Title
Transposable elements in drosophila and yeast

Author
Mossie, Kevin Mxolisi Godwin

Publication Date
1987

Peer reviewed|Thesis/dissertation

TRANSPOSABLE ELEMENTS IN DROSOPHILA AND YEAST - MOLECULAR MECHANISMS OF TRANSPOSITION

by

KEVIN MXOLISI GODWIN MOSSIE

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

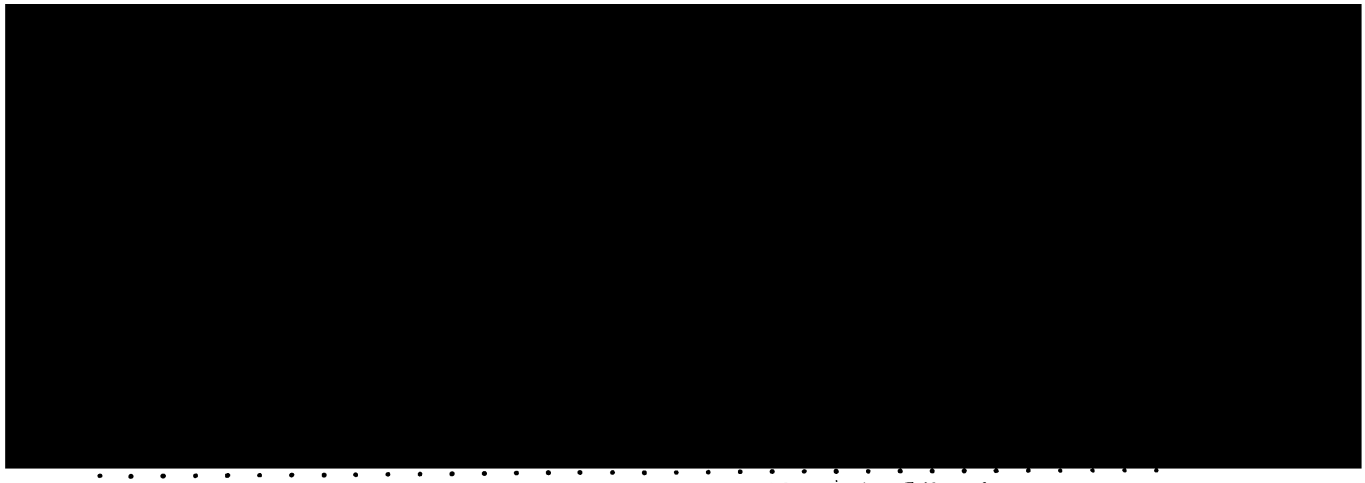
in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco



Date . . . . . University Librarian

Degree Conferred: . . . . . .

'. . . . . almost all great scientists - those who learn to cultivate insight - learn also to respect its mysterious workings. It is here that rationality has its limits.'

BARBARA MC CLINTOCK

Dedicated to the memories of Steven Bantu Biko, Robert Mangaliso Sobukwe, and Nelson Rolihlahla Mandela and all the heroes of our struggle.

The struggle continues....

ACKNOWLEDGEMENTS

Without the help of Herb Boyer, none of this would have happened. It is to Herb Boyer, that offer my sincerest gratitude.

During the six years I spent in the Biochemistry & Biophysics Department at UCSF, I have been fortunate in meeting a number people have had an influence in shaping my life. I therefore wish to thank

Tom Kornberg and Christine Guthrie, who while charged with the task of supervising, among others, a bewildered African, were gracious enough to have both been patient and encouraging during my initial months in San Francisco. Susan Adams, who has been for years doubled as an administrator and a mother for graduate students, and tried to be as helpful as she could.I am grateful to Ira Herskowitz, whose constant enthusiasm always worked in lifting up our spirits even when things were not going well

To my collegues John Majors, Ron Swanstrom, Paul Luciw, the late Richard Parker, Tyler Jacks and Greg Shackelford and my classmates for being part of my life. To my room/labmate, Suzanne Ortiz, for putting up with me for five years and the many hours we have spent from the ghetto to Trump Plaza. To David Westaway, Mike Scott and the entire Yancees clientel, for the many gallons of 'you know what' we've consumed. I am also grateful to Janine Marinos for all her help and for not yelling at me for two years.

I will forever remain indebted to my supervisor, Harold Varmus for being both an advisor and a friend. I am thankful for the many hours and miles we have travelled together.

TABLE OF CONTENTS

ABSTRACT:

Structural similarities between vertebrate retroviral proviruses and copia-like transposable elements in Drosophila and yeast led to suggestions that these elements were evolutionarily related (Temin, 1980). Furthermore, this hypothesis predicted that the mechanism of transposition of these elements could closely resemble the RNA-dependent reverse transcriptase mediated pathway utilised during the replication of vertebrate retroviruses. The obvious prediction of this model, therefore was that transposition intermediates closely resembling retroviral replication intermediates, could exist for these elements. This dissertation describes our efforts in determining the mechanism of transposition of copia-like elements in Drosophila and yeast by initially identifying such intermediates, and the examination of factors which influenced their formation.

While we were succesful in identifying these intermediates for a large number of copia-like elements in Drososphila, we were not able to detect any free, or extrachromosomal Ty forms in yeast, and have isolated and characterised abherrant Ty structures integrated in the genome. We describe the structure of these elements and their possible origin from putative Ty transposition intermediates whose structures resemble those described for retroviruses and the Drosophila elements.

We also describe studies on the structural organisation, and transcriptional

regulation of a differently-structured class of transposable elements in Drosophila which is also believed to transpose via an RNA intermediate. Our results indicate that F elements behave in a manner that closely resemble their mammalian conterparts (the LINE sequences) both in terms of scrambled arrangement of sequences in the genome, and transcriptional properties. Furthermore, we have found that the F RNA strand that is regulated during development (suggesting that it is F-specific), is the same oligoadenylate-terminated strand that has has been proposed to serve as template for the proposed reverse transciptase mediated reaction. Finally, we describe the theoretical basis and preliminary construction of a copia-based 'hopping vector' we hope to use to monitor the transposition of this element in Drosophila.

CHAPTER 1

INTRODUCTION

Reverse-transcriptions is now believed to form a significant pathway by which genetic elements and other sequences in eucaryotes are mobilized (reviewed by Temin, 1985; Varmus and Swanstrom, 1984). While this process has been directly demonstrated for some of these genetic elements, notably retroviruses, para-retroviruses (hepatitis B-like viruses) and the Ty retrotransposon (see Temin, 1985), the bulk of the evidence comes from analysis of structures of various DNAs isolated by molecular DNA cloning. Genetic elements implicated in transposition via an RNA intermediate belong to different, but distinct structural classes which have been conserved throughout evolution. These different structural features suggested that different, but evolutionarily conserved mechanisms must exist for the dispersal of these elements.

The best studied class of transposable genetic elements is the retroviral-like class of elements that include the *copia*-like elements in *Drosophila* (see Spradling and Rubin, 1981; Rubin, 1983) and Ty elements in yeast (Roeder and Fink, 1983). It is generally believed that the transposition pathway for this class of elements, which have been termed retrotransposons (Boeke et al., 1985), is an intracellular version of the retrovirus life cycle. While the details of this pathway are not as well-characterized as to retroviral transposition/replication (reviewed by Varmus and Swanstrom, 1984), the general pathway is thought to be similar to that depicted in Fig. 1. Reconstruction of this pathway comes from our understanding of the retroviral

life cycle and from analysis of cloned DNA segments and characterization of isolated "transposition intermediates" as described below:

1-1.a. *General properties and primary structure of retrotransposons*: Approximately 5-8% of the *Drosophila* genome is made up of 30-50 families of transposable elements which are believed to be structurally homologous (Manning, 1975; Young, 1979; Spradling and Rubin, 1981). The best characterized of these are the *copia* (Levis et al., 1980), *297* (Kugimiya et al, 1983), *412* (Will et al., 1981), B104 or *roo* (Scherer et al., 1982), *gypsy* (Bayer et al., 1984); Modolell et al., 1984), *mdg1* and *mdg3* (Kulguskin et al., 1981; Bayer et al., 1980), *HMS Beagle* (Snyder et al., 1982), and *17.6* elements (Saigo et al., 1984). The first indication that these elements were transposable came from comparison of their location in polytene chromosomes of different strains of *Drosophila melanogaster*. Generally, there is one element at each site, but the number and location varies between *D. melanogaster* stocks, between *D. melanogaster* and other *Drosophila* strains and between embryonic and tissue culture cells (Young, 1979; Young and Schwartz, 1981; Tschurikov et al., 1981). Direct evidence for transposition has been obtained from recently derived mutations where *copia*-like elements have been implicated (Levis et al., 1984; Rubin et al., 1982; Gerasimova et al., 1984).

The initial clues to possible transposition mechanisms came from examination of the primary structure of these elements (Spradling and Rubin, 1981; Rubin, 1983). Members belonging to this class of transposable elements are usually large (5-9 kb) and terminate in direct repeats, approximately 300-600 bp.

termed long terminal repeats (LTRs). Like most other transposable elements of procaryotic and eucaryotic origin (Kleckner, 1977; Calos and Miller, 1980), these elements are usually found flanked by short nucleotides, 4-6 bp. of host sequence which are believed to be duplicated upon integration. The long terminal repeats bounding these elements are themselves flanked by near perfect inverted inverted repeats. Most share the olinucleotides TG at their 5' termini and CA at their 3' termini. In this regard, they are structurally homologous to integrated vertebrate retrovirus proviruses (Varmus and Swanstrom, 1984; Varmus, 1983) and the Ty element in yeast (Roeder and Fink, 1983). Interestingly, the *Drosophila* elements *297*, *17.6*; *gypsy* and HMS Beagle seem to be part of a distinct subgroup; they share AGT or AAT at their 5' ends and ACT or ATT at their 3' termini. Furthermore, they share the collective property of showing preference to integrate at AT stretches (Kugimiya et al., 1983; Bayer et al., 1984; Modolell et al., 1984; Snyder et al., 1982; Saigo et al., 1984; Inouye et al., 1984; Freund and Meselson, 1984).

1-1.b. *DNA sequence homologies*:

Nucleotide sequence analyses of long open reading frames encoded by these elements have further strengthened the evolutionary link (Temin, 1980) between retrotransposons and vertebrate retroviruses (see Varmus, 1983). Most genomes of retroviruses usually consist of three genes required for a complete cycle of viral infection and replication (Varmus, 1983). These code foı the viral core (*gag*), polymerase (*pol*), and envelope (*env*) polyproteins and are arranged in the 5'-LTR-gag-pol-env-3'LTR configuration. While no homologies necessarily exist between the *gag* and *env* genes of retroviruses, homology are observed

within the *pol* region (Chiu et al., 1984). Using computer aided searches, Toh et al. (1983) were the first to report sequence homologies and relatedness between the *pol* genes of retroviruses and the long, unassigned open reading frames of hepatitis B and cauliflower mosaic viruses - 2 classes of DNA viruses, (or pararetroviruses) (Temin, 1985)), that replicate via an RNA intermediate (see Seeger et al., 1985). Using the same approach, sequence homologies between the *pol* region of retroviruses and the retrotransposons *copia* (Emori et al., 1985; Mounat and Rubin, 1985), *17-6* (Saigo et al., 1984), *297* (Inouye et al., 1986), and *gypsy* (Marlor et al., 1986) in *Drosophila*, and the Ty element (Clare and Farabaugh, 1985) in yeast. Using synthetic oligomers corresponding to amino acid sequences best conserved among retroviral reverse transcriptases, York et al. (1986) extended these approaches and successfully identified putative reverse transcriptase genes from two *Drosophila* retrotransposons, *gypsy* and *412*.

