array of telomere-associated sequences. The many insertions and/or deletions that mark the otherwise strongly conserved HeT-A sequence suggest that these sequences may represent sites of exceptional rearrangement activity. If so, transposition of HeT-A may only be the first step in rebuilding a telomere.

Is HeT-A a retroposon? All of the junctions that have been observed between HeT-A and the DNA to which it has transposed have involved an oligo(A) segment of 3 to 28 bp. It is possible that this junction could indicate a novel transposition mechanism involving either the oligo(A) segment, its complement oligo(T), or the double-stranded DNA. However, the junction could also have been created by an RNA transposition intermediate, thus classifying HeT-A as a non-long terminal repeat retroposon (1, 6, 29). The possibility of the RNA intermediate is attractive because it involves mechanisms that have been shown in eukaryotic cells. In addition, any of several mechanisms for attaching the poly(A) tail of an RNA to a terminal break on a chromosome would provide an explanation for the polarity of the addition of HeT-A to the chromosome. The observation that HeT-A has a potential polyadenylation signal 54 nucleotides 5′ to the oligo(A) end lends some support to the suggestion that the oligo(A) segment is derived from the poly(A) tail of an RNA. However, 54 nucleotides is longer than the usual distance (9 to 19 bp) between the polyadenylation signal and the site of poly(A) addition in known *Drosophila* retroposons (9–11, 20, 23).

If HeT-A is a retroposon, it should have an RNA transcript; however, we have not yet found HeT-A RNA. We have used HeT-A to probe RNA blots both from cultured cells and from different developmental stages of *D. melanogaster*. Results were negative in both cases (27a). There are more powerful techniques for finding RNA transcripts, but we have not yet tried them. Our results do allow us to conclude that if the RNA exists, it is not an abundant transcript, despite the apparent number of HeT-A elements in the genome. (However, we do not know how many of these HeT-A elements are sufficiently intact to yield functional transcripts.) These negative results do not eliminate the possibility that there is an RNA transposition intermediate. There are several other explanations for our results. Perhaps the most interesting of these is the possibility that HeT-A transcription is induced by chromosome damage. McClintock (19) found that chromosome breakage mobilized transposable elements. HeT-A might be under similar control. We note that the original λT-A clone (31) was selected by nuclear RNA from heat-shocked cells. However, until the RNA is not only found but also shown to be a functional rather than an accidental transcript, we can draw no conclusion from this.

Do HeT-A elements contain ORFs? The DNA of most known transposable elements in *Drosophila* spp. tends to be largely employed in coding for proteins involved in transposition (1). We have found no such ORFs in any of the HeT-A elements that we have sequenced. It is possible that there are only a few HeT-A elements with functional ORFs and that those we have analyzed were rearranged to the point that ORFs were unrecognizable. Because of this possibility, we have used a variety of approaches to look for decayed or partial ORFs. The evidence strongly suggests that the elements studied do not encode proteins. The elements sequenced are very conserved, yet none of the long runs of sequence homology show the expected ORFs. Two of the elements had recently transposed, suggesting that successful transposition does not require coding potential. Finally, the sequences show a residual repeat pattern, which suggests that the sequence is conserved for a structural rather than a coding function.

The arguments above apply to the >2 kb of HeT-A sequence immediately upstream of the oligo(A) segment. However, we do not know the limit of HeT-A at the opposite end, leaving open the possibility that more complete elements will prove to have ORFs. This possibility is now very attractive because of recent work in two laboratories (1a, 8a). Both groups have found HeT DNA fragments with an ORF that encodes a protein with some similarity to the *gag* proteins found in many transposable elements. The fragment identified by Danilevskaya et al. (8a) contains a second ORF that lies upstream of the first and overlaps for several codons in a different reading frame. Such overlapping reading frames are typical of several types of transposable elements, including retroposons (6). The ORF sequences are very similar to the sequence in fragment B of λT-A (28), although many of the sequence differences result in stop codons in fragment B (7b). It is particularly interesting that fragment B lies just 5′ of the point on the A1 sequence where the comparison to 473 has stopped. Thus, the A1 element may actually extend further 5′ and include these (nonfunctional) ORFs. Although we have no proof that the ORF-containing region is actually part of the HeT-A element, we note that fragment B homology has never been seen, either with in situ hybridization or in cloned DNA fragments, unlinked from homology to the 394/473 HeT-A consensus sequence. In contrast, homology to the 394/473 consensus sequence has

been detected without fragment B in both situations (7a, 25a).

