The Gag protein of the Drosophila telomeric retrotransposon TAHRE collaborates with HeT-A and TART Gags for nuclear localization

A. M. Fuller*¹, E. G. Cook*, K. J. Kelley, and M.-L. Pardue

Department of Biology
Massachusetts Institute of Technology
Cambridge, MA 02139

*These authors contributed equally.

1. Present address: Department of Science, Urban Assembly Academy of Civic Engagement, 650 Hollywood Avenue, Bronx, NY 10465
Running head: The Drosophila TAHRE Gag protein

Key words: telomeres, nuclear localization, retrotransposon, MHR and zinc knuckle motifs, pre_C2HC motif.

Corresponding Author:

Mary-Lou Pardue
Department of Biology, 68-670
Massachusetts Institute of Technology
77 Massachusetts Ave.
Cambridge, MA 02139
Phone: (617) 253-6741
FAX: (617) 253-8699
email: mlpardue@mit.edu
ABSTRACT

TAHRE, the least abundant of the three retrotransposons forming telomeres in Drosophila melanogaster, has high sequence similarity to the gag gene and untranslated regions of HeT-A, the most abundant telomere-specific element. Despite TAHRE’s apparent evolutionary relationship to HeT-A, we find TAHRE Gag cannot locate to telomere-associated “Het dots” unless collaborating with HeT-A Gag. TAHRE Gag is carried into nuclei by HeT-A or TART Gag, but both TART and TAHRE Gags need HeT-A Gag to localize to Het dots. When coexpressed with the appropriate fragment of HeT-A and/or TART Gags, TAHRE Gag multimerizes with either protein. HeT-A and TART Gags form homo- and hetero-multimers using a region containing MHR and zinc knuckle (CCHC) motifs, separated by a pre_C2HC motif (motifs common to other retroelements). This region’s sequence is strongly conserved among the three telomeric Gags, with precise spacing of conserved residues. Non-telomeric Gags neither interact with the telomeric Gags nor have this conserved spacing. TAHRE Gag is much less able to enter the nucleus by itself than HeT-A or TART Gags. The overall telomeric localization efficiency for each of the three telomeric Gag proteins correlates with the relative abundance of that element in telomere arrays, suggesting an explanation for the relative rarity of TAHRE elements in telomere arrays and supporting the hypothesis that Gag targeting to telomeres is important for the telomere-specific transposition of these elements.
INTRODUCTION

Drosophila telomeres are maintained by a remarkable variant of the telomerase mechanism that maintains telomeres in almost all organisms (MELNIKOVA and GEORGIEV 2005; PARDOE and DEBARYSHE 2003). As in other organisms, Drosophila telomeres are elongated by tandem repeats that are reverse-transcribed onto the ends of the chromosomes. What makes Drosophila telomeres unusual is the RNA template that is reverse-transcribed to produce these repeats: Drosophila telomere repeats are copied from full length retrotransposons (HeT-A, TART, and TAHRE), rather than from a short segment of the RNA molecule that makes up part of the telomerase holoenzyme (Figure 1).

Although clearly related to other retrotransposons in the D. melanogaster genome, the three retrotransposons that make up the telomeres have several characteristics that set them apart from the more typical retrotransposable elements. One of these characteristics is their localization to telomere arrays. The euchromatic regions of the D. melanogaster genome have been completely sequenced (CELNIKER et al. 2002). Analysis of these gene-rich regions reveals no sequence from any of the three telomeric elements (GEORGE et al. 2006), although these euchromatic regions are littered with other retrotransposons (KAMINKER et al. 2002). Conversely, the long arrays of telomeric retrotransposons do not contain their non-telomeric relatives. Thus, the telomeric and non-telomeric elements have distinctly different genomic distributions, except for small “transition zones” at the proximal ends of telomere arrays where fragments of both kinds of elements are mingled (GEORGE et al. 2006).
The telomere-specific transposition of HeT-A and TART appears to depend on the intranuclear targeting of the Gag proteins encoded by each element. These Gags share amino acid sequence motifs with retroviral Gags, proteins known to be important in intracellular transport of viral RNA. The sequence similarities with retroviral Gags suggest that the telomeric Gags are important in intracellular transport of the retrotransposon RNA. This suggestion is supported by studies of the intracellular localization of HeT-A and TART Gag proteins. Transient expression of tagged Gag proteins in D. melanogaster cells showed that Gags of both HeT-A and TART localize to nuclei very efficiently. Gags of non-telomeric retrotransposons were also tested in these experiments and found predominantly, if not entirely, in the cytoplasm (RASHKOVA et al. 2002b). Preventing Gags of non-telomeric retrotransposons from entering the nucleus may be one of the mechanisms cells use to protect their genomes from parasitic invaders. In contrast, the telomeric retrotransposons have an essential role in the nucleus and the cell benefits from facilitating nuclear localization of their Gags.