The structure of the *pol* gene in the different retrotransposons and retroviruses is interesting in that it is made up of three functional domains (see Fig. 2). These include the protease, reverse transcriptase and associated RNase H, and the endonuclease domains (see Varmus, 1985). The *Drosophila* retrovirus *17-6*, *412*, *297*, and *gypsy* have protein sequences homologous to retroviral protease, reverse, reverse transcriptase and endonuclease domains encoded within the second of three long open reading frames similar to the domains for the Murine Leukemia Virus (MLV) *pol* gene (Saigo et al. 1984; Yuki et al., 1986; Inouye et al., 1986; Marlor et al., 1986; see Varmus, 1985); (See Fig. 2). It is interesting that the order of the homologies in *copia* and

Ty are reversed compared to the MLV *pol* gene in that they are in the order 5' integrase-protease-polymerase-3' (Mount and Rubin, 1985; Emori et al., 1986; Clare and Farabaugh, 1985), suggesting that these elements may have evolved differently from the others. It is also noteworthy to point out that the DNA viruses, hepatitis B viruses and the cauliflower mosaic virus share no homology with the integrase domain, as expected since DNA integration for this class of viruses is not required for virus replication (see Varmus, 1985; Fig. 2).

The expression of the *pol* gene is another feature unique to this class of elements. In most retroviruses (reviewed by Varmus, 1985), the *pol* gene is released from the *gag-pol* polyprotein, a reaction which is believed to be processed by the element-encoded protease. Since in most retroviruses the *gag* and *pol* reading frames are either different, or separated by a termination codon, the manner in which the gag-pol polyprotein was synthesized had remained an enigma. By examining the amino acid sequence of the purified 13KD MLV protease, whose amino terminus is encoded by the last four amino acids of *gag* and the 5' end of *pol*, Yoshinaka et al., (1985), demonstrated that nonsense suppression was the primary pathway for the synthesis of the MLV gag-pol polyprotein. Similarly, Jacks and Varmus (1985) demonstrated that the *gag-pol* polyprotein for the Rous Sarcoma Virus (RSV) were fused via a frameshift suppression pathway. Frameshift suppression has also been shown to similarly express the *pol* genefor other *pol* genes including those for the human immunodeficiency virus (HIV), mouse-mammary tumor virus (T. Jacks, pers. comm.) and the putative *pol* gene for the yeast Ty element (Mellor et al., 1985).

1-2. *Transcription of retrotransposons*:

All of the retrotransposons studied to date have been found to be transcribed (see Spradling and Rubin, 1981). Both full-length and shorter transcripts have been observed for *copia* (Carlson and Brutlag, 1978; Young and Schwartz, 1981), *mdg* 1 and *mdg* 3 (Ilyin et al., 1980; Giorgiev et al., 1981), *412* (Finnegan et al., 1978; Mossie et al., 1985; Young and Schwartz, 1981), *B104* (Scherer et al., 1982), *297* (Rubin, unpublished; Mossie, unpublished), *gypsy* (Parkhurst and Corces, 1986; Mossie, unpublished), and the yeast retrotransposon Ty (Elder et al., 1983). The transcripts are polyadenylated (Falkenthal and Lengyel, 1980; Young and Schwartz, 1981; Scherer et al., 1982; Elder et al., 1983; Emori et al., 1985), and have modified 5' termini (Rubin et al., 1981; Flabell et al., 1981a). Although most of *copia* and *A12* RNAs are associated with polyribosomes (Young and Schwartz, 1982), total *copia* RNA can be translated *in vitro* to produce a 51,000 dalton polypeptide as well as smaller polypeptides (Flavell et al., 1980; Falkenthal and Langyel, 1980). Although mRNA activity has been previously assigned to the smaller 2kb *copia* mRNA (Flavell et al., 1980), isolated virus-like particles enriched in the 5kb *copia* RNA have also been reported to have mRNA activity (Shiba and Saigo, 1983).

Close structural similarities between retrotransposons and retroviruses have led to suggestions that these classes of elements may not only be evolutionary related, but also that the pathway for DNA synthesis may be similar (Temin, 1980; Varmus, 1983; Kugimiya et al., 1983). For retroviruses, viral RNA

synthesis has been shown to initiate at the beginning of R sequences at the left-hand LTR and terminate at the end of the R region in the right LTR (reviewed by Varmus and Swanstrom, 1983). Sequences capable of initiating RNA transcription, putative eukaryotic promotor (TATATAA) and polyadenylation signals (AATAA) have been observed in most LTRs sequenced at appropriate positions (Varmus and Swanstrom, 1983). This terminal redundancy of RNA templates is an important prerequisite for a reverse-transcriptase reaction. The relationship of proviral RNA to proviral DNA is shown in Fig. 3. Retroviral LTRs can therefore be divided into the U_3-R-U_5 format where U_3 represent sequences unique to the 3' end of viral RNA, the R region being short sequences common at both ends of viral RNA, and U5 being unique to the 5' end of the RNA, respectively.

Full-length retrotransposon transcripts have been shown to have their ends within the LTRs. The 3' ends of RNAs specific for the *copia* (Emori et al., 1985), Ty 1 (Elder et al. 1983), and *B104* (Scherer et al., 1982) have been determined by nucleotide sequence analyses of cDNA clones derived from poly A_+ RNA in cells containing these elements. Primer extension studies have been used to map the 5' termini of *copia* RNAs. Other approaches used include filter hybridization of cloned DNA segments to demonstrate that the transcripts encoded by *mdg*1 and *mdg*3 elements (Georgiev et al., 1981) and *copia* (Young and Schwartz, 1981) initiated and terminated within the LTRs. For *copia*, the R region is between 77-116 bp (Flabell et al., 1980; Emori et al., 1985), and 45 bp for Ty1 (Elder et al., 1983). Nucleotide sequence analyses of LTRs of other retrotransposons (see Rubin, 1983) confirm that promoter and polyadenylation

signals are present at positions to place them in the same category as that of other analysed elements. These findings, therefore, provide evidence that full-length retrotransposon transcripts are suitable candidates to serve as templates for a reverse-transcriptase mediated reaction, analogous to retroviral RNAs.

1-3. *Reverse transcription and synthesis of transposition intermediates*:

1-3-1: *Compartmentalization of RNA templates may be a required step for initiating the reverse transcription reaction*: A common characteristic of Tyl and the *Drosophila* retrotransposons is that they are heaviliy transcribed to give abundant poly A^+ RNAs (reviewed by Rubin, 1983; Elder et al., 1983). *Copia* and Ty for example, can account for up to 3% of total Poly A^+ RNA and yet, the transposition frequency is low in these cells (Roeder and Fink, 1983). This lack of correlation between the amount of detectable RNA template and putative transposition intermediates has been confirmed for other *Drosophila* retrotransposons (Mossie et al., 1985). The initial indications that reverse transcription may occur in specialized particles came with the detection of intracellular viral-like particles with associated reverse transcriptase activity in cultured *Drosophila* cells (Heine et al., 1980). Associated with this particlate fraction was a high molecular weight poly A^- containing RNA ca. 30-35S. These observations were confirmed by Shiba and Saigo (1983) who also reported that not only was the large, or heavy RNA (H-RNA) homologous to the *copia* retrotransposon, but could also be translated, *in vitro*, to give virus-like particle-specific polypeptides. Furthermore, these virus-like particles had detectable reverse transcriptase activity using

synthetic oligomeric templates (Shiba and Saigo, 1983). Similar observations have recently been reported for the Ty element in yeast (Garfinkel et al., 1985; Mellor et al., 1985) where a direct correlation between Ty transposition and the appearance of Ty-RNA-containing intracellular particles has been noted. Taken together, these observation suggest that among the rate-limiting steps in the transposition process, efficient packaging of retrotransposon RNA templates could be one of the rate limiting steps. There is evidence that by analogy with vertebrate retroviruses, retrotransposons also contain sequences which may be required in *cis* to package template RNA into viral-like particles. The first evidence is that Ty1 elements do not induce transposition of Ty2 elements in yeast (G.R. Fink-pers. comm.). The main structural difference between Ty1 and Ty2 has been mapped to a short ca. 800 bp fragment which has been reported to be required in *cis* to package Ty RNAs containing marked beta-interferon sequences (Kingsman et al., pers. comm.). Similar studies have not been extended to the *Drosophila* retrotransposon, although it is tempting to speculate that, by analogy to vertebrate retroviruses, they are likely to be present.

Analysis of the reverse transcriptase reactions in purified Ty RNA-containing particles has indicated that an additional step may also be limiting - the activity of the reverse transcriptase enzyme (Garfinkel et al., 1985). These studies were prompted by observations that the rate of Ty transposition to defined loci was temperature-dependent (Paquin and Williamson, 1984; 1986). Garfinkel et al. (1985) observed that the activity of intracellular particle-associated reverse transcriptase in yeast was optimal at low temperatures,

being highest at 15 C and drastically reduced and practically undetectable at 37 C. This effect of temperature correlates well when *in vivo* transposition frequencies are compared with *in vitro* enzymatic activities (Paquin and Williamson, 1984; 1986; Boeke et al., 1986; Garfinkel et al., 1985).The temperature-effect on the transposition frequency or enzymatic activity of purified *copia* RNA-containing particles in *Drosophila* has not been determined. It must be noted, however, that in *Drosophila* cultured cells, only the *copia* RNA, and none homologous with other *Drosophila* retrotranspons has been detected associated with these intracellular particles. Furthermore, the *copia*-RNA-containing particles were exclusively found in nuclei, whereas the Ty containing particles were exclusively found in the cytoplasm (Garfinkel et al., 1985; Shiba and Saigo, 1983). The evolutionary significance of this difference is not apparently clear, but it should be noted that retroviral reverse transcription also occurs in the cytoplasm of infected cells (Varmus and Shank, 1976; Varmus et al., 1974).

1-3-2: *Reverse transcription: Primers and primer-binding sites*:
In addition to promoter sequences and polyadenylation signals present in each LTR, retroviruses have several sequences which are known to be important for reverse transcription. These are i) the minus (-) primer binding site (-PBS), a sequence immediately downstream from the left-land LTR that permits binding of a tRNA primer molecule used to initiate synthesis of the first DNA strand (minus strand) after infection; ii) a polypurine tract (+PBS) adjacent to the right-hand LTR, that is involved in priming the second strand (plus strand DNA) (Varmus, 1982).

If retrotransposons are related to retrovirus, one would expect to find within retrotransposons some of the conserved functional sequences characteristic of retroviruses. That has indeed proved to be true. Comparison of putative primer binding sites of sequenced retrotranspons has revealed similarities with the primer binding site in avian leukosis sarcoma virus for the retrotransposons *297*, *412*, *17.6*, and *mdg*1 and similarity with murine sarcoma virus for B104 (Lugimiya et al., 1983; Scherer et al., 1982). This approach did not, however, offer such information in the case of the *copia* and *mdg*3 elements which were reported to be more closely related (Shiba and Saigo, 1983). There were, however, no reports published at this stage to support the notion that, as in the case of host cells of retroviruses, *Drosophila* and yeast cells had a set of tRNAs whose 3' terminal, 18 nucleotides were complementary to the putative PBS of the retrotransposons analyzed. The first evidence for this has been with the report that a short (11 base pair long) perfect homology exist between the putative PBS of *gypsy* (*mdg*4) and the *Drosophila* lysine tRNA (Bayer et al., 1984). Inouye et al. (1986), using a synthetic oligomer homologous to the putative primer binding site for the retrotransposon *297*, have identified a species of *Drosophila* serine tRNA that has two distinct properties characteristic of retroviruses: its 3' terminal 18 nucleotides are exactly complementary to the putative PBS of 297 and its 19th base was modified. Furthermore, products of this serine tRNA species could be detected in *Drosophila* cells (Inouye et al., 1986). Recently, Kikuchi et al. (1986) has identified among the VLP associated tRNAs a methionine tRNA capable of serving as a primer for VLP *copia* H-RNA template (Shiba and Saigo, 1983). The

surprising result, however, was that an unusual priming mechanism was involved: minus strand DNA extension did not occur at the 3' end of the tRNA at the internal site of $tRNA^{Met}$ (Kikuchi et al., 1986). 15 nucleotides, 25-39 of $tRNA^{Met}$ were found to be complementary to the region just 3' of the left-hand *copia* LTR, i.e. no nucleotides were inserted between the 5' LTR and the (-) PBS.

