The termini of extrachromosomal linear *copia* elements

Andrew J. Flavell* and Cathy Brierley

Department of Biochemistry, The University, Dundee DD1 4HN, UK

Received 27 February 1986; Revised and Accepted 8 April 1986

ABSTRACT

Extrachromosomal linear *copia* elements were isolated and cloned from cultured *Drosophila melanogaster* cells. Four full length *copia* recombinants were characterised. Each contains a full sized copy of *copia* bounded by its direct repeats. The sequences of the ends of the *copia* inserts were determined. All of the termini are identical to those of genomic *copias* except for a single base deletion from one end of one clone. These results suggest that the priming for initiation of *copia* reverse transcription differs from normal retroviral models. Southern blot analysis of uncloned full length *copia* circles shows that approximately 50% correspond to the predicted circularisation product of such linears. We propose that this class of circles and the linear DNAs described here are precursors to integrated genomic *copia* elements.

INTRODUCTION

The *copia* mobile genetic element of *Drosophila* has strong similarities to vertebrate retroviruses (1). Three lines of evidence link these two classes of DNA element. First, they share common structural features of long direct terminal repeats (LTRs) which contain both transcriptional control signals and small inverted terminal repeats. The LTRs flank an internal region of several thousand nucleotides. There is sequence homology between retroviral genes contained in the internal region, including the reverse transcriptase gene, and sections of the *copia* genome (2,3). Second, retrovirus-like particles containing *copia* RNA and a reverse transcriptase-like activity have been detected in cultured *Drosophila* cells (4). Third, both cultured *Drosophila* cells and flies contain small amounts (0.01-0.1 copies per cell) of extrachromosomal *copia* DNAs with structures very similar to extrachromosomal retrovirus proviral DNAs. These include closed circular molecules with one or two LTRs (5-7) and full length linear *copias* bounded by LTRs (8). The synthesis of linear extrachromosomal *copias* in cultured cells is unaffected by aphidicholin, an inhibitor of DNA polymerase α but not of reverse transcriptases (8), suggesting strongly that the latter enzyme is responsible for their synthesis.

copia shares some of these properties not only with vertebrate retroviruses but also with other transposable elements in Drosophila, collectively termed copia-like elements, retrotransposons or mdg elements (1). These elements also appear to transpose via RNA intermediates. DNA-RNA hybrids representing putative reverse transcription intermediates of the mdg 1 and mdg 3 elements have been isolated from Drosophila (9). The most convincing evidence of RNA mediated transposition of copia-like elements comes from yeast where an experimentally modified Ty element transposes with loss of an added intron and reconstitution of the altered 5' LTR (10). It is therefore highly likely that the retrovirus related mobile elements are united by a common reverse transcription transposition mechanism.

Extrachromosomal linear DNAs are intermediates in the life cycles of vertebrate retroviruses. They are the product of reverse transcription of virion RNA and are precursors to circular DNAs, some of which become integrated into the host cell genome (1). We have cloned copia extrachromosomal linears for three reasons: First, to unambiguously characterise these molecules. Surprisingly, only one linear extrachromosomal retrovirus provirus (and no other copia-like element) has been cloned previously (11). Second, to investigate the copia reverse transcription mechanism. copia does not possess a conventional (-) strand primer binding site (see Results section and Figure 5). This raises the question of how reverse transcription initiates in this element. The termini of the extrachromosomal linear DNA are created by the initiation of (+) and (-) strand synthesis. Therefore, by determining the sequences of these termini we could address this question. Third, to test whether some copia extrachromosomal circles might be formed by an alternative mechanism to reverse transcription. We previously cloned several full length copia circles (2 LTR circles) from cultured Drosophila cells. By analogy with retroviruses we presumed that these molecules were the circularization products of linears derived from reverse transcription. However, when we analysed the LTR-LTR putative circularization junctions of these circles we discovered that the majority contain insertions between one and fifteen nucleotides (6). The largest two inserts (of fifteen bases each) are not present in copia and appear to be of random sequence (the others are 1-3 nucleotides long). Such random insertions have not been seen in any of the many retroviral circle junctions that have been studied (1) and are inconsistent with accepted reverse transcription models. As we were confident that copia extrachromosomal linears were reverse transcription products we could test our hypothesis that these 2LTR circles were synthesized

by an alternative mechanism by looking for the presence or absence of extra sequences at the ends of the linears.