After moving from the cytoplasm into the nucleus, HeT-A Gags form aggregates (Het dots) associated with telomeres in interphase nuclei. HeT-A and TART are intermingled in D. melanogaster telomere arrays so it was surprising that TART Gags formed loose intranuclear clusters with no obvious telomere-associations. However, cotransfection experiments showed that when the two Gags are expressed in the same cells, HeT-A Gag dominates the localization and both proteins are moved into telomere-associated Het dots (RASHKOVA et al. 2002a). Presumably this localization is necessary for transposition to telomeres.
The collaborative localization of the two Gags suggests an explanation for two puzzling observations. The first observation is that all *D. melanogaster* stocks and cell lines have both *HeT-A* and *TART* in their telomeres, suggesting that both elements are needed by the cell. However, the two elements seem to be distributed randomly in telomere arrays, giving no indication that either has a special role. The second observation is that *HeT-A* elements do not encode reverse transcriptase, while *TART* does. Most, if not all, other retrotransposons encode this enzyme. Having the enzyme sequence associated with the element’s RNA would be expected to allow more efficient transposition, as has been shown for human *Lines-1* elements (Wei et al. 2001).

Nevertheless *HeT-A* transposes efficiently and is significantly more abundant than *TART* in telomeres of all *D. melanogaster* stocks and cell lines studied (George et al. 2006). The finding that *HeT-A* Gag positions *TART* for transposition to telomeres suggested that *TART* might provide the reverse transcriptase for both elements, thereby explaining the need for both elements in the genome. We suggest that *HeT-A* predominates because it has strong telomere targeting.

After these localization studies were finished, a third *D. melanogaster* telomeric retrotransposon, *TAHRE*, was reported (Abad et al. 2004). *TAHRE* has both a *HeT-A*-related Gag protein and a reverse transcriptase gene closely related to that of *TART* and thus presumably with the same activity. *TAHRE*’s sequence predicted that it should combine the localization activity of *HeT-A* with the enzyme activity of *TART* and transpose more efficiently than either of the other elements, yet *TAHRE* is actually much less abundant than either *HeT-A* or *TART*. Only one full-length copy of this element has been reported and only one full-length copy of its Gag gene is found in the
D. melanogaster database. In this study we have examined the intracellular localization of TAHRE Gag to see whether the sequence similarity to HeT-A Gag yields a protein with the remarkable telomere targeting of HeT-A Gag and to shed light on TAHRE’s relative rarity in telomeric arrays.

MATERIALS AND METHODS

Cultured Drosophila Cells: Drosophila Schneider line 3 cells (S3) were maintained at room temperature in Schneider’s medium (GIBCO Invitrogen) with heat-inactivated 10% Fetal Bovine Serum (Hyclone) and 10 units/ml penicillin, 10 µg/ml streptomycin solution (Mediatech).