1-3-3: *Synthesis of transposition intermediates*:

The replication of all retroviruses is initiated by the covalent addition of a deoxynucleotide to the 3' end of a cellular tRNA molecule that is hydrogen bonded to the primer binding site near the 5' end of viral RNA. The point at which synthesis is begun defines the 5' boundary of U_5. The first DNA product of this reaction (-strong stop DNA) therefore consists of R and U_5 sequences (Varmus, 1982). When the template has been copied, the first "jump" occurs using DNA complementary to the R region to base-pair at the 3' end of viral RNA, and the nascent (-) strand is then extended. Meanwhile, using an unidentified RNA primer, thought to be of viral origin (Smith et al., 1984 a, b), the plus strand is initiated at a position that defines the 5' boundary of U_3, a region that is recognized by the presence of a polypurine tract in viral RNA (Varmus, 1982). The major plus strand, or plus strong stop DNA, is predicted to begin at a specific site in this region, precisely with AATG (or GATG) forming the 5' end of the plus strand (Mitra et al., 1979; Varmus and Swanstrom, 1982). When the minus strand has been suffiently extended to provide template for the arrested plus strand, a duplex is formed between the 3' end of the growing (-) strand and the arrested (+) strand, the so-called

"second jump." Extension of both strands yields a product that is a linear double-stranded DNA, that is colinear and indistinguishable from integrated proviral DNA (Varmus, 1982). Linear proviral DNA in the cytoplasm of infected cells is then transported to the nucleus where it is circularized to produce circular DNA containing one or two copies of the LTR (Guntaka et al., 1976; Fritsch and Temin, 1977; Shank and Varmus, 1978; Yoshimura and Weinberg, 1979). Panganiban and Temin (1984) has showed that the circular DNA species containing two LTRs able to integrate into the host genome. Furthermore, they demonstrated that only a small portion of the circle junction that spans the inverted repeats, that have been previously shown to be essential for viral infectivity (Panganiban and Temin, 1983), can act as a substrate for the virus-encoded endonuclease (Panganiban and Temin, 1984).

The first evidence that similar unintegrated retrotransposon DNA forms existed was the discovery of unintegrated circular DNA molecules hybridizing to *copia* (Flavell and Ish-Horowicz, 1981; 1983). The majority (30/39) of the clones analyzed were identical, consisting of an intact circular DNA molecule containing a single LTR (one had undergone an 800bp deletion of internal sequence, ending at the LTR border). 7/39 of the total clones consisted of circular DNA molecules containing two LTRs in tandem. Although 6 of these clones had between 1-15bp inserts, one clone contained perfectly fused LTRs with no intervening sequences. The other two clones appeared to have undergone autointegration events, since the LTRs are inverted and separated by 1.9kb and 199bp of internal copia sequence, and bounded by a 5bp duplication (Flavell and Ish-Horowicz, 1983) in a manner similar to that observed for circles of

MLV DNA (Shoemaker et al., 1980; Van Beveren et al., 1982). Similarly, unintegrated circular DNA forms have been observed for the other retrotransposons studied (Shepherd and Finnegan, 1985; Mossie et al. 1985; Ilyin et al., 1984). While all of the cloned *mdg*1 and *mdg*3 circles (Ilyin et al., 1984) and most of the *412* circles (17/24) only contained 1 LTR, a *412* containing inverted *412* sequences (4kb) was also identified (Shepherd and Finnegan, 198). Mossie et al. (1985) also demonstrated the existence of circular DNA forms for other retrotransposons (including *copia*, *412*, *297*, *mdg*1, *mdg*3, and *gypsy* (*mdg*4) as well as a number of elements that have not been characterized structurally in either cultured cells or embryos. It is likely that these represent and probably account for most of the heterogenously-sized circular DNA molecules observed in embryos by Stanfield and Lengyel (1979). Similar attempts to identify circular Ty DNA forms were unsuccessful (K. Mossie - unpublished).

While the structure of these molecules offered indirect evidence that these arose by a reverse transcriptase-like pathway similar to that of retroviruses, it was difficult to exclude the possibility that an unknown fraction could have been formed via other events e.g. recombination between direct repeats in integrated elements (Flavell and Ish-Horowicz, 1983). The first direct evidence that unintegrated DNA molecules were formed by a non-conservative mechanism consistent which the reverse transcriptase pathway came from analysis of free DNA in density labelled cells (Flavell, 1984). Most free *copia* DNA, which included linear *copia* molecules that were colinear with integrated *copia* elements, were labelled on both strands (HH) and banded away

from genomic *copia* sequences (HL, LL) (Flavell, 1984). The structure of linear *copia* DNA molecules was later elucidated after molecular cloning and had ends that were identical to those of genomic *copia* elements (Flavell and Brierley, 1986). Cloned *copia* linear DNA molecules provided compelling evidence that they were products of a reverse transcriptase-mediated reaction and were related to previously characterized circular forms (Flavell and Brierley, 1986).

Although evidence for the retrovirus-like transposition pathway was overwhelming, detailed aspects of this pathway were still unclear. Arkhipova et al. (1984) reported on the detection of DNA-RNA complexes containing *mdg*1 and *mdg*3 sequences. Contained within these were perfectly hybrid molecules of *mdg*1 and *mdg*3 associated with poly A^+ RNA. From these complexes, three species of single-stranded DNAs were found: (i) full-sized *mdg*1 and *mdg*3 containing one or two LTRs (both minus strand) and (ii) full-sized plus strand DNA containing a single LTR. In a recent report, they detailed a careful examination of these complexes where they also detected the presence of both minus strong stop and plus strong stop DNAs for the retrotransposons *mdg*1, *mdg*3, and *gypsy* (mdg4) (Arkhipova et al., 1986). They were able to align these DNAs to mdg sequences and, in combination of S1 nuclease analysis of the RNA initiation and termination sites, were able to provide evidence that not only was the synthesis of these molecules linked to a tRNA-sized RNA primer, but also demonstrated that the LTRs of these retrotransposons could be divided into the U_3-R-U_5 retroviral format. The RNA terminal redundancy determined for *mdg*1, *mdg*3, and *gypsy* was 60-70 bp., 50 bp., and 55 bp., respectively

(Arkhipova et al., 1986). It is significant that these studies demonstrated the existence of an RNA primer for *mdg*3, a retrotransposon whose nucleotide sequence analysis failed to reveal any tRNA primer homology at this region, suggesting that utilization of such primers may be different between retroviruses and these elements.

In a classic set of experiments, Boeke et al., (1985) demonstrated that transposition of the yeast retrotransposon, Ty, occurs via an RNA intermediate. Using a marked Ty element containing a ribosomal protein intron, they were able to demonstrate in their transposition assay that Ty-induced revertants at the *HIS* locus had Ty elements which had precisely lost the intron fragment. They also demonstrated the direction of the transposition pathway by showing a mutation at the U_5 region of the 5' delta (LTR) is transferred to the 3' delta in newly transposed Ty elements - consistent with the retroviral replication model (Varmus, 1982; Varmus and Swanstrom, 1982), that U_5 information at the upstream LTR and U_3 information at the right are used during DNA replication, to construct LTR sequences.

1-4: *Genetic interactions between retrotransposons and other genes*:
The expression of retrotransposons studied to date appears to be regulated during development. In *Drosophila*, the different retrotransposons behave as developmentally-regulated transcriptional units, being expressed at different times during development (see Spradling and Rubin, 1981). At present, there is no evidence to indicate how many members of each family are expressed at any one time, nor is it known whether their expression is controlled autonomously

or is merely a reflection of the state of adjacent genes (reviewed by Finnegan, 1986). A number of spontaneous mutations resulting from the insertion of these elements have been described in *Drosophila* (Zachar and Bingham, 1982); Bender et al., 1983; Scott et al., 1983; O'Hare et al., 1984; Kidd et al., 1983; Snyder et al., 1982; Freund and Meselson, 1984; Gerasimova et al., 1984). Suppressor loci exist whose allelic states strongly influence the phenotypes produced by mutant alleles resulting from insertion of *Drosophila* retrotransposons (Green, 1959; Bingham et al., 1981; Bender et al., 1981; Modollel et al., 1983).

The best studied of the extragenic suppressors is the *suppressor of Hairy wing, su(Hw)* (Lewis, 1949). This can suppress the phenotypic effects of particular mutations at least eleven loci scattered throughout the genome, including all mutations in the bithorax complex (Bender et al., 1983) and the *yellow* locus (Parkhurst and Corce, 1986) associated with the *gypsy* insertions. It is noteworthy that suppression occurs without any alteration to the *gypsy* elements themselves (Bender et al., 1983; Parkhurst and Corces, 1986; Modollel et al., 1983). The effect of *su (Hw)* on the *gypsy* elements has been demonstrated as partial inhibition of transcription of *gypsy* elements, and it has been suggested that the mutational effect of *gypsy* is due to developmentally-specific transcriptional interference on *yellow* transcription and that mutations at *su (Hw)* locus reverse this effect by altering the quantitative expression of *gypsy* (Parkhurst and Corces, 1986).

The effect of *su (Hw)* mutation on gypsy insertions is one example of interactions between retrotransposons and other cellular genes. The phenotypic effect of the *white apricot* mutation is reduced by the *suppressor of white apricot su* (w^a), and is increased by the *suppressor of forked su(f)* (Green, 1959). The *white apricot* mutant allele at the *white* locus results from insertion of the *copia* retrotransposon (Bingham and Judd, 1981; Bingham et al., 1981; Rubin et al., 1982). The insertion of the *copia* element in the intervening sequence of the *white* transcriptional unit has been shown to cause premature termination of the *white* transcript within the LTRs of the *copia* insert (Pirrotta and Brockl, 1984; Levis et al., 1984). Zachar et al., (1985) have demonstrated that in *su (Hw)* mutant alleles transcriptional readthrough occurs through the *copia* element, prompting the suggestion that the product of the *su (Hw)* gene was a *copia* LTR-specific transcriptional anti-terminator.

Similar extragenic suppressors have been identified for Ty-induced mutations in yeast. These have been identified as single-copy genes, designated *SPT*, *ROC*, and *TYE* (Ciriacy and Williamson, 1981; Dubois et al., 1982; Winston et al., 1984a) which interact with Ty elements. The best studied of these is the *SPT3* gene (Winston et al., 1984a,b). The product of the *SPT3* gene has been should to influence Ty transcription initiating from delta sequences, since in *spt3* mutants there are little, if any, delta-delta transcripts. For this reason therefore, the *SPT3* gene is also required for Ty transposition (Bocke et al., 1986). The sites of interaction of these transacting genes appear to be located with the long terminal repeats in the case of *copia su* (W^a), and Ty (SPT3) (Winston et al., 1984a,b; Zachar et al., 1986).

The work described in this thesis attempts to molecularly define the mechanism by which retroviral-like elements transpose in *Drosophila* and yeast. Chapter II describes the search for transposition intermediates for *copia*-like retrotransposons in *Drosophila*, the relationship between these, and parameters which could influence their formation. The results indicate that neither the levels of free cellular RNA nor the amounts of element-specific sequences integrated in the genome correlates with the levels of unintegrated, element-specific, circular DNAs present in *Drosophila* cultured cells and embryos. A surprising finding from our screen has been the detection of free circular DNAs specific to F-elements, a class of *Drosophila* transposable elements, which, though not retroviral-like structurally, has also been suggested to transpose *via* an RNA intermediate. Analysis of uncharacterized dispersed repetitive DNA elements indicate that these too were probably *copia*-like, since they displayed similar characteristics to characterized *copia*-like elements.

Chapter III describes analysis of transcriptional regulation during development of F-elements in *Drosophila*. These studies were motivated by observations described above that different retrotransposons appeared to behave as individual transcriptional units that were regulated at different times during development. We describe the results of those analyses and the identification of the putative F-coding strand.

Chapter IV describes our efforts to understand the mechanism of transposition of the yeast retrotransposon Ty, especially since we have been unable to detect unintegrated Ty in different yeast strains. Using a novel approach, we have cloned and characterized Ty forms integrated in the yeast genome whose structure is consistent with the hypothesis that they arose from free TyDNAs.

Appendix A is a description of the construction of a *copia*-based vector that we hope to use to monitor transposition in *Drosophila*.

Literature Cited

CHAPTER 1

1. Temin, H.M. (1985). Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts. Mol. Biol. Evol. *2*, 455-468.

2. Varmus, H.E. and Swanstrom, R. (1985). Supplement: Replication of retroviruses. In the Molecular Biology of Tumor Viruses, part III. RNA Tumor Viruses (Weiss, R.A., Teich, N., Varmus, H.E., and Coffin, J., eds.) pp. 75-134. Cold Spring Harbor Laboratory Press. Cold Spring Harbor.