The potential ORFs are of interest because they may provide clues to the transposition of HeT-A, if they are proven to have function. This should not detract from interest in the >2 kb of conserved sequence with an underlying repeated sequence structure and the marked oscillations in base composition. These features suggest that a significant part of the HeT-A sequence may be involved in chromatin structure. If HeT-A elements do prove to contain coding regions external to the sequences reported here, it may be important that the potentially structural portions of this element lie on the end that is oriented toward the chromosome. Because they are at the internal end of the element, the structural regions will tend to become enriched on the telomere by the combined actions of the polar orientation of HeT-A addition, terminal truncation of attached elements, and subsequent addition of other HeT-A elements.

Conclusion. These sequence comparisons provide the beginnings of analysis of a novel transposable element. The apparently obligate association of HeT-A with telomeres and the suggestions of an underlying structural pattern in the DNA sequence raise the possibility that the element may play a role in chromosome structure.

ACKNOWLEDGMENTS

This work was supported by grants from the National Science Foundation (DMB 89-03731 to H.B. and DMB 90-05047 to R.W.L.), the National Institutes of Health (GM38259 to R.W.L.), and the American Cancer Society (CD-487H to M.-L.P.).

We acknowledge the excellent technical assistance of Karen Jakes, Kathy Houtchens, and Robin Ganesan. We thank the MIT Whitaker College Computing Facility for computer use.