The data presented here show that copia extrachromosomal linears are precise copies of the genomic elements with respect to their terminal sequences. This is in contrast to both the normal retroviral model of reverse transcription initiation and the hypothesis that the insert-containing copia circles are derived from linears. We also show that approximately one half of uncloned copia 2 LTR circles lack junction inserts and therefore correspond to the circularization product of these linears. We suggest that the linears described here are the product of copia reverse transcription and precursors to these insert-lacking copia circles and integrated genomic elements.

MATERIALS AND METHODS

Isolation and analysis of extrachromosomal DNA

Extrachromosomal DNA for exonuclease assay was isolated from 5×10^8 Kc cultured Drosophila cells (12) labelled for 24 hours with 5-bromodeoxyuridine (8). The 5-bromodeoxyuridine substituted DNA was predigested with Hpa II restriction endonuclease. These two steps were necessary to eliminate contamination of the preparation with high molecular weight chromosomal DNA, similar results were obtained with unsubstituted DNA (not shown). One half of the preparation was digested with three units of ATP-dependent exonuclease from Micrococcus lysodeikticus for one hour (13; a gift from Toyobo Co. 8 Dojima Hamadori, Osaka, Japan). Samples were assayed for copia sequences by Southern blotting (14). For Nsp H1 digestions, a sample of extrachromosomal DNA isolated from 5×10^8 Kc cells by the method of Hirt (15) was digested with Hpa II and ATP-dependent exonuclease to remove chromosomal DNA and extrachromosomal linear DNAs. The sample was then digested with restriction enzymes as shown in Figure 4. Densitometric scanning of Southern blot autoradiographs used an LKB 2002 Ultrosan instrument.

Cloning and Sequence Analysis of copia extrachromosomal linear DNAs

Extrachromosomal DNA was isolated from 10^9 confluent Kc cells (8,15). The sample was treated with RNAase A in TE buffer ($50 \mu\text{g} \cdot \text{ml}^{-1}$, 37° 30 min, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA), phenol extracted, ethanol precipitated, redissolved in 500 μl TE and size fractionated by 5-20% sucrose gradient ultracentrifugation (16). Fractions were assayed for copia linears by Southern blotting (14). Half of the recovered linear sample was treated with Klenow DNA polymerase (16). Both halves (0.5 ng copia linears each) were

FIGURE 1 Exonuclease sensitivity of extrachromosomal copia elements. Lane 1, untreated sample. Lane 2, Hpa II digest. Lane 3, Hpa II and exonuclease. copia supercoils, nicked circles and linears plus chromosomal DNA and a 5.2 kb copia linear marker fragment are marked (5,6,8).

ligated to Bam HI linkers (New England Biolabs) at 16° 16 hr (16) and then inserted into 0.25 µg Bam HI-digested, dephosphorylated bacteriophage NM 570BV2 DNA in 6 µl solution at 4° for 16 hr (16,17). During these experiments we observed that carrier RNA markedly inhibits linker ligations. Three different samples of tRNA and one of polyadenylated mRNA produced the same effect. Recombinants were packaged, plated and screened for copia homology (16). Positive clones were grown up and those containing correctly sized

FIGURE 2 Restriction maps of cloned copia extrachromosomal linear DNAs. Maps for Taq I (T), Hind III (HIII), Eco RI (RI), Eco RV (RV) and Hind II (HII) sites. A deletion of approximately 110 base pairs of pCB 6(+) relative to the other inserts and genomic copias (2,3,20) is shown. Sequencing of the ends of the inserts was from Eco RI and Hind III sites in the polylinker which flank the Bam HI insert. Sequencing of the putative (-) primer binding regions of pCB 18(-) and pCB 21(-) was from an Apa I site at 283 nucleotides. All sequences were derived by the method of Maxam and Gilbert (19) from 5'-end labelled DNA (●).

inserts subcloned into pUC 8 (18). Terminal sequences were determined by the method of Maxam and Gilbert (19) from the flanking Eco and Hind sites in the polylinker. The (-)-strand primer binding region of pCB 18 (-) and pCB 21(-) were determined from the Apa I site at 283 nucleotides on the copia sequence (2,3). Clones designated (+) and (-) derived from Klenow treated and untreated samples respectively.