3. Spradling, A.C., and Fink, G.M. (1981). Drosophila genome organization: conserved and dynamic aspects. Ann. Rev. Genet. *15*, 219-264.

4. Rubin, G.M. (1983). Dispersed repetitive DNAs in *Drosophila*. In Mobile Genetic Elements (Shapiro, J.A., ed.) pp. 329-361. Academic Press, Inc., Orlando, Fla.

5. Roeder, G.S., and Fink, G.R. (1983). Transposable elements in yeast. In Mobile Genetic Elements (Shapiro, J.A. ed.) pp. 299-328. Academic Press, Inc., Orlando, Fla.

6. Boeke, J.D., Garfinkel, D.J., Styles, C.A., and Fink, G.R. (1985). Ty elements transpose through an RNA intermediate Cell *40*, 491-500.

7. Manning, J.E., Schmid, C.W., and Davidson, N. (1975). Interspersion of repetitive and non-repeptitive DNA sequences in the *Drosophila melanogaster* genome. Cell 4, 141-155.

8. Young, M.W. (1979). Middle repetitive DNA: A fluid component of the *Drosophila* genome. Proc. Natl. Acad. Sci. (USA) *76*, 6274-6278.

9. Levis, R., Dunsmuir, P., and Rubin, G.M. (1980). Terminal repeats of the *Drosophila* transposable *copia*: nucleotide sequence and genomic organization. Cell *21*, 581-588.

10. Kugimiya, W., Ikenaga, H., and Saigo, K. (1983). Close relationship between the long terminal repeats of avian leukosis-sarcoma and *copia*-like movable genetic elements of *Drosophila*. Proc. Natl. Acad. Sci. (USA) *80*, 3193-3197.

11. Will, B.M., Bayev, A.A., and Finnegan, D.J. (1981). Nucleotide sequence of the terminal repeats of 412 transposable elements of *Drosophila melanogaster*. J. Mol. Biol. *153*, 897-915.

12. Scherer, G. Tschudi, C., Perera, J., Delius, H., and Pirrotta, Y. (1982). B104, a now dispersed repeated gene family in *Drosophila melanogaster* and its analogies with retroviruses. J. Mol. Biol. *157*, 435-451.

13. Bayev, A.A., Krayev, A.S., Lyubomirskaya, N.V. Ilyin, Y.V., Skryabin, K.G., and Georgiev, G.P. (1980). The transposable element mdg3 in *Drosophila melanogaster* is flanked with perfect direct and mismatched inverted repeats. Nucleic Acids Res. *8*, 3263-3273.

14. Kulguskin, V.V., Ilyin, Y.V., and Georgiev, G. P. (1981). Mobile dispersed genetic element mdgl of *Drosophila melanogaster*: nucleotide sequence of long terminal repeats. Nucleic Acids Res. *14*, 3451-3464.

15. Snyder, M.P., Kimbrell, D., Hunkapiller, M., Hill, R. Fristrom, J., and Davidson, N. (1982). A transposable elemtn that splits the promoter region inactivates a *Drosophila* cuticle protein gene. Proc. Natl. Acad. Sci. (USA) *79*, 7430-7434.

16. Saigo, K., Kugimiya, W., Matsuo, Y. Inouye, S. Yoshioka, K., and Yuki, S. (1984). Identification of the coding sequence for reverse transcriptase-like enzyme in a transposable genetic element, 17.6, in *Drosophila melanogaster*. Nature (London) *312*, 659-661.

17. Young, M.W., and Schwartz, H.E. (1981). Nomadic gene families in *Drosophila*. Cold Spring Harbor Symp. Quant. Biol. *45*, 629-640.

18. Tschurikov, N.A., Ilyin, Y.V., Skryabin, K.G., Ananiev, E.V., Bayev, A.A., Krayev, A.S., Zelentskova, E.S., Kulgushkin, V.V., Lyubomirskaya, N.V., and Georgiev, G. P. (1981). General properties of mobile dispersed genetic elements of *Drosophila melanogaster*. Cold Spring Harbor Symp. Quant. Biol. *45*, 655-665.

19. Rubin, G.M., Kidwell, M.G., and Bingham, P.M. (1982). The molecular basis of hybrid dysgenesis: the nature of induced mutations. Cell *29*, 987-994.

20. Gerasimova, T.I., Ilyin, Y.V., Mizrokhi, L.J., Semjonova, L.V., and Georgiev, G.P. (1984). Mobilization of the transposable element mdg 4 by hybrid dysgenesis generates a family of unstable *cut* mutations in *Drosophila melanogaster*. Mol. Gen. Genet. *193*, 488-492.

21. Levis, R., O'Hare, K., and Rubin, G.M. (1984). Effects of transposable element insertions on RNA encoded by the *white* gene of *Drosophila*. Cell *38*, 471-481.

22. Kleckner, N. (1977). Translocatable elements in procaryotes. Cell *11*, 11-23.

23. Calos, M.P., and Miller, J.H. (1980). Transposable elements. Cell *20*, 575-595.

24. Bayev, A.A., Lyvbomirskaya, N.V., Dzhumagdiev, E.B., Ananiev, E.V., Amiant, I.G., and Ilyin, Y.V. (1984). Structural organization of transposable element *mdg4* from *Drosophila melanogaster* and a nucleotide sequence of its long terminal repeats. Nucleic Acids Res. *12*, 3707-3723.

25. Modolell, J., Bender, W., and Meselson, M. (1983). *Drosohpila melanogaster* mutations suppressible by the suppressor fo hairy-wing are insertions of a 7.3. kilobase mobile element. Proc. Natl. Acad. Sci. (USA) *80*, 1678-1682.

26. Temin, H.M. (1980). The origin of retroviruses from cellular movable genetic elements. Cell *21*, 599-600.

27. Chiu, I.M., Callahan, R., Tronick, S.R., Schlom, J., and Aaronson, S.A. (1984). Major *pol* gene progenitors in the evolution of oncoviruses. Science *223*, 364-370.

28. Toh, H. Hayashida, H., and Miyata, T. (1983). Sequence homology between retroviral reverse transcriptases and putative polymerases of hepatitis B and cauliflower mosaic virus. Nature (London) *305*, 827-829.

29. Seeger, C. Ganem, D., and Varmus, H.E. (1985) Biochemical and genetic evidence for the hepatitis B virus replication strategy (1985). Science.

30. Inouye, S., Yuki, S., and Saigo, K. (1984). Sequence-specific insertion of the *Drosophila* transposable genetic element 17.6. Nature (London) *310*, 332-333.

31. Mount, S.M., and Rubin, G.M. (1985). Complete sequence of the *Drosohpila* transposable element *copia*: Homology between *copia* and retroviral proteins. Mol. Cell. Biol. *5*, 1630-1638.

32. Emori, Y., Shiba, T., Kanaya, S., Inouye, S. Yuki, S., and Saigo, K. (1985). The nucleotide sequences of *copia* and *copia*-related RNA in *Drosophila* virus-like particles. Nature (London) *315*, 773-776.

33. Marlor, R.L., Parkhurst, S.M., and Corces, V.G. (1986). The *Drosophila melanogaster gypsy* transposable element encodes putative gene products homologous to retroviral proteins. Mol. Cell. Biol. *6*, 1129-1134.

34. Clare, J., and Farabaugh, P. (1985). Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. Proc. Natl. Acad. Sci. (USA) *82*: 2829-2833.

35. Yuki, S., Ishimarv, S., Inouye, S., and Saigo, K. (1986). Identification of genes for reverse-transcriptase-like enzymes in two *Drosophila* retrotransposons, 412 and *gypsy*; a rapid detection method of reverse transcriptase genes using YXDD box probes. Nucleic Acids Res. *14*, 3017-3029.

36. Freund, R., and Meselson, M. (1984). Long terminal repeat nucleotide sequence and specific insertion of the gypsy tranposon. Proc. Natl. Acad. Sci. (USA) *81*, 4462-4464.

37. Yoshinaka, Y., Katoh, I., Copeland, T.D., and Oroszian, S. (1984). Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. (USA).

38. Mellor, J.S., Fulton, S.M., Dobson, M.J., Wilson, W. Kingsman, S.M., and Kingsman, A.J. (1985). A retrovirus-like strategy for expression of a fusion protein encoded by yeast retrotransposon Ty 1. Nature (London) *313*, 243-246.

39. Carlson, M., and Brutlag, D.L. (1978). One of the *copia* genes is adjacent to satellite DNA in *Drosophila melanogaster*. Cell *15*, 733-742.

40. Ilyin, Y.V., Chmeliauskaite, V.G., Kulguskin, V.V., and Georgiev, G.P. (1980). Mobile genetic element *mdg1* of *Drosophila melanogaster*: transcription pattern. Nucleic Acids Res. *8*, 5333-5346.

41. Georgiev, G.P., Ilyin, Y.V., Chmeliauskaite, Y.G., Ryskov, A.P., Kramekov, D.A., Skryabin, K.G., Krayev, A.S., Lukanidin, E.M., and Grigoryan, M.S. (1981). Mobile dispersed genetic elements and other middle repetitive DNA sequences in the genomes of *Drosohpila* and mouse: transcription and biological significance. Cold Spring Harbor Symp. Quant. Biol. *45*, 641-653.

42. Finnegan, D.J., Rubin, G.M., Young, M.W., and and Hogness, D.S. (1978). Repeated gene families in *Drosophila melanogaster*. Cold Spring Harbor Symp. Quant. Biol. *42*, 1053-1063.

43. Parkhurst, S.M., and Corces, V.G. (1986). Interactions among the *gypsy* transposable element and the *yellow* and the suppressor fo hairy-wing loci in *Drosohpila melanogaster*. Mol. Cel. Biol. *6*, 47-53.

44. Elder, R.T., Loh, E.H., and Davis, R.W. (1983). RNA from the yeast transposable element Ty has both ends in direct repeats, a structure similar to retrovirus RNA. Proc. Natl. Acad. Sci.(USA) *80*, 2432-2436.

45. Falkenthal, S., and Lengyel, J.A. (1980). Structure, translation, and metabolism of the cytoplasmic *copia* ribonucleic acid of *Drosophila melanogaster*. Biochemistry *19*, 5842-5850.

46. Rubin, G.M., Brorein, W.J., Dunsmuir, P., Flavell, A.J., Levis, R., Strobel, E., Toole, J.J., Young, E. (1981). *Copia*-like transposable elements in the *Drosophila* genome. Cold Spring Harbor Symp. Quant. Biol. *45*, 619-628.

47. Flavell, A.J., Levis, R., Simon, M.A., and Rubin, G.M. (1982). The 5' termini of RNAs encoded by the transposable element *copia*. Nucleic Acids Res. *9*, 6279-6291.

48. Flavell, A.J., Ruby, S.W., Toole, J.J., Roberts, B.W., Rubin, G.M. (1980). Translation and developmental regulation of RNA encoded by the eukaryotic transposable element *copia*. Proc. Natl. Acad. Sci. (USA) *77*, 7107-7111.

49. Shiba, T., and Saigo, K. (1983). Retrovirus-like particles containing RNA homologous to the transposable element *copia* in *Drosohpila melanogaster*. Nature (London) *302*, 119-123.

50. Heine, C.W., Kelly, D.C., and Avery, R.J. (1980). The detection of intracellular retrovirus-like entities in *Drosophila melanogaster* cell cultures. J. Gen. Virol. *49*, 385-395.

51. Garfinkel, D.J., Boeke, J.D., and Fink, G.R. (1985). Ty element transposition: reverse transcriptase and virus-like particles. Cell *42*, 507-517.

52. Mellor, J., Malim, M.H., Gull, K., Tuite, M.F., McCready, S. Dibbayawan, T., Kingsman, S.M., and Kingsman, A.J. (1985b). Reverse transcriptase activity and Ty RNA associated with virus-like particles in yeast. Nature (London) *318*, 583-586.

53. Boeke, J.D., Styles, C.A., and Fink, G.R. (1986). Saccharomyces cereviciae SPT3 gene is required for transposition and transpositional recombination of chromosomal Ty elements. Mol. Cell. Biol. *6*, 3575-3581.