Recombinant DNA and Plasmid Construction: TAHRE gag coding sequence was amplified by PCR from BACR40C07, a kind gift from Alfredo Villasante, Centro de Biología Molecular, Severo Ochoa. PCR primers (forward: 5'-GAAGAATTCCCACCATGTCCACGTCCGACCAC--3’ and reverse: 5-GATGGATCCTATCCGGAGGTCATGAGGTGTGCT-3’) amplified the entire coding region, preceded by a Drosophila Kozak sequence and minus the stop codon,. This was flanked by 5’ EcoR1 and 3’ Kpn1 sites so that the product could be inserted in frame with the gene for fluorescent protein (GFP, YFP, or CFP) in expression vectors. The PCR product was TA cloned into a Strataclone PCR cloning vector (Stratagene) and verified by sequencing. The gag sequence was cut out with EcoR1 and Kpn1 and inserted into pPL17. The resulting construct expressed Gag tagged with EGFP at the C-terminus, driven by the armadillo promoter. EcoR1-Kpn1 fragments were also
inserted in the pPL17 variants, pSR24 and pSR25, to express Gag with C-terminal tags of YFP and CFP, respectively.

**Transfection:** Two ml of S3 cells (5 x 10^6 cells/ml) were seeded onto six-well tissue culture dishes and transfected with 1 µg of DNA plus 25 µl of Effectene Transfection Reagent (Qiagen), used as the manufacturer directs. For co-transfections, 0.5 µg of each DNA was used. After 72 hr at rt, transfected cells were resuspended in their medium. 40 µl were used for cytological preparations while the rest was washed in PBS, pelleted, and frozen for immunoblotting.

**Cytology:** 20 ml of transfected cells were spotted onto slides that had been dipped in 0.005% poly-L-lysine and dried. Cells were allowed to settle for 20 min, fixed with 3.7% formaldehyde/ PBS for 30 min., washed in PBS for 5 min, 0.2% triton/PBS 10 min, and PBS 2 min. Preparations were blocked with 1% BSA/PBS for 1 hr and incubated with anti-lamin (Dmo ADL67.10-c, DHSB, Univ. of Iowa) 1:200 in 1% BSA/PBS at 4º C. overnight. Slides were washed with PBS 4 x 5 min. and incubated with Cy3-labeled anti-mouse IgG (Jackson Labs) 1:200 in BSA/PBS at rt for 1 hr, washed 4 x 5 min with PBS and mounted in 70% glycerol/PBS. Slides were examined with a Nikon ECLIPSE E 600 microscope, photographed with a Spot RT slider camera, and false-colored with Photoshop.

**Immunoblots:** Cells from 2 ml cultures were washed in PBS and resuspended in 100 µl of PBS. 100 µl of 2X Laemmli buffer was added and samples were boiled for 5 min. 2 µl β-mercaptoethanol was added to each 20 µl of sample. Proteins were resolved by SDS-polyacrylamide gel electrophoresis in 8% gels and transferred to nitrocellulose membranes. Fusion proteins were visualized by immunoblotting with anti-
GFP antiserum from guinea pig (Rashkova et al. 2002b) and secondary anti-guinea pig antibody conjugated to alkaline phosphatase (Jackson ImmunoLaboratories, Inc.)

**Sequences used for analysis:** TAHRE, AJ542581; TART A, AY561850; TART B, U14101; TART C, AY600955; canonical HeT-A, U06920 (nt 1015-7097); HeT-A elements from 4R telomere, AC010841 (4-5=nt 27992-33790, 4-4= nt 33948-39772, 4-3= nt 40382-46298, 4-2=53189-59033, 4-1= nt 60023-65592); HeT-A element from XL telomere, CP000372 (nt 12874-18879); jockey, M22874; Doc, X17551; I Factor, M14954.

**RESULTS**

The **TAHRE gag** gene and untranslated regions have strong nucleotide sequence similarity to the corresponding regions of **HeT-A**: Their nucleotide sequences suggest that **TAHRE** and **HeT-A** are derived from the same ancestor (Abad et al. 2004). These elements differ in that **TAHRE** has a reverse transcriptase coding sequence (**pol** gene) plus 200 bp of novel 3′UTR sequence (see Figure 1). The ancestral sequence is unknown: the **pol** gene may have been either deleted to produce **HeT-A** or inserted to produce **TAHRE**. The **pol** gene has more sequence similarity to **TART pol** than to **pol** genes in the non-telomeric retrotransposons that we have compared (data not shown).