RESULTS

Cloning and sequence analysis of extrachromosomal copia linears

We cloned copia extrachromosomal linears from cultured Drosophila cells by the addition of Bam HI linkers. To ensure that we were isolating a representative sample of these molecules we took two precautions. First we assayed for the presence of any blocked ends by exonuclease digestion (13; Figure 1). This treatment degraded all copia linears in the preparation, leaving copia circles intact (prior treatment with Hpa II aided the exonuclease digestion of contaminating genomic DNA by reducing its size, copia is not cleaved by Hpa II). Second, to avoid any possible problems with recessed ends, one half of the preparation was treated with Klenow DNA polymerase before cloning. This turned out to be unnecessary, indeed the cloning efficiency of linears from the untreated sample was higher than that of the treated fraction. We isolated four cloned copia extrachromosomal linears, one from the Klenow-treated sample (pCB 6+) and three from the untreated sample (pCB 7(-), 18(-) and 21(-)). Restriction mapping with six enzymes gave identical maps to the sequenced genomic elements (2,3) with the single exception that pCB 6(+) bears a deletion of approximately 110 base pairs between a Taq I site at 2539 and an Eco RI site at 2766 on the copia sequence (2,3; Figure 2). This region contains two tandem 108 base repeats of unknown function in the interior of the element (2,3,20). We presume pCB 6(+) contains only one copy of this sequence. Such deletion variants were not detected in Southern blots (data not shown) and are either rare or the deletion took place during the cloning.

The terminal sequences of the four clones are shown in Figure 3. Seven of the eight termini are identical to the ends of genomic copia (2,3), the eighth bears a one base pair deletion from the 3' end of the element (as defined by the polarity of the full length copia transcript). This clone derived from the sample treated with Klenow polymerase and the deletion may have been caused by this treatment. Alternatively, this deletion may have resulted from an aberrant step in the reverse transcription process. 2 LTR

FIGURE 3 Sequences of the ends of cloned copia extrachromosomal linears. The terminal sequence of genomic copia (2,3) flanked by an unspecified five base pair duplication of host sequence (►) is compared with the termini of four cloned extrachromosomal linears.

retroviral circles sometimes contain small deletions at their LTR-LTR junctions (1) which may be similarly derived. In no case do we see any extra bases at the ends of the cloned linears. This contrasts with our data for 2 LTR copia circle junctions where six out of seven clones contained inserts (6). We conclude that the copia linears which we have cloned are not precursors to the insert-containing circles. However the single cloned 2 LTR circle which lacks an insert, pBB 58 (6), may have derived by circularization of this type of molecule.

Approximately 50% of 2 LTR copia circles lack junction inserts

Our previous cloning studies had characterised seven 2 LTR copia circles, one of which lacked a junction insert. However, we could not be sure that this was representative of the total circle population owing to the small number of molecules characterised. To determine whether 2 LTR circles lacking inserts comprise a significant proportion of full length copia circles we digested a sample of Kc cell extrachromosomal DNA with Nsp HI restriction endonuclease (Figure 4). This enzyme cleaves the LTR-LTR junction of insert-lacking copia circles, leaving insert-containing junctions intact (unless they contain a fortuitous Nsp HI site). Additionally, copia has one other Nsp HI site which serves as a useful control for the digestion. 2 LTR copia circles containing junction inserts and 1 LTR circles are linearised by this enzyme whereas 2 LTR circles lacking inserts yield two fragments of 3.1 and 2.0 kilobase pairs. Figure 4 shows that fragments of this size were present in the digest in addition to linearised 2 LTR circles.

FIGURE 4 Approximately 50% of *copia* extrachromosomal linears lack sequence inserts at their LTR-LTR junctions. **A.** Predicted restriction maps of 1 LTR and 2 LTR *copia* circles are shown linearised at the unique internal Nsp HI site at 2011 nucleotides (2,3). Bracketed fragments are present only in 2 LTR circles containing an Nsp HI site at the LTR-LTR junction (□). **B.** Fragments obtained from digestion of an extrachromosomal DNA sample. Lane 1; Undigested nicked *copia* circles. Lane 2; Nsp HI and Xba I. Lane 3; Nsp HI. Lane 4; Nsp HI and Pvu II. Fragment sizes are deduced by comparison with parallel size markers. The band X is an irreproducible artefact in this preparation which is unaffected by any of the enzymes used.

We mapped the extra site to the LTR-LTR junction by double digestion with Pvu II or Xba I. These enzymes cleaved the two fragments in the expected manner (Figure 4). The Nsp I digestion was shown to be complete by the total linearization of 1 LTR circles. Furthermore, digestion of the cloned insert-lacking *copia* circle pBB 58 shows that the junction site is as sensitive to Nsp HI as the internal site (data not shown). By densitometric scanning of these bands we estimate that $50 \pm 10\%$ of 2 LTR *copia* circles in Kc cells lack junction inserts and therefore correspond to the predicted circularisation product of the *copia* linears which we have cloned.