54. Paquin, C., and Williamson, V.M. (1984). Temperature effects on the rate of Ty transposition. Science *226*, 53-55.

55. Paquin, C.E., and Williamson, V.M. (1986). Ty insertions at two loci account for most of spontaneous antimycin A resistance mutations during growth at 15 C of *Saccharomyces cerevisiae* strains lacking ADH 1. Mol. Cell. Biol. *6*, 70-79.

56. Finnegan, D.J. (1983). Retroviruses and transposable elements - which came first? Nature (London) *302*, 105-106.

57. Varmus, H.E., and Shank, P.R. (1976). Unintegrated viral DNA is synthesized in the cytoplasm of Avian Sarcoma virus-transformed duck cells by viral DNA polymerase. *J. Virol. 18*, 567-573.

58. Varmus, H.E. (1982). Form and function of retroviral proviruses. Science *216*, 812-820.

59. Mitra, S.W., Goff, S., Grilboa, E., and Baltimore, D. (1979). Synthesis of a 600 nucleotide-long plus strand DNA by virions of Moloney murine leukemia virus. Proc. Natl. Acad. Sci. (USA) *76*, 4355-4359.

60. Inouye, S., Saigo, K. Yamada, K., and Kuchino, Y. (1986). Identification and nucleotide sequence determination of a potential primer tRNA for reverse transcription of a *Drosophila* retrotransposon, 297. Nucleic Acids Res. *14*, 3031-3043.

61. Kikuchi, Y., Ando, Y., and Shiba, T. (1986). Unusual priming mechanism of RNA-directed DNA synthesis in *copia* retrovirus-like particles of *Drosophila*. Nature (London) *323*, 824-826, 1986.

62. Fritch, E., and Temin, H. (1977). Formation and structure of infectious DNA of spleen necrosis virus. J. Virol. *21*, 119-130.

63. Guataka, R.V., Richard, O.C., Shank, P.R., Kung, H.J., Davidson, N., Fritsch, E., Bishop, J.M., and Varmus, H.E. (1976). Covalently closed circular DNA of avian sarcoma virus: purification from nuclei of infected quail tumor cells and measurement by electron microscopy and gel electrophoresis. J. Mol. Biol. *106*, 337-357.

64. Yoshimura, F.K., and Weinberg, R.A. (1979). Restriction endonuclease cleavage of linear and closed circular murine leukemia viral DNAs: discovery of a smaller circular form. Cell *16*, 323-332.

65. Panganiban, A.T., and Temin, H.M. (1983). The terminal nucleotides of retrovirus DNA are required for integration but not virus production. Nature (London) *306*, 155-160.

66. Panganiban, A.T., and Temin, H.M. (1984). Circles with two LTRs are precursors to integrated retrovirus DNA. Cell *36*, 673-679.

67. Flavell, A.J., and Ish-Horowicz, D. (1981). Extrachromosomal circular copies of the eucaryotic transposable element *copia* in cultured *Drosophila* cells. Nature (London) *292*, 591-595.

68. Flavell, A.J. and Ish-Horowicz, D. (1983). The origin of extrachromosomal circular *copia* elements. Cell *34*, 415-419.

69. Ilyin, Y.V., Schuppe, N.G., Lyubomirskaya, N.V., Gorelova, T.V., and Arkhipova, I.R. (1984). Circular copies of mobile dispersed genetic elements in cultured *Drosophila melanogaster* cells. Nucl. Acids. Res. *12*, 7515-7531.

70. Shepard, B.M., and Finnegan, D.J. (1985). Structure of circular copies of the *412* transposable element present in *Drosohpila* melanogaster tissue culture cells and isolation of a free *412* long terminal repeat. J. Mol. Biol.

71. Flavell, A.J. (1984). Role of reverse transcription in the generation of extrachromosomal *copia* mobile genetic elements. Nature (London) *310*, 514-516.

72. Flavell, A.J., and Brierley, C. (1986). The termini of extrachromosomal linear *copia* elements. Nucleic Acids Res. *14*, 3659-3669.

73. Arkhipova, I.R., Gorelova, T.V., Ilyin, Y.V., and Schuppe, N.G. (1984). Reverse transcription of *Drosophila* mobile dispersed genetic element RNAs: detection of intermediate forms. Nucl. Acids. Res. *12*, 7533-7548.

74. Shoemaker, C., Goff, S., Gilboa, E., Paskind, M., Mitra, S.W., and Baltimore, D. (1981). Structure of a cloned circular Moloney murine leukemia virus molecule containing an inverted segment: implications for retrovirus integration. Proc. Natl. Acad. Sci. (USA) *72*, 3932-3936.

75. Van Beveren, C., Rands, E. Chattopadhyay, S.K., Lowy, D.R., and Verma, L.M. (1982). Long terminal repeat of murine retroviral DNAs: sequence analysis, host-proviral junctions, and pre-integration site. J. Virol. *41*, 542-556.

76. Stanfield, S., and Lengyel, J.A. (1979). Small circular DNA of *Drosophila melanogaster*: chromosomal homology and kinetic complexity. Proc. Natl. Acad. Sci. (USA) *76*, 6142-6146.

92. Zachar, Z., Davison, D., Garza, D., and Bingham, P.M. (1985). A detailed developmental and structural study of the transcriptional effects of insertion of the *copia* transposon into the *white* locus of *Drosophila melanogaster*. Genetics.

93. Ciriacy, M., and Williamson, V.M. (1981). Analysis of mutations affecting Ty-mediated gene expression in *Saccharomyces cerevisiae*. Mol. Gen. Genet. *182*, 159-163.

94. Dubois, E., Jacobs, E., and Jauniaux, J-C. (1982). Expression of RoAM mutations in *Saccharomyces cerevisiae*: involvement of trans-acting regulatory elements and relation to Ty 1 transcription. Embo J., *1*, 1133-1139.

95. Winson, F., CHaleff, D.T., Valent, B., and Fink, G.R. (1984a). Mutations affecting Ty-mediated expression of the HS4 gene of *Saccharomyces cerevisiae*. Genetics *107*, 179-197.

96. Winson, F., Durbin, K.J., and Fink, G.R. (1984b). The *SPT3* gene is required for normal transcription of Ty elements in *S. cerevisiae*. Cell *39*, 675-682.

97. Boeke, J.D., Styles, C.A., and Fink, G.R. (1986). *Saccharomyces cerevisiae SPT3* gene is required for transposition and transpositional recombination of chromosomal Ty elements. Mol. Cel. Biol. *6*, 3575-3581.

98. Varmus, H.E., Guntaka, R.V., Fan, W.J., Heasly, S., and Bishop, J.M. (1974). Synthesis of viral DNA in the cytoplasm of embryo fibroblasts and in enucleated cells after infection by avian sarcoma virus. Proc. Natl. Acad. Sci. USA *71*, 3874-3878.

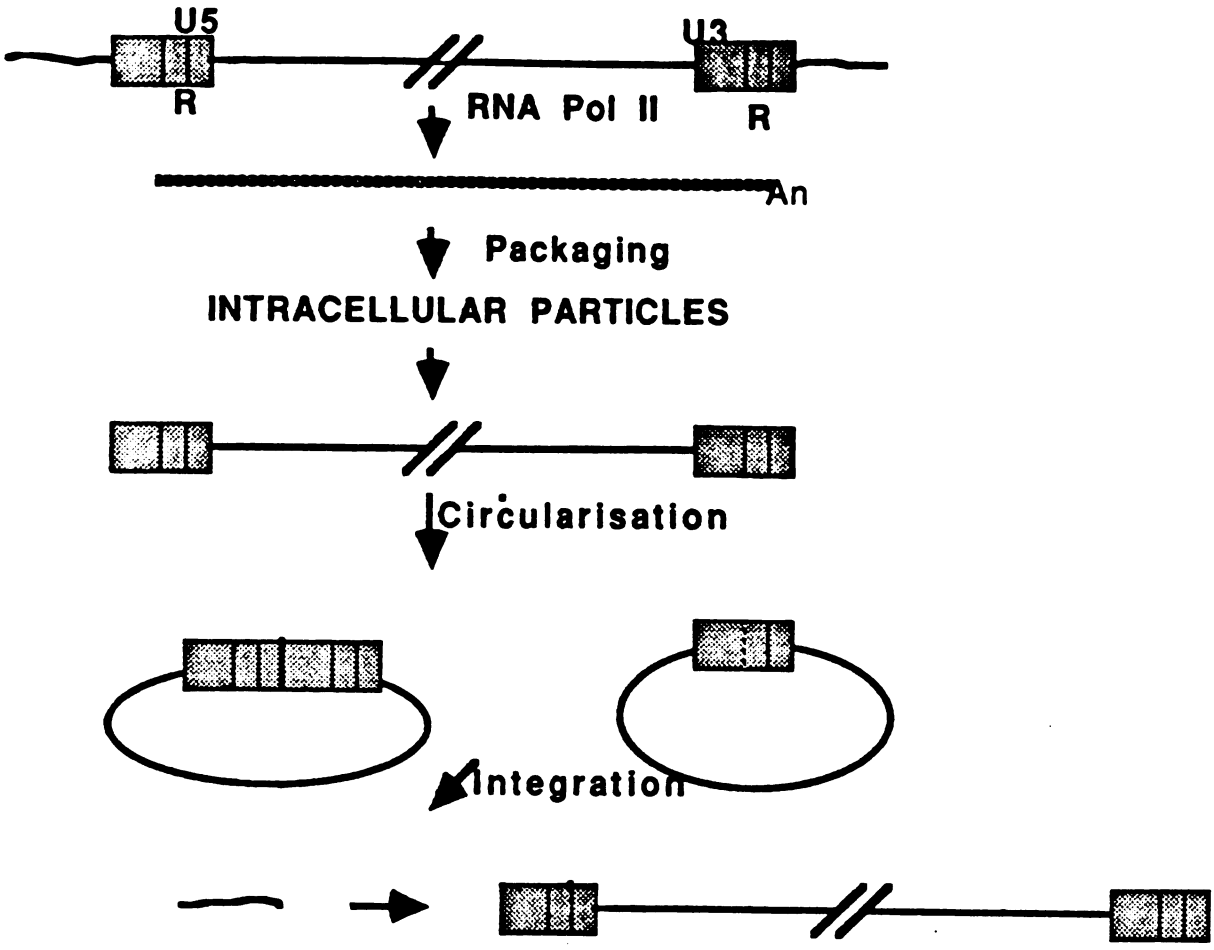
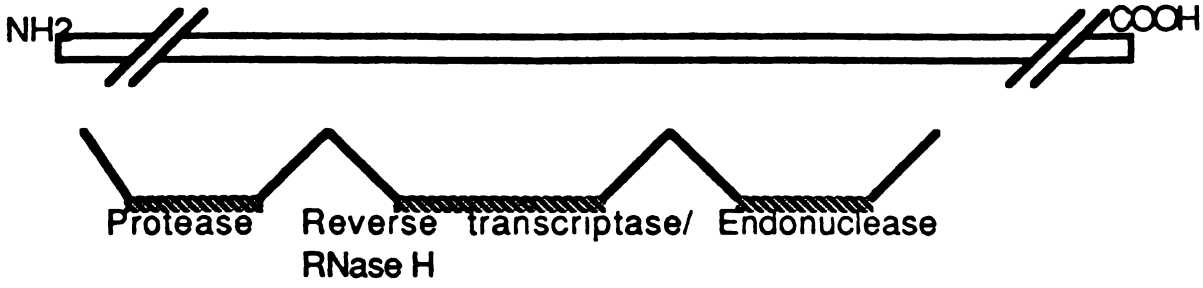


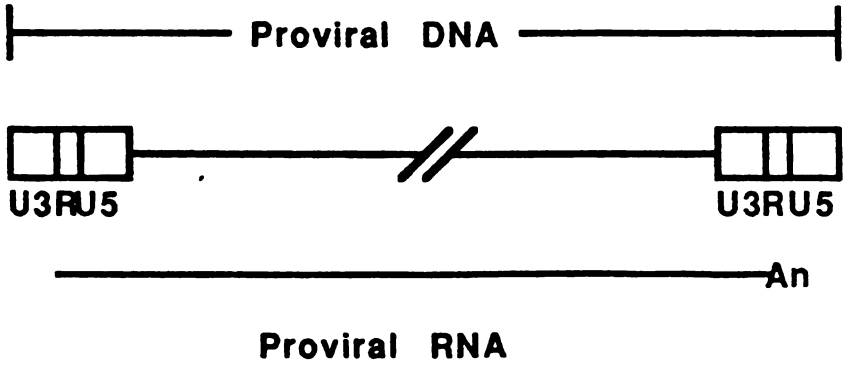
Figure 1: General pathway for the transposition of retrotransposons in Drosophila and yeast.