REFERENCES

1. **Berg, D. E., and M. M. Howe (ed).** 1989. Mobile DNA. American Society for Microbiology, Washington, D.C.
1a. **Biessmann, H., L. E. Champion, M. O'Hair, K. Ikenaga, B. Kasravi, and J. M. Mason.** Submitted for publication.
2. **Biessmann, H., and J. M. Mason.** 1988. Progressive loss of DNA sequences from terminal chromosome deficiencies in Drosophila melanogaster. EMBO J. **7:**1081–1086.
3. **Biessmann, H., and J. M. Mason.** Genetics and molecular biology of telomeres. Adv. Genet., in press.
4. **Biessmann, H., J. M. Mason, K. Ferry, M. d'Hulst, K. Valgeirsdottir, K. L. Traverse, and M. L. Pardue.** 1990. Addition of telomere-associated HeT DNA sequences "heals" broken chromosome ends in Drosophila. Cell **61:**663–673.
5. **Blackburn, E. H.** 1990. Telomeres and their synthesis. Science **249:**489–490.
6. **Boeke, J. D., and V. G. Corces.** 1989. Transcription and reverse transcription of retrotransposons. Annu. Rev. Microbiol. **43:** 403–434.
7. **Cabot, E. L., and A. T. Beckenbach.** 1989. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Comput. Appl. Biosci. **5:**233–234.
7a. **Danilevskaya, O.** Unpublished data.
7b. **Danilevskaya, O., A. Lofsky, and M.-L. Pardue.** Unpublished data.
8. **Danilevskaya, O. N., E. V. Kurenova, M. N. Pavlova, D. V. Bebehov, A. J. Link, A. Koga, A. Vellek, and D. L. Hartl.** 1991. He-T family DNA sequences in the Y chromosome of Drosophila melanogaster share homology with the X-linked stellate genes. Chromosoma **100:**118–124.
8a. **Danilevskaya, O. N., D. A. Petrov, M. N. Pavlova, A. Koga, E. V. Kurenova, and D. L. Hartl.** Submitted for publication.
9. **Di Nocera, P. P.** 1988. Close relationship between non-viral retroposons in Drosophila melanogaster. Nucleic Acids Res. **16:**4041–4052.
10. **Di Nocera, P. P., and G. Casari.** 1987. Related polypeptides are encoded by Drosophila F elements, I factors, and mammalian L1 sequences. Proc. Natl. Acad. Sci. USA **84:**5843–5847.
11. **Di Nocera, P. P., F. Graziani, and G. Lavorgna.** 1986. Genomic and structural organization of Drosophila melanogaster G elements. Nucleic Acids Res. **14:**675–691.
12. **Farr, C., J. Fantes, P. Goodfellow, and H. Cooke.** 1991. Functional reintroduction of human telomeres into mammalian cells. Proc. Natl. Acad. Sci. USA **88:**7006–7010.
13. **Fickett, J. W.** 1982. Recognition of protein coding regions in DNA sequences. Nucleic Acids Res. **10:**5303–5318.
14. **Gillies, C. B.** 1975. Synaptonemal complex and chromosome structure. Annu. Rev. Genet. **9:**91–109.
15. **Gribskov, M., J. Devereux, and R. R. Burgess.** 1984. The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Res. **12:**539–549.
16. **Hazelrigg, R., R. Levis, and G. M. Rubin.** 1984. Transformation of white locus DNA in Drosophila: dosage compensation, zeste interaction, and position effects. Cell **36:**469–481.
16a. **Levis, R., R. Ganeson, K. Houtchens, and F. Sheen.** Unpublished data.
17. **Levis, R. W.** 1989. Viable deletions of a telomere from a Drosophila chromosome. Cell **58:**791–801.
18. **Livak, K.** 1984. Organization and mapping of a sequence on the Drosophila melanogaster X and Y chromosomes that is transcribed during spermatogenesis. Genetics **107:**611–634.
18a. **Mason, J.** Unpublished data.
19. **McClintock, B.** 1978. Mechanisms that rapidly reorganize the genome. Stadler Genet. Symp. **10:**25–47.
20. **O'Hare, K., M. R. K. Alley, T. E. Cullingford, A. Driver, and M. J. Sanderson.** 1991. DNA sequence of the Doc retroposon in the white-one mutant of Drosophila melanogaster and of secondary insertions in the phenotypically altered derivatives white-honey and white eosin. Mol. Gen. Genet. **225:**17–24.
21. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
22. **Potter, S. S.** 1984. Rearranged sequences of a human KpnI element. Proc. Natl. Acad. Sci. USA **81:**1012.
23. **Priimagi, A. F., L. J. Mizrokhi, and Y. V. Ilyin.** 1988. The Drosophila mobile element jockey belongs to LINES and contains coding sequences to some retroviral proteins. Gene **70:** 253–262.
24. **Rouyer, F., A. de la Chapelle, M. Andersson, and J. Weissenbach.** 1990. An interspersed repeated sequence specific for human subtelomeric regions. EMBO J. **9:**505–514.
25. **Scott, A. F., B. J. Schmeckpeper, M. Abdelrazik, C. T. Comey, B. O'Hara, J. P. Rossiter, T. Cooley, P. Heath, K. D. Smith, and L. Margolet.** 1987. Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics **1:**113–125.
25a. **Traverse, K., and M.-L. Pardue.** Unpublished data.
26. **Traverse, K. L., and M. L. Pardue.** 1988. A spontaneously opened ring chromosome of Drosophila melanogaster has acquired HeT DNA sequences at both new telomeres. Proc. Natl. Acad. Sci. USA **85:**8116–8120.
27. **Traverse, K. L., and M. L. Pardue.** 1989. Studies of He-T DNA sequences in the pericentric regions of Drosophila chromosomes. Chromosoma **97:**261–271.
27a. **Valgeirsdottir, K., and M.-L. Pardue.** Unpublished data.
28. **Valgeirsdottir, K., K. L. Traverse, and M. L. Pardue.** 1990. HeT DNA: a family of mosaic repeated sequences specific for heterochromatin in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA **87:**7998–8002.
29. **Weiner, A. M., P. L. Deininger, and A. Efstratiadis.** 1986.

Nonviral retrotransposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. **55**:631–661.

30. **Xiong, Y., and T. H. Eickbush.** 1988. Functional expression of a sequence-specific endonuclease encoded by the retrotransposon R2Bm. Cell **55**:235–246.
31. **Young, B. S., A. Pession, K. L. Traverse, C. French, and M. L. Pardue.** 1983. Telomere regions of Drosophila share complex DNA sequences with pericentric heterochromatin. Cell **34**:85–94.
32. **Zakian, V. A.** 1989. Structure and function of telomeres. Annu. Rev. Genet. **23**:579–604.