DISCUSSION

The lack of any extra sequences at the ends of *copia* extrachromosomal linears is interesting for two reasons. First, it suggests that the initiation of *copia* reverse transcription differs from that of the vertebrate

FIGURE 5 Extrachromosomal intermediates in the retrovirus life cycle. Reverse transcription is initiated from a bound tRNA (←⊖). The plus strand initiation site is shown (⊕). The gap nucleotides between the minus and plus initiation sites and the LTRs appear at the 3' (■) and 5' (□) ends respectively of the linear provirus in the one case studied (11) and are at the LTR-LTR junction of the 2 LTR circle. Integration of the 2 LTR circle is accompanied by the loss of these bases plus a duplication of several nucleotides of host sequence (▶). The minus strand primer site sequence of Moloney murine leukaemia virus (MuLV) is compared with the corresponding regions of *copia*, the human endogenous retrovirus Clone 4-1 and two *Drosophila copia*-like elements (21-24). The LTRs are boxed and the TGG 5' end sequences of the primer regions are underlined.

retroviruses. The first step in reverse transcription of retroviruses involves elongation of a specific tRNA primer from its 3' end which is base paired to the virion RNA (1; Figure 5). DNA polymerisation proceeds to the 5' end of the RNA and, after a temporary halt, continues at the 3' end of the same or another virion RNA to complete the (-) DNA strand. The first deoxyribonucleotide base synthesised eventually becomes the 3' terminal nucleotide of the extrachromosomal linear. Nearly all retroviruses studied to date possess a two base gap between the tRNA-homologous region and the LTR boundary (as defined by the genomic provirus; Figure 5). These two bases are present at the 3' end of the only linear retrovirus provirus to be cloned previously (11). The analogous region of *copia* shares some similarities with other primer regions, including a TGG triplet corresponding to the CCA 3' end of tRNAs. *copia* has a five base gap between this triplet and the LTR boundary (Figure 5). If *copia* reverse transcription were initiated from this triplet

we would expect to see five extra bases (GGTTA) at the 3' ends of copia linears. Why do we not see these nucleotides? One possibility is that the particular molecules which we cloned lack these sequences at the LTR-interior junction. We eliminated this possibility by determining the (-) primer region sequence of two of the four cloned linears. Both sequences are identical to the normal primer region sequence shown in Figure 5, indeed all three genomic copias which have been sequenced in this region possess the same five nucleotides (2,3,25). Therefore, either these bases are removed during maturation of copia linears or reverse transcription of copia is initiated by an RNA lacking a CCA 3' end or another molecule entirely. We cannot exclude any of these alternatives at present. Reverse transcription of the mdg 3 element is initiated from an RNA but mdg 3 lacks a TGG triplet at the LTR-interior junction (26). copia may also use an RNA primer without a CCA 3' end for initiating its reverse transcription. We are presently studying copia nascent DNA-RNA hybrids to resolve this question.

Our interest in the copia (-) strand primer region drew our attention to the same region of other retrovirus related elements (Figure 5). copia-like elements of Drosophila have gaps of various sizes and some seem to contain a part of the primer region inside the LTR (23,24; Figure 5). Additionally, the human endogenous retrovirus-like element Clone 4-1 contains a five base pair gap between its LTR and a 16 out of 18 base complementary match with rat tRNA^{Glu}. Such variation demonstrates both the considerable evolutionary divergence between these elements and the tolerance of the reverse transcription mechanism to this heterogeneity. It would be interesting to investigate the termini of extrachromosomal linears of the other copia-like elements to see whether they too are flush-ended with respect to the genomic elements.

copia circles lacking inserts presumably derive from circularisation of the linears which we cloned. At least the majority, and possibly all copia extrachromosomal linears are synthesised by a reverse transcriptase-like enzyme (8). We therefore suggest that copia circles lacking inserts are the normal product of reverse transcription. Such circles are more prevalent in the 2 LTR circle population than we supposed from our cloning experiments. This may be a statistical anomaly due to the small number of circles we originally cloned or a result of alterations in the Kc cell line in the period since the circles were cloned. If copia circles which lack sequence inserts at their LTR junctions are the normal reverse transcription product we assume by analogy with 2 LTR retrovirus proviruses (27) that they are the

precursors to integrated copias.