Figure 2 : Arrangement of the polymerase domains in retroviral and retrotransposon genes.



Regions of Homology			
Retroviruses	+	+	+
297,412, 17.6,gypsy	+	+	+
copia,Ty	+	+	+
CaMV	+	+	
HBV		+	

Figure 3 : Relationship of retroviral RNA to provial DNA forms



CHAPTER II

EXTRACHROMOSOMAL DNA FORMS OF COPIA-LIKE TRANSPOSABLE ELEMENTS, F ELEMENTS AND MIDDLE REPETITIVE DNA SEQUENCES IN DROSOPHILA MELANOGASTER : VARIATION IN CULTURED CELLS AND EMBRYOS

(Reprinted from J.Mol.Biol.182, 31-43 (1985))

CHAPTER II

EXTRACHROMOSOMAL DNA FORMS OF COPIA-LIKE TRANSPOSABLE ELEMENTS, F ELEMENTS AND MIDDLE REPETITIVE DNA SEQUENCES IN DROSOPHILA MELANOGASTER : VARIATION IN CULTURED CELLS AND EMBRYOS

(Reprinted from J.Mol.Biol.182, 31-43 (1985))

J. Mol. Biol. (1985) 182, 31-43

Extrachromosomal DNA Forms of *copia*-like Transposable Elements, F Elements and Middle Repetitive DNA Sequences in *Drosophila melanogaster*

Variation in Cultured Cells and Embryos

Kevin G. Mossie[1], Michael W. Young[2] and Harold E. Varmus[1,3]

[1]*Division of Genetics, Department of Biochemistry and Biophysics, and the*
[3]*Department of Microbiology and Immunology*
University of California, San Francisco, Calif. 94143, and
[2]*Rockefeller University, New York, N.Y. 10021, U.S.A.*

(Received 25 June 1984 and in revised form 12 October 1984)

Drosophila melanogaster embryos and cells in culture were screened for the presence of unintegrated covalently closed circular DNA forms that hybridize to *copia*-like transposable elements, the F element and uncharacterized dispersed middle repetitive DNA elements. Our results indicate that the majority of *copia*-like elements (including *copia*, 297, 412, *mdg*1, *mdg*3 and *gypsy*), the F elements, and 9 of 12 middle repetitive DNA elements are present as free DNA forms in cultured cells and embryos. An 18 base-pair inverted repeat has been reported to flank the long direct repeat of *mdg*3, implying that *mdg*3 is not an orthodox *copia*-like element; however, we have sequenced two independently isolated *mdg*3 clones and shown that the inverted repeat is not part of the element.

The relative abundance with which free DNA forms are found varies between the cultured cells used, and between cultured cells and embryos. This variation, which can be up to 20-fold for some elements, does not correlate well with either the amount of element-specific poly(A)$^+$ RNA present per cell or the number of element-specific sequences integrated in the genome.

1. Introduction

Many transposable elements in *Drosophila melanogaster* belong to the *copia*-like class of dispersed middle repetitive DNA elements (Young & Schwartz, 1980; Spradling & Rubin, 1981; Rubin, 1983). Transposable elements in this family are usually long (5 to 9 kb†) and terminated with direct repeats of some 200 to 500 bp (Spradling & Rubin, 1981). Structurally, they resemble members of the Tn*9* class of bacterial transposons (Kleckner, 1981), the Ty-1 element in yeast (Cameron *et al.*, 1979) and the integrated forms of retroviral proviruses (Varmus, 1983). The other structural classes of transposable elements in *Drosophila* include those terminating in very short inverted repeats, e.g. P elements and *hobo* (O'Hare & Rubin, 1983; McGinnis & Beckendorf, 1983), repetitious inverted repeats (Potter *et al.*, 1980; Truett *et al.*, 1981), and a class that lacks terminal repeats, e.g. the F family (Dawid *et al.*, 1981; DiNocera *et al.*, 1983). While the manner in which these classes of transposable elements proliferate in the genome is unknown, structural differences between these elements suggest that different mechanisms may be employed.

Structural similarities between the *copia*-like transposable elements of *Drosophila* and integrated forms of vertebrate retrovirus proviruses (Levis *et al.*, 1980; Varmus, 1983; Spradling & Rubin, 1981; Will *et al.*, 1982) have led to speculations that they may be evolutionarily related (Temin, 1980). The notion that *copia*-like transposable elements could transpose *via* an RNA intermediate has been strengthened from two lines of investigation: (1) virus-like particles with reverse transcriptase-like activity have been detected in *Drosophila* cultured cells (Heine *et al.*, 1980) and contain RNA with sequences homologous to the *copia* element (Shiba & Siago, 1983). (2) Cultured cells contain covalently closed circular and unintegrated linear DNA molecules that anneal to *copia*-specific probes (Flavell & Ish-Horowicz, 1981, 1983; Flavell, 1984).

† Abbreviations used: kb, 10^3 base-pairs; bp, base-pairs; LTR, long terminal repeat.

0022-2836/85/050031-13 $03.00/0

The circles, containing one or two terminal repeats, and the linear DNA resemble unintegrated retroviral DNA (see Varmus & Swanstrom, 1982), prompting the suggestion that they may be intermediates in the transposition cycle (Flavell & Ish-Horowicz, 1981, 1983; Flavell, 1984).

A different structural class of mobile elements, also ubiquitous in eukaryotes, has also been proposed to transpose *via* a reverse transcriptase-mediated pathway (see Rogers, 1983). This class of elements, which includes the F family in *Drosophila*, differs from the *copia*-like elements in that the members lack terminal repeated sequences at the ends. Members of this class terminate at their 3′ end with oligo-adenylated sequences (7 to 20 A residues) and encode mRNAs that are not only heterogeneous in size but also appear to be enriched in the poly(A)$^-$ fraction (Dawid *et al.*, 1981; Kole *et al.*, 1983; Shafit-Zagardo *et al.*, 1983). It is believed, though not proven, that transposition of these elements occurs by reverse transcription of element-specific transcripts to form cDNA copies, which then integrate elsewhere in the genome (Van Arsdell *et al.*, 1981; Sharp, 1983).

In this paper, we present evidence showing that members belonging to the *copia*-like and F-like structural classes are present as free DNA circular forms in *Drosophila* cultured cells and embryos. Similar free DNA forms hybridized to other uncharacterized dispersed middle repetitive DNA elements. The relative abundance of free DNA varies among different elements and among the cells used. This variation does not correlate with either the amount of element-specific poly(A)$^+$ RNA present per cell or the number of element-specific sequences integrated in the genome.

2. Materials and Methods

(a) D. melanogaster *embryos and tissue culture cells*

Embryos (0 to 24 h old) were collected from wild-type Oregon R strain of *D. melanogaster* and were a gift from T. Kornberg (UCSF). The two established *D. melanogaster* tissue culture lines, K_{c0} (Echalier & Ohanessian, 1970) and the Schneider line 2 (Schneider, 1972), were obtained from A. Blumenthal (UCSF).

(b) *Media and growth conditions*

Flies were grown on standard cornmeal–agar medium at 22°C supplemented with yeast paste. Embryos were collected by allowing flies in population cages to lay eggs, washed with 0·7% NaCl, 0·4% Triton X-10 (Triton/NaCl), dechorionated for 90 s with 50% (v/v) Chlorox, washed again with Triton/NaCl, and either used directly or quick-frozen immediately in liquid nitrogen before storage at −70°C. K_{c0} cells were grown as monolayer cultures on Echalier medium (Gibco) at 24°C in T-flasks, or as suspension cultures in 200 to 500 ml side-arm flasks, with mild stirring at 22°C. Schneider line 2 cells were grown as monolayer cultures on Schneider medium (Gibco), supplemented with 0·5% bacteriological peptone (Gibco) and 15% heat-inactivated fetal calf serum (Gibco) in T-flasks at 24°C. The pH of the medium was adjusted to 6·8 with 1 M-KOH before addition of the serum. Both media contained 100 units of penicillin streptomycin/ml.

(c) *DNA clones and subclones*

A description of DNA clones used as hybridization probes is presented in Table 1. Genomic clones containing *mdg*3 sequences were isolated from a library of *D. melanogaster* Canton S DNA in bacteriophage Charon 4A (Maniatis *et al.*, 1978) by screening with a probe containing *mdg*3 terminal repeat sequences as described

Table 1

Catalogue of DNA clones

DNA clone	Description and comment	Source/Reference
cDm2042	Oregon R DNA in pBR322; contains entire 412 element	G. M. Rubin (unpublished)
cDm4006	Oregon R DNA in pBR322; contains entire 297 element	G. M. Rubin (unpublished)
cDm5002	Oregon R DNA in pBR322; contains entire *copia* element	G. M. Rubin; Dunsmuir *et al.* (1980)
λbDm2030	Oregon R DNA in λSep6; contains entire *roo* element	D. Meyerowitz; Meyerowitz & Hogness (1982)
λbx^{34e}	Oregon R DNA in λSep6; contains entire *gypsy* element	W. Bender (unpublished)
λDm2/3 LCP1	Canton S DNA in λL47; contains entire HMS Beagle element	N. Davidson; Snyder *et al.* (1982)
Dm58	Canton S DNA in λL47; contains entire *mdg*1 element	G. Georgiev; Georgiev *et al.* (1980)
Dm86	Canton S DNA in λL47; contains entire *mdg*3 element	G. Georgiev; Georgiev *et al.* (1980)
pDm27	Oregon R DNA in pBR325	M. Young, Dowsett & Young (1982)
pDm34	Oregon R DNA in pBR325	M. Young; Dowsett & Young (1982)
pDm39	Oregon R DNA in pBR325	M. Young, Dowsett & Young (1982)
pDm54	Oregon R DNA in pBR325	M. Young, Dowsett & Young (1982)
pDm73	Oregon R DNA in pBR325	M. Young; Dowsett & Young (1982)
pDm151	Oregon R DNA in pBR325	M. Young, Dowsett & Young (1982)
pDm185	Oregon R DNA in pBR325	M. Young; Dowsett & Young (1982)
pDm298	Oregon R DNA in pBR325	M. Young; Dowsett & Young (1982)
pDm321	Oregon R DNA in pBR325	M. Young; Dowsett & Young (1982)
cDm2028	Oregon R DNA in Col E1 plasmid	M. Young; Young (1979)
cDm2041	Oregon R DNA in Col E1 plasmid	M. Young; Young (1979)
cDm2067	Oregon R DNA in Col E1 plasmid	M. Young; Young (1979)
p125	5·9 kb *Eco*RI fragment from phage A19 containing the 3′ end of F element in pBR322	I. Dawid; DiNocera *et al.* (1983)
p127	4·3 kb *Eco*RI fragment from phage A19; contains the 5′ end of F element in pBR322	I. Dawid; DiNocera *et al.* (1983)

by Benton & Davis (1977). Restriction fragments were subcloned in pUC8 (Vieira & Messing, 1982).

(d) *Enzymes and isotopes*

Restriction enzymes were obtained from New England Biolabs Inc. (Beverly, MA) and Bethesda Research Laboratories (Bethesda, MD). Reaction conditions and buffers were as recommended by the manufacturers. DNA polymerase I (Kornberg enzyme) and the Klenow fragment were purchased from Boehringer (Mannheim, GmbH). DNase I was from Sigma (St Louis, MO). ^{32}P-labeled deoxynucleotice triphosphates (dNTPs) were purchased from the Radiochemical Centre (Amersham, England). Labeled DNA probes were prepared according to the nick-translation method described by Rigby *et al.* (1977).

(e) *DNA sequencing*

Restriction fragments containing *mdg*3 terminal repeats were labeled at their 3′ ends using the Klenow fragment of DNA polymerase I and [α-^{32}P]dNTPS, or at their 5′ ends using phage T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]dATP as described by Maxam & Gilbert (1980). DNA sequencing by partial chemical cleavage was as described (Maxam & Gilbert, 1980) except that precipitation with ethanol rather than lyophilization was used to remove piperidine (Smith & Calvo, 1980).

(f) *Isolation of nucleic acids*

Chromosomal DNA was prepared from cultured cells essentially as described by Stanfield & Helinski (1979). Embryonic chromosomal DNA was prepared according to McGinnis & Beckendorf (1983). Covalently closed circular DNA molecules were isolated from cultured cells and embryos using a procedure initially developed by Casse *et al.* (1979) and modified by Devenish & Newlon (1982). Covalently closed circular DNA molecules were purified by ethidium bromide/CsCl equilibrium centrifugation (Radloff *et al.*, 1968) as follows: 18 g solid CsCl (Beckman Instruments Inc., CA) was dissolved in 17 g TE buffer (10 mM-Tris·HCl (pH 8·0), 1 mM-EDTA) containing the nucleic acids, and ethidium bromide was added to 300 mg/ml. Contents were mixed, topped with light mineral oil, and centrifuged to equilibrium in a Ti 42.1 rotor at 33,500 revs/min for 60 to 70 h at 15°C. Fractions were collected and DNA samples were analyzed on agarose gels. The supercoiled DNA peak was pooled, ethidium bromide was removed by extraction with NaCl-saturated isopropanol, and the DNA was precipitated with ethanol. Samples were stored at −20°C until use. Poly(A)$^+$ RNA was prepared from cultured cells and eluted on oligo(dT) columns as described by Varmus *et al.* (1981).

(g) *Analysis of nucleic acid preparations*

Both circular DNA and restriction enzyme-digested chromosomal DNA preparations were fractionated on agarose gels in either 0·02 M-Tris-acetate or 0·04 M-Tris-acetate containing 0·005 M-EDTA. The size(s) of unintegrated DNAs was estimated by comparing the migration of circular plasmid DNA markers relative to linear phage DNA markers fractionated under similar conditions. DNA was transferred to nitrocellulose filters as described by Southern (1975). RNA samples were denatured, fractionated on formaldehyde-containing agarose gels and transferred to nitrocellulose filters as described by Nusse & Varmus (1982) and Thomas (1980). Filters with transferred DNA or RNA species were baked *in vacuo* at 80°C for 2 to 3 h, and prehybridized in a mixture of 50% formamide, 3 × SSC (SSC is 0·15 M-NaCl, 0·015 M-sodium citrate), 1 × Denhardt's (1966) solution, 50 mM-HEPES (pH 7·4), 150 μg yeast tRNA/ml, and 200 μg sheared salmon sperm DNA/ml for 4 to 6 h at 41°C. Filters were hybridized for 48 h at 41°C in the same mixture containing 10^6 to 4×10^6 cts/min per ml. The filters were washed in 0·1 × SSC, 0·1% sodium dodecyl sulfate at 50°C, air dried and exposed to X-ray film. To re-use the blot, filters were washed in 2 changes of 0·1 M-NaOH for 5 min, rinsed thoroughly in distilled water, then washed at room temperature in 2 × SSC for 5 min.

3. Results

(a) *Screening* copia*-like transposable elements for free DNA forms in cultured* Drosophila *cells*

Both the K_{co} and Schneider 2 *Drosophila* cell lines were used to prepare extrachromosomal forms of DNA using a modification of the procedure described by Devenish & Newlon (1982). This procedure greatly enriches for covalently closed circular DNA molecules since both linear DNA and nicked circular DNA molecules are removed during extraction with phenol in high salt solution. Nucleic acid from approximately 5×10^7 cells was used to seek unintegrated forms of several cloned transposable elements. To distinguish the elements from each other and from previously uncharacterized middle repetitive DNA clones, one microgram of *Hha*I-digested Oregon R chromosomal DNA was loaded next to lanes containing partially purified, uncut free DNA. (The restriction enzyme *Hha*I cuts many times in the *Drosophila* genome and, since members of each *copia*-like class of elements are fairly homogeneous, each family will have distinct fragments of *Hha*I-cut DNA.) Typical results of these analyses are shown in Figure 1. Probes prepared from five of the eight tested *copia*-like elements clearly detected unintegrated DNA species in K_{co} cells (lane 1, (a) to (e)). The sizes of the free DNA molecules, presumed to be covalently closed circles, correspond approximately to the sizes of integrated genomic forms (see Rubin, 1983, Table 2).

We were not successful in demonstrating the presence of free DNA forms using *gypsy*, HMS Beagle and *roo* transposable elements in these cells, indicating that less than 0·01 copy is present per cell. However, as will be described in later sections of this paper, *gypsy* is present as unintegrated DNA forms in embryos (see Fig. 6(d)). The number of unintegrated copies per cell also appears to vary among the elements, and between different preparations (data not shown), with optimum yields of one copy in 20 cells for 297 to almost a copy per cell in K_{co} cells for 412 (see Table 2; Figs 5 and 6). With some of the elements, notably 297 and *mdg*1 (Fig. 1(b) and (d)), two species of similar mobilities are often observed. Although these have not been analyzed structurally, we believe that they may be similar to those described for *copia* (Flavell & Ish-

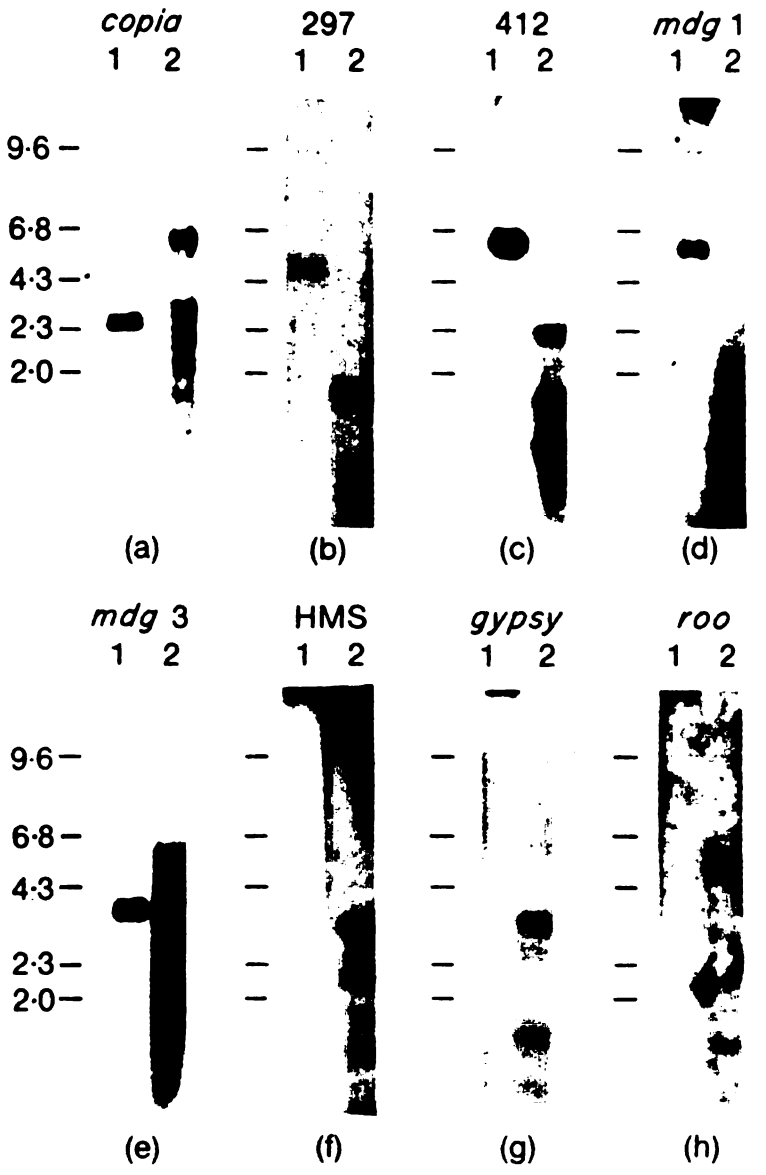


Figure 1. Analysis of free DNA preparations for unintegrated forms of *copia*-like elements. Samples containing free DNA were prepared from K_{co} cells, fractionated on agarose gels in 0·02 M-Tris-acetate, 0·005 M-EDTA (TAE) alongside *Hha*I-digested Oregon R chromosomal DNA, transferred to nitrocellulose and hybridized with ^{32}P-labeled probes containing the indicated *copia*-like transposable elements. For tests with each element, lane 1 contains undigested free DNA samples from 5×10^7 to 8×10^7 cells, and lane 2 contains 1 μg of *Hha*I-digested Oregon R DNA. Linear size markers are shown alongside in kb.

Horowicz, 1981, 1983), and thus may contain one or two copies of the LTR.

(b) *Structure of the* mdg3 *transposable element*

The finding of free circular forms of *mdg*3 in both K_{co} (Fig. 1(e)) and Schneider cells (not shown) was unexpected. The structure of *mdg*3 (Bayev *et al.*, 1980) has been thought to differ from that of other *copia*-like elements in that 18 bp imperfect inverted repeats at the ends of one cloned *mdg*3 element were not part of the 268 bp direct repeat but were considered to be part of the element. It is also possible, however, that the ends of the element were not correctly assigned in the single sequenced sample (see Table 1 of Rubin, 1983). To examine this assignment further, we have sequenced the termini of additional isolates of *mdg*3 clones. We reasoned that the correct ends of *mdg*3 repeats would be defined by divergence in DNA sequence between the clones we have sequenced and the published *mdg*3 repeat sequence. Labeled fragments containing *mdg*3 repeats were used as hybridization probes to isolate phage clones from a genomic DNA library from the Canton S strain. Two clones were chosen, and *Eco*RI fragments containing the repeat sequences were subcloned into pUC8 for sequencing.

Figure 2(a) shows the restriction map of these

Table 2

Characterization of transposable element DNA and RNA

	Copies per cell of circular transposable element DNA[a] (and integrated transposable element DNA[b])						Poly(A)$^+$ RNA (approx. $^0_0$ per cell)[f]			
	K_{co}		Sch. L2		Embryos		Approx. size (kb)[d]	K_{co}	Sch. L2	Predominant transcripts (kb)[g]
copia	0·1	(90)[b]	0·03	(170)[b]	1–2	(60)[b]	(4–6)[e]	3–4	3–4	5·2; 2·1
297	0·02	(110)[b]	0·03	(170)[b]	1–2	(30)[b]	(6–7)[e]	0·1	0·3	~7
412	0·7	(120)[b]	0·1	(40)[b]	0·20	(40)[b]	(7–8)	1	0·08	~7
*mdg*1	0·1	(245)[b]	0·2	(245)[b]	0·1	(35)[b]	(7–8)[e]	1	1	~7; numerous
*mdg*3	0·15	(236)[b]	0·2	(236)[b]	0·2	(17)[b]	(5–6)	1	0·15	5
gypsy	<0·01		<0·01		0·06	(10)[b]	(7–8)[e]	0·1	0·3	8
roo (B104)	<0·01	(79)	<0·01		N.T.[c]	(95)[b]	—	<0·01	<0·01	N.T.[c]
HMS Beagle	<0·01		<0·01		N.T.[c]		—	<0·01	<0·01	N.T.[c]
pDm27	<0·01		0·2		<0·01	(21)[b]	(5–6)	0·05	<0·01	N.T.[c]
pDm34	<0·01		0·03		N.T.[c]	(10)[b]	(5–6)	<0·01	<0·01	~7
pDm39	0·4		0·1		0·20	(34)[b]	(7–8)	1	0·08	~7
pDm54	<0·01		0·03		N.T.[c]	(59)[b]	(5–6)	0·16	0·24	Numerous
pDm73	<0·01		<0·01		N.T.[c]	(18)[b]	—	0·08	0·32	N.T.[c]
pDm151	<0·01		0·1		0·06	(41)[b]	(5–6)	<0·01	<0·01	Heterogeneous
pDm185	<0·01		0·1		N.T.[c]	(41)[b]	(4–5)	<0·01	<0·01	N.T.[c]
pDm298	<0·01		0·1		N.T.[c]	(20)[c]	(5–6)	0·04	0·04	N.T.[c]
pDm321	<0·01		<0·01		N.T.[c]	(20)[b]	—	<0·01	<0·01	N.T.[c]
cDm2028	<0·01		0·08		N.T.[c]	(10)[b]	(7–8)	0·04	0·04	N.T.[c]
cDm2041	0·1		0·15		N.T.[c]	(59)[b]	(5–6)	0·04	0·04	N.T.[c]
cDm2067	0·06		0·06		N.T.[c]	(205)[b]	(5–6)	<0·01	<0·01	N.T.[c]
F element	<0·01		0·01		0·05	(50)[b]	(5–6)	0·03	0·03	N.T.[c]

[a] Estimated by comparison with known concentrations of marker plasmids. Sch., Schneider.

[b] Published numbers of element-specific sequences integrated in the genome (see Potter *et al.*, 1979; Tchurikov *et al.*, 1980; Young, 1979; Dowsett & Young, 1982; DiNocera *et al.*, 1983; Scherer *et al.*, 1982).

[c] Not tested.

[d] Size(s) calculated from circular plasmid DNA markers and linear phage DNA markers.

[e] Instances where a double band was observed in the position of closed circular DNA.

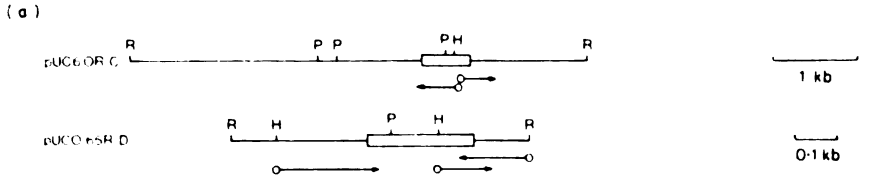
[f] Estimated by RNA dot-blot analyses and values normalized with respect to published values for actin (1^0_0) and *copia* (3 to 4^0_0).

[g] Examined by Northern gel analyses as in Fig. 7.

subclones and the sequencing strategy used to identify the termini. The DNA sequence of the *mdg*3 repeat termini is shown in Figure 2(b), and is compared to the published sequence (Bayev *et al.*, 1980). The first subclone pUC6.OR/C contains the left half of the *mdg*3 transposable element plus 5′ flanking cellular DNA. From this we conclude that the 5′ end of the *mdg*3 terminal repeat should read 5′-T-T-T-A-G-C-C . . . 3′ (Fig. 2(b)). This conclusion is confirmed by examining the termini of the repeat in subclone pUO.65R/D, which appears to contain a solo terminal repeat with the ends flanked by non-*mdg*3 sequences. Thus the sequence of the clone permits definition of the 3′ end of the elements as well (Fig. 2(b)). It should be noted that the 18 bp inverted repeat present in Dm86 is absent in both clones sequenced. We conclude therefore that this sequence is not part of the *mdg*3 LTR. We believe that the correct 3′ end assignment is . . . A-C-T-A-C-A instead of . . . A-C-T-A-C-A-A for the following reasons: (1) this produces a more extensive inverted terminal repeat (4/5) at ends of *mdg*3 and would also make the terminal penta-nucleotides similar to those of *mdg*1 and other *copia*-like elements (see Fig. 2(c)). (2) *mdg*3, like *mdg*1, would then appear to generate a 4 bp (rather than an unusual 3 bp) duplication of target sequence. It is interesting that a 4 bp repeat A-T-T-A flanks what we propose to be the ends of *mdg*3 in Dm86 (Fig. 2(b)). We believe, therefore, that there has been a point mutation in a 4 bp target duplication in pUCO.65R/D at either of the positions shown in Figure 2(b). We have not, however, sequenced an "empty" site to establish this.

(c) *Analysis of uncharacterized repetitive DNA elements*

One of us (M. W. Young) previously isolated a collection of dispersed middle repetitive elements from a genomic library of *D. melanogaster*, some of which were shown to be transposable (Young, 1979; Dowsett & Young, 1982). Although none of these elements has been characterized structurally, we decided to test several for extrachromosomal DNA forms using the same methods employed for the defined elements. *Hha*I digest fragments were used to compare these elements to each other and to known *copia*-like elements. Again, as with *copia*-like elements, most of the elements tested (9/12) had unintegrated DNA forms in *Drosophila* cultured cells (Fig. 3 and Table 2). The number of element-specific free DNA forms varied from element to element. Examination of *Hha*I digest patterns confirmed *in situ* hybridization data (Young, 1979;



(b) *Nucleotide sequence*

Dm 86 (Bayev *et al.*, 1980)

Left LTR

ATAGAAGTGTATTTTTAAGCTATTATTTAGCC ... CTACTACAACTCGGCCGTCCTGACAACTAGATCTC

Right LTR

GATTTTATAATCCCACTTCTGAGATTTTAGCC ... CTACTACAATTAATAGCTCTGAATCAGAATCAGAAAGC

pUC6.OR C

TTTCGGACACACAAATAAAATAAAGTTTAGCC ... CTACTACAACTCGGCCGTCCTGACAACTAGATATC

pUC0.65R/D

ACACTATATGCGCCCTAGCTGGAAGTTTAGCC ..... CTACTACAAAAGTATACACTCTCTCGGCTATATGAA

A G

(c) *Similarities with ends of structurally related elements*

Element	*5′ end of LTR*	*3′ end of LTR*
Mdg 1	TGTAGT	ACTACA
Mdg 3	TTTAGT	ACTACA
297	GTGACGTATT	AATTTTACA
412	TGTACT	ATTACA
copia	TGTTGGAATA	TTACAACA
Ty-1	TGTTGGAATA	ATTTCTCA
B104	TGTTCA	ATTTTACA

Figure 2. DNA sequence analysis of the ends of the *mdg*3 terminal repeats. Phage clones containing *mdg*3 terminal repeats were isolated from the *Drosophila* library and *Eco*RI fragments containing the terminal repeats subcloned into pUC8 as described in Materials and Methods. Hyphens have been omitted from sequences for clarity. (a) Restriction maps and sequencing strategies used to define the ends of *mdg*3 repeats in subclones pUC6.OR/C (top) and pUC0.65R/D (bottom). The restriction sites shown are *Eco*RI (R), *Hin*dIII (H), and *Pst*I (P). (b) Nucleotide sequence of the ends of *mdg*3 repeats in subclones. The LTR sequences are aligned with those published for Dm86 (Bayev *et al.*, 1980) and are heavily underlined. Only the terminal nucleotides are shown and the internal LTR sequences are represented as dots. In Dm86, shown in bold face letters are the 5 bp (A-G-A-A-G) of target sequences that have been proposed to be duplicated (Bayev *et al.*, 1980). These bound the 18 bp palindrome (lightly underlined), which has been proposed to be part of the element. Note that this sequence is missing from the left end in both of our subclones, and from the right end in pUC0.65R/D. In Dm86, immediately flanking what we propose to be the real LTR termini, is a 4 bp duplication A-T-T-A. pUC0.65R/D appears to contain a sole LTR and to have suffered a mutation in the duplicated 4 bp sequence (either G → A on the left or A → G on the right). (c) Comparison of *mdg*3 LTR termini to termini of other *copia*-like elements.

Dowsett & Young, 1982) that these elements were different from each other.

One of the elements tested, pDm39 (Fig. 3(c)), had restriction digest patterns that were very similar to those seen with 412 (Fig. 1(c)). Homology between these two clones was confirmed by cross-hybridization studies (data not shown).

(d) *Unintegrated DNA forms annealing to F elements*

F elements belong to a structural class proposed to transpose *via* a reverse transcriptase-mediated pathway (DiNocera *et al.*, 1983; Sharp, 1983). We therefore decided to test F elements for the presence of free DNA forms in these cells. Labeled probes prepared from both the 5′ half (p127) and poly(dA)-containing 3′ end (p125) were hybridized to filters containing unintegrated DNAs from Schneider cells. It is clear from Figure 3 (i) and (j) that both probes hybridize to extrachromosal DNA of the same size. pDm151 also hybridizes to free DNA of approximately the same size (Fig. 3(d)), and to a clone (p125) containing the 3′ end of an F element (data not shown). From the *Hha*I digest patterns, it is clear that pDm151, p125 and p127 share similar-sized *Hha*I fragments. We believe, therefore, that pDm151 is a truncated F element. The nature of F-specific circles is currently under investigation.

(e) *Unintegrated forms in* Drosophila *embryos*

The genomic copy number of some *copia*-like elements is amplified three- to fourfold in tissue culture cells compared to embryos (Potter *et al.*, 1979; Tchurikov *et al.*, 1981). This raises the question of whether the circular DNAs we have

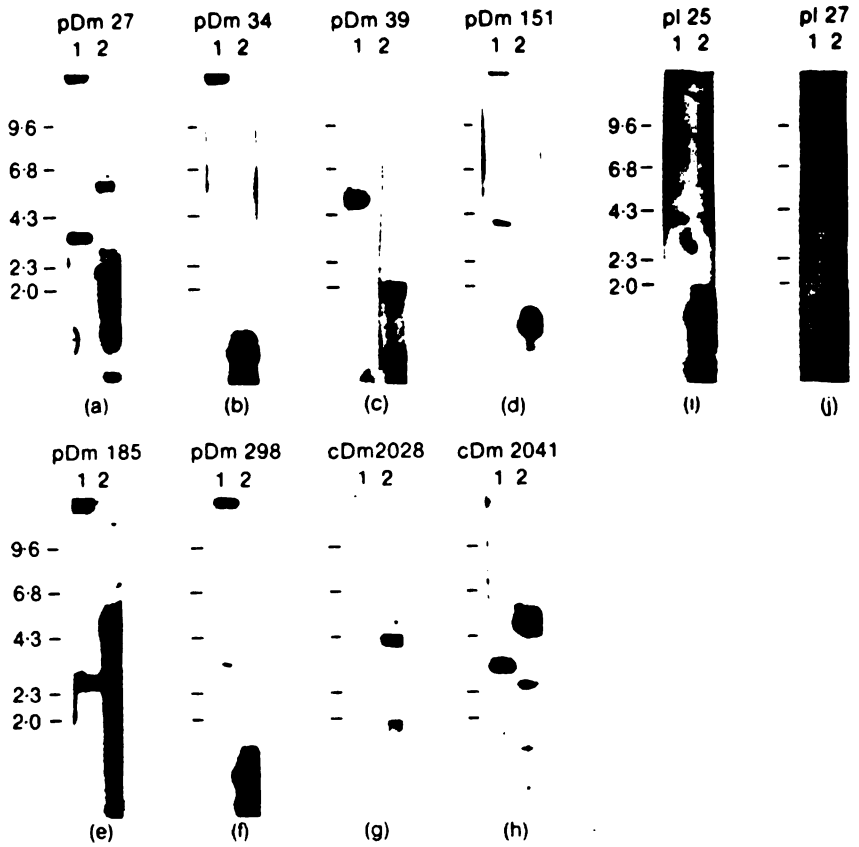


Figure 3. Analysis of free DNA preparations from cultured cells using middle repetitive DNA elements as labeled probes. Samples containing free DNA were prepared from either K_{co} or Schneider cells and fractionated on agarose gels in 0·04 M-Tris-acetate, 0·005 M-EDTA (except (e), which was fractionated in 0·02 M-Tris-acetate, 0·005 M-EDTA). DNA was then transferred to nitrocellulose filters and hybridized with ^{32}P-labeled probes containing middle repetitive DNA elements. For a description of lanes see Fig. 1.

observed are dependent upon this amplification process or are for some other reason peculiar to cultured cells. We therefore examined *D. melanogaster* embryos harvested en masse 0 to 24 hours after fertilization for circular DNA forms. Nine of ten probes tested were homologous to unintegrated DNA (Table 2). Figure 4 shows unintegrated DNA forms specific for *copia*, *mdg*1 and pDm39 elements. The following points can be made from these results: (1) *copia*-like and middle repetitive DNA elements are present as extrachromosomal DNA forms in embryos, (2) *copia*-specific and 297-specific circles are much more abundant (10 to 20-fold) in embryos than in cultured cells and considerably more abundant than other elements (e.g. 412) in embryos: (3) *copia*, *mdg*1 and pDm39-specific circles are present as two similar but separable forms, again presumed to be circles containing one or two LTRs (Fig. 1(b) and (d); and Flavell & Ish-Horowicz