Even within the same telomere, **HeT-A** elements differ in sequence, as shown by the six full-length elements in the partially assembled telomeres on chromosomes X and 4 (George et al. 2006). Surprisingly, all these Gag proteins show less conservation of the amino acid sequence than of the nucleotide sequence. Pairwise comparisons of the
Gag proteins from these HeT-A elements show a broad distribution of between 75.5% and 99.7% amino acid identity. Interestingly, pairwise comparisons between TAHRE and these HeT-A Gags show that TAHRE has about the same amount of amino acid identity (69.5%-72.3%) with each of the divergent HeT-A proteins.

There are three subfamilies of TART elements in the D. melanogaster genome. In pairwise comparisons of amino acids, Gags of these subfamilies differ by 82.5%-94.7%. These TART Gags have less amino acid identity with TAHRE Gag (44.8%-45.9%) than do the HeT-A proteins. HeT-A and TART Gags have less amino acid identity with each other (21.1% -23.9%) than either has with TAHRE.

Amino acid identities between TAHRE and HeT-A Gags are most concentrated in the region containing the MHR (major homology region) and zinc knuckle motifs (see Figure 2 and Supporting Material Figures S1, S2, S3). This same region also has the highest concentration of identities of HeT-A and TAHRE Gags to TART Gags.

To investigate the significance of these sequence similarities, we subcloned the TAHRE Gag coding region (Figure 2), minus its stop codon, from BACR40C07. This coding region was fused in frame to sequence for a fluorescent protein tag in the expression vector, pPL17, that has been used to study the HeT-A and TART Gags (RASHKOVA et al. 2002a). Three constructs were made, adding GFP, CFP, or YFP (Green, Cyan, or Yellow Fluorescent Protein) to the C-terminus of TAHRE Gag. When transfected into cultured Drosophila cells, these constructs expressed full-length fusion protein, as confirmed by immunoblot analysis (Figure 3).

Unlike other telomeric Gags, TAHRE Gag is found mostly in the cytoplasm: Because of the strong sequence similarity to HeT-A Gag, we assumed that TAHRE
Gag, like *HeT-A* Gag, would enter the nucleus and localize to telomeres (*Rashkova et al. 2002b*). Surprisingly, *TAHRE* Gag formed clusters but almost all of these clusters were in the cytoplasm, although they tended to concentrate around the nucleus (Fig 4A, B). Lamin staining to define the nuclear envelope showed many Gag clusters superimposed on the lamin stain, suggesting that they were stuck within the nuclear lamina. A few *TAHRE* Gag clusters were seen inside the lamin staining; however these remained near the periphery of the nucleus, rather than moving to more interior regions. The *TAHRE* Gag distribution differed from that of *HeT-A* Gag in a second way. In cells with nuclear Het dots, *HeT-A* Gag is sometimes also seen in the cytoplasm. This cytoplasmic protein is always concentrated into a single characteristic body, which we refer to as a Het body. We have suggested that Het bodies may be the result of overload of the transport system in these cells which are overexpressing *HeT-A* Gag (*Rashkova et al. 2002b*). Cytoplasmic *TAHRE* Gag has a very different distribution, a second indication that it is not transported as is *HeT-A* Gag.

The cytoplasmic distribution of *TAHRE* Gag differs from that of the Gags of non-telomeric retrotransposons that have been studied in this transient transfection assay. These non-telomeric Gags do not enter the nucleus but they are distributed more evenly through the cytoplasm and do not concentrate near the nucleus like *TAHRE* Gag (*Rashkova et al. 2002b*). This difference suggests that *TAHRE* is actively localized toward the nucleus and has some ability to enter; however it is either unable to move independently within the nucleus or it is impeded by interactions with the nuclear periphery.
**HeT-A and TART Gags can move TAHRE Gag into the nucleus:** As noted above, the amino acid sequence identities between TAHRE Gag and HeT-A Gag are especially strong in the region containing the MHR-zinc knuckle motifs (Fig 2). This region of HeT-A and TART Gags has been shown to be responsible for the protein-protein interactions that allow HeT-A Gag to move TART Gag to telomere-associated Het dots (RASHKOVA et al. 2003). The sequence similarity suggested that TAHRE Gag might interact with the Gags of the other two telomeric retrotransposons to enable it to localize to the nucleus. We tested this possibility by coexpression experiments similar to those used to characterize the interaction of HeT-A and TART Gags.

HeT-A Gag has been shown to enter the nucleus and localize to telomere-associated Het dots. It also forms a single, distinctive cytoplasmic Het body in some cells that have several Het dots (RASHKOVA et al. 2002a). We saw similar localizations when we expressed HeT-A Gag (Figure 4C). When we coexpressed TAHRE Gag with HeT-A Gag, the two Gags colocalized to nuclear Het dots (Figure 4E) and, in some cells, a cytoplasmic Het body, showing that HeT-A Gag could direct TAHRE Gag to telomeres. Interestingly, HeT-A Gag had also cleared TAHRE Gag from its diffuse distribution in the cytoplasm: when TAHRE Gag was seen in the cytoplasm, it was always gathered into the single Het body. Although HeT-A Gag has a strong influence on TAHRE Gag, TAHRE Gag also affects the interaction. In the cotransfections, nuclear Het dots are less round and regular and most tend to be located near the nuclear membrane (Figure 4E), while cells expressing only HeT-A Gag have round Het dots both at the membrane and in interior positions (Figure 4C).
The similarity of the MHR-zinc knuckle regions (Supporting Figures S1 to S3) suggested that TART Gag might also be able to localize TAHRE Gag to the nucleus. When expressed alone, TART Gag enters the nucleus and forms many small clusters. None of the Gag remains in the cytoplasm but the nuclear clusters do not show preferential association with telomeres (RASHKOVA et al. 2002a) (see also Figure 4D). Coexpression of TART and TAHRE Gags showed that TART Gag was able to move TAHRE Gag into the nucleus. In the nucleus the two Gags colocalized but their distribution is much more diffuse than the clusters that TART Gag makes when transfected alone (Figure 4F). TAHRE Gag changes the intranuclear organization of TART Gag more dramatically than it changes the intranuclear localization of HeT-A Gag. This suggests that HeT-A Gag, which dominates localization of the other two Gags, has stronger targeting within the nucleus than either of the other Gags.

**TAHRE Gag association with HeT-A and TART Gags involves the region containing the MHR-zinc knuckle motifs:** Rashkova et al. (2003) have shown that the regions of TART and HeT-A Gags containing the MHR-zinc knuckle motifs facilitate both homologous and heterologous associations to form clusters and co-localize. Deletion derivatives of the two Gags that retain the MHR-zinc knuckle internal association region can be carried into the nucleus by full length Gags of either element, even though the deletion derivatives lack the N-terminal residues necessary to independently enter the nucleus. The deletion derivatives used in those experiments were HeT-A Gag, amino acids 482-689, and TART Gag, amino acids 533-970 (see Figure 2). Neither of these deletion derivatives was able to enter the nucleus when
expressed individually (Rashkova et al. 2003). Instead they were distributed through the cytoplasm and formed aggregates of various sizes (Figures 5A and B).

To see whether TAHRE Gag's association with HeT-A and TART Gags involved the MHR-zinc knuckle regions, we co-expressed TAHRE Gag with the deletion derivatives of HeT-A Gag and TART Gag used by Rashkova et al. When co-expressed with either HeT-A Gag: 482-689 or TART Gag 533-950, TAHRE Gag colocalized tightly with the deletion derivative (Figures 5C and D). This colocalization shows that TART and HeT-A Gags interact with TAHRE Gag, whose functional regions have not yet been characterized, using the same regions in which they interact with each other, the regions containing the MHR and zinc knuckles.

None of the proteins in these colocalization experiments were moved into the nucleus. However each protein had some modulating effect on its partner's localization. When expressed with the HeT-A derivative, TAHRE Gag was pulled away from the nuclear membrane into the HeT-A cytoplasmic aggregates, producing larger aggregates, some of which resembled Het bodies (Figure 5C). When TAHRE Gag was coexpressed with the TART Gag derivative, the proteins colocalized to form one or more large Het body-like aggregates in the cytoplasm (Figure 5D). These bodies sometimes became large enough to distort the shape of the nucleus but were not concentrated around the nuclear membrane or seen inside the nucleus. We observed a range of sizes of cytoplasmic clusters in both single transfections and cotransfections of TART Gag 533-970; however the clusters of individually transfected TART Gag 533-970 were much more diffuse than the smooth, dense clusters seen in the cotransfectants. These results demonstrate TAHRE Gag’s inability to localize other proteins to the
nuclear membrane but show that this Gag is able to affect the conformation of the aggregates it forms with other proteins.

**DISCUSSION**

*HeT-A, TART and TAHRE make up an unusual trio in terms of the intracellular localization of their Gag proteins:* Each Gag protein has a different pattern of localization when expressed by itself, although the three elements appear to have similar roles in forming telomere arrays. *HeT-A* Gag localizes to Het dots associated with telomeres in interphase nuclei. *TART* Gag moves into nuclei but does not show preferential association with telomeres. *TAHRE* Gag remains predominantly in the cytoplasm with a tendency to concentrate around the nucleus and to colocalize with nuclear lamin. Neither *TAHRE* nor *TART* Gags localize to telomeres independently. Both require interaction with *HeT-A* Gag to reach this localization.

**Association between the telomeric Gags is directed by a well-conserved segment of these highly variable proteins:** Studies of deletion derivatives of Gag proteins (Rashkova *et al.* 2003) show that association between *HeT-A* and *TART* Gags depends on a highly-conserved region of each protein that contains the MHR (major homology region) and the zinc knuckle (CCHC box) motifs (Figure 2A). The experiments reported here show that this same region directs associations of these two telomeric Gags with *TAHRE*

The MHR and zinc knuckle amino acid motifs are hallmarks of retroviral Gag proteins. The MHR (QGX\textsubscript{2}EX\textsubscript{7}R) is so named because it is the only region of significant homology among different groups of retroviruses (Cra\text{\textsc{v}}en *et al.* 1995). The zinc
knuckle motif has the general formula $\text{CX}_2\text{CX}_4\text{HX}_7\text{C}$, although the spacing of the conserved residues may differ in different elements (Covey 1986; Summers et al. 1990). Retroviral Gags usually have one or two zinc knuckles; the *D. melanogaster* retrotransposons studied here each have three. In both retroviral and retrotransposon Gags, the MHR is slightly N-terminal of the zinc knuckle region. These two regions and the sequence between them are strongly conserved, in contrast to the marked sequence variability seen in much of the amino acid sequence of Gag proteins. The MHR-zinc knuckle region appears to have several roles in the retroviral life-cycle, including involvement in multimerization of Gags (Franke et al. 1994; Orlinsky et al. 1996; Singh et al. 2001; Strambio-de-Castillia and Hunter 1992). In the insect retrotransposons discussed here this region also contains a domain, pre C2HC, of unknown function (Doerks et al. 2002). This domain occupies most of the sequence between the MHR and the zinc knuckles (see Supporting Figure S1).

*HeT-A* Gags from the same *D. melanogaster* genome can differ significantly in amino acid sequence, yet the 151 amino acids of their MHR-zinc knuckle regions align with no gaps in spacing and only 15 residues where one or more of the amino acids differ from the consensus (Supporting Figure S1). The only available *TAHRE* Gag sequence is very similar, having only 20 residues which are not identical to all of the *HeT-A* Gags in the alignment. Interestingly, 15 of these *TAHRE* residues are at sites where *HeT-A* Gags are not all identical and for most sites *TAHRE* has the amino acid found in the majority of the *HeT-A* Gags (Figure S1). Therefore most of the differences in the *TAHRE* sequence are ones that are tolerated in *HeT-A* Gag as well.
Sequence variation in TART elements is concentrated in the untranslated regions, which define three subfamilies, TART A, TART B, and TART C (SHEEN and LEVIS 1994). The MHR-zinc knuckle regions in Gags of the TART subfamilies also have 151 amino acids, all identical except for two residues in TART C. The TART sequence in this region aligns with the sequences from HeT-A and TAHRE with no gaps and no misalignment of CCHC residues; however, there are more amino acid differences between TART and HeT-A than between TAHRE and HeT-A (Figure S2). TAHRE and the canonical HeT-A have 95% identity in this region but only 50 and 52% identity, respectively with TART. Because HeT-A Gag interacts as efficiently with TART Gag as with TAHRE Gag, it appears that these amino acid differences are tolerated.

The D. melanogaster genome has many non-LTR retrotransposons that do not transpose into telomeric arrays (KAMINKER et al. 2002). Gags of these non-telomeric elements also have a MHR-zinc knuckle region with three zinc knuckles. However the MHR-zinc knuckle regions of Gag in the non-telomeric elements differ more from the regions in HeT-A, TART, and TAHRE than the regions in the telomeric Gags differ from each other. These differences are easily seen in the spacing of the CCHC residues and in their spacing relative to the MHR (Supporting Figure S3). All of the sequences from telomeric Gags have identical spacing while the other three sequences differ in spacing from the telomeric Gags and from each other. Jockey and Doc have only 27-31% amino acid identity with each other or any of the telomeric Gags, while I factor has approximately 17% amino acid identity with any of the other sequences. HeT-A Gag does not form functional associations with Gag proteins from Doc, jockey, or I factor (PARDUE et al. 2005). This specificity is similar to that the MHR-zinc knuckle region of
retroviruses which forms heteromultimers only between genetically related retroviruses (FRANKE et al. 1994). The sequence differences between the non-telomeric Gags and telomeric Gags support the hypothesis that the MHR-zinc knuckle region is involved in the association between telomeric elements.

**In conclusion.** These studies provide new evidence that Gag protein localization is important in the transposition of the three telomere-specific retrotransposons of *D. melanogaster*. Of the three telomere-specific retrotransposons only *HeT-A* encodes a Gag protein that is specifically localized to telomeres. Nevertheless both *TAHRE* and *TART* Gags can be directed to telomeres by association with *HeT-A* Gag. Interactions between any of the three Gag proteins depend on the segment containing the MHR, pre_C2HC, and zinc knuckle motifs. The amino acid sequence in this region has a highly conserved pattern that is specific for the telomere retrotransposons. The conservation of this segment in these unusually variable proteins suggests the importance of the Gag interactions for the retrotransposons.

Gags of the three telomere elements differ in their ability to localize to telomere Het dots. *HeT-A* Gag is independently able to achieve this. *TART* Gag must have the help of *HeT-A* Gag yet is able to localize to the nucleus independently (RASHKOVA et al. 2002a). Thus it moves into an optimal position to encounter *HeT-A* Gag for localization to Het-dots. *TAHRE* Gag requires assistance to move from the cytoplasm so is less efficient than *TART* Gag in encountering *HeT-A* Gag for localization to Het dots and thus less likely to have carried in the template for addition to telomeres. This could explain the rarity of *TAHRE* in telomeres, one complete and three truncated copies in
the *D. melanogaster* genome sequenced by the genome project (Abad et al. 2004).

Similarly, the observation that *HeT-A* is consistently more abundant than *TART* in different stocks of *D. melanogaster* (George et al. 2006) may reflect the fact that *TART* Gag needs *HeT-A* Gag for telomere localization. This correlation between the abundance of each element and the efficiency of its Gag in localizing to Het dots provides additional support for the hypothesis that Gag localization is important for targeting telomere-specific transposition.

**ACKNOWLEDGMENTS**

We are grateful to Alfredo Villasante for the BAC containing the complete *TAHRE* element. Members of the Pardue lab and Elena Casacuberta have provided helpful discussion and comments on this manuscript. We particularly thank Laura Croal for helpful advice. The work was initiated in a student laboratory supported by a grant to MIT from Howard Hughes Medical Institute and has been supported in part by National Institutes of Health Grant GM50315 (to M.-L.P.).
LITERATURE CITED


KAMINKER, J. S., C. M. BERGMAN, B. KRONMILLER, J. CARLSON, R. Svirskas et al., 2002


FIGURE LEGENDS

FIGURE 1. The telomere retrotransposons of D. melanogaster. These three non-LTR retrotransposons are drawn, approximately to size, as the RNA molecules that are reverse-transcribed onto the telomere. Thick black lines: untranslated regions; gray ovals: Gag and Pol coding sequences; AAAAA: poly(A) 3’ tail. Dotted lines enclose regions of significant nucleotide identity between TAHRE and the other elements. The range of identity in pairwise comparisons of TAHRE with different copies of HeT-A or TART is given for each region. The TAHRE 5’ UTR is shorter and the 3’ UTR is longer that the URTs of HeT-A. The extra sequences were not included in calculating identity.

FIGURE 2. TAHRE Gag proteins. a) Comparison of TAHRE Gag with HeT-A and TART Gags showing locations of amino acid motifs shared with retroviral Gags. White hexagon: major homology region (MHR); Gray bars: zinc knuckle motifs. The region between the hexagon and the gray bars in insects has been named the pre_C2HC domain. Lines below HeT-A and TART diagrams indicate regions of those proteins expressed by deletion constructs designed to test specific parts of the Gags. First and last amino acids of each deletion construct are given. For both HeT-A and TART, Gag proteins can differ in length between different copies of the element. The amino acid numbers given here refer to the sequences used for the tagged proteins in this study.

FIGURE 3. Immunoblot of proteins expressed by cultured cells transfected with constructs carrying TAHRE sequences. Lanes are labeled with the fluorescent tag on the protein expressed by the transfected construct. Control cells were not
transfected. The filter was probed with an antibody to GFP, which also recognizes CFP and YFP. All constructs express a protein of the size calculated from the sequence of the fusion protein, 133 kD. The weaker bands of non-specific antibody binding seen in all lanes serve as a loading control.

**FIGURE 4. Intracellular localization of TAHRE, HeT-A, and TART Gag proteins in transiently transfected D. melanogaster cultured cells.** A and B) GFP- tagged TAHRE Gag in single transfections, showing most of the protein clustered in the perinuclear cytoplasm or superimposed on the lamin. C) YFP-tagged HeT-A Gag, forming large round Het dots in the nucleus and a single round cytoplasmic Het body. D) YFP-tagged TART Gag, forming small irregular dots completely contained within the nucleus or superimposed on the lamin. E) CFP- tagged TAHRE Gag plus YFP-tagged HeT-A Gag, colocalizing to regular nuclear Het dots preferentially near the edge of the nucleus. F) CFP- tagged TAHRE Gag plus YFP-tagged TART Gag, colocalizing in diffuse clusters mostly within the nucleus. Panels from right to left show: 1) false-colored Gag fluorescence (green, yellow, or yellow and cyan) plus fluorescent antibody staining lamin (red), 2) anti-lamin staining (red) plus DIC of the cell, 3) black and white view of GFP channel in A and B and YFP channel in C-F, 4) black and white view of CFP channel in E and F.

**FIGURE 5. Intracellular localization of Gag protein deletion derivatives in transiently transfected D. melanogaster cultured cells.** A) YFP- tagged HeT-A Gag 482-689 in single transfection, forming irregular clusters completely excluded from the
nucleus. **B)** YFP- tagged *TART* Gag 533-970 in single transfection, forming varied size clusters in the cytoplasm. **C)** Co-transfected YFP- tagged *HeT-A* Gag 482-689 plus CFP-tagged *TAHRE* Gag, colocalizing completely in cytoplasmic aggregates. **D)** Co-transfected YFP- tagged *TART* Gag 533-970 plus CFP-tagged *TAHRE* Gag, colocalizing completely in cytoplasmic aggregates. Panels from right to left show: 1) False colored Gag fluorescence (yellow or superimposed cyan and yellow) plus fluorescent antibody staining lamin (red), 2) anti-lamin staining (red) plus DIC of cell, 3) black and white view of YFP channel, 4) black and white view of CFP channel in E and F.
Fig. 1

HeT-A

```
| Gag | AAAA |
```

74-78%

TAHRE

```
| Gag | Pol  | AAAA |
```

61-70%

56%

TART

```
| Gag | Pol  | AAAA |
```

**Fig. 2**

- **TAHRE**: 965 aa
- **HeT-A**: 921 aa
- **TART**: 970 aa
Figure 5

A  HeT-Δ

B  TARTΔ

C  HeT-Δ  TAHRE

D  TARTΔ  TAHRE