Our second reason for interest in the terminal sequences of copia linears was to test whether these were the precursors to the copia circles containing sequence inserts. Our data strongly suggest that they are not. Insert-containing circles do not seem to be directly produced from linears. Furthermore, such molecules have not been detected in any retrovirus circle preparations examined to date (1). We have suggested that they may be the result of excision of genomic copias (6). The data presented here are consistent with, but do not prove this suggestion. However, excision of the copia-like gypsy element from the Drosophila genome has been observed (28) and excision of Drosophila P transposons, both precise and imprecise has been reported (29). We think it likely that insert-containing copia circles are incapable of integration owing to the inability of the copia integrase enzyme to recognise and cleave the LTR-LTR junction. They therefore may accumulate in the Kc cells while functional circles become integrated.

ACKNOWLEDGEMENTS

We thank Dr Yuri Ilyin for communicating his laboratory's results prior to publication. We are also grateful to Luke Alphey and Sarah Newbury for criticisms of the manuscript. This work was supported by MRC Project Grant G8329473CB.

***To whom correspondence should be addressed**

REFERENCES

- 1 Varmus, H.E. (1982) in 'Mobile Genetic Elements' Ed Shapiro, J.A. (Academic Press, New York) 411-503.
- 2 Mount, S.M. and Rubin, G.M. (1985) Mol. Cell. Biol. 5, 1630-1638.
- 3 Emori, Y., Shiba, T., Kanaya, S., Inouye, S., Yuki, S. and Saigo, K. (1985) Nature 315, 773-776.
- 4 Shiba, T. and Saigo, K. (1983) Nature 302, 119-124.
- 5 Flavell, A.J. and Ish-Horowicz, D. (1981) Nature 292, 591-595.
- 6 Flavell, A.J. and Ish-Horowicz, D. (1983) Cell 34, 415-419.
- 7 Ilyin, Y.V., Schuppe, N.G., Lyubomirskaya, N.V., Goreleva, T.V. and Arkhipova, I.R. (1984) Nucleic Acids Res. 12, 7517-7531.
- 8 Flavell, A.J. (1984) Nature 310, 514-516.
- 9 Arkhipova, I.R., Goreleva, T.V., Ilyin, Y.V. and Schuppe, N.G. (1984) Nucleic Acids Res. 12, 7533-7548.
- 10 Boeke, J.D., Garfinkel, D.J., Styles, C.A. and Fink, G.R. (1985) Cell 40, 491-500.
- 11 Scott, M.L., McKereghan, K., Kaplan, H.S. and Fry, K.E. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 4213-4217.
- 12 Echallier, G. and Ohanessian, A. (1970) In Vitro 6, 162-172.
- 13 Anai, M., Hirahashi, T. and Takagi, Y. (1970) J. Biol. Chem. 245, 767-774.

- 14 Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
- 15 Hirt, B. (1967) J. Mol. Biol. 26, 365-369.
- 16 In 'Molecular Cloning' Eds Maniatis, T., Fritsch, E. and Sambrook, J. Cold Spring Harbor Laboratory (1982).
- 17 Klein, B. and Murray, K. (1979) J. Mol. Biol. 133, 289-294.
- 18 Vieira, J. and Messing, J. (1982) Gene 19, 259-268.
- 19 Maxam, A.M. and Gilbert, W. (1980) Meth. Enzymol. 65, 499-560.
- 20 Fouts, D.L., Manning, J.E., Fox, G.M. and Schmid, C.W. (1981) Nucleic Acids Res. 9, 7053-7063.
- 21 Van Beveren, C., Goddard, J., Berns, A. and Verma, I. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 3307-3311.
- 22 Repaske, R., Steele, P.E., O'Neill, R.R., Rabson, A.R. and Martin, M.M. (1985) J. Virol. 54, 764-772.
- 23 Will, B.M., Bayev, A.A. and Finnegan, D.J. (1981) J. Mol. Biol. 153, 897-915.
- 24 Saigo, K., Kugiyama, W., Matsuo, Y., Inouye, S., Yoshioka, K. and Yuki, S. (1984) Nature 312, 659-661.
- 25 Flavell, A.J., Levis, R., Simon, M.A. and Rubin, G.M. (1981) Nucleic Acids Res. 9, 6279-6291.
- 26 Arkhipova, I.R., Mazo, A.M., Cherkasova, V.A., Gorelova, T.V., Schuppe, N.G. and Ilyin, Y.V. (1986) Cell 44, 555-563.
- 27 Panganiban, A. and Temin, H.M. (1984) Cell 36, 673-679.
- 28 Gerimasova, T.I., Matjunina, L.V., Mizrokhi, L.J. and Georgiev, G.P. (1985) 13, 3781-3787.
- 29 Rio, D.C., Laski, F.A. and Rubin, G.M. (1986) Cell 44, 21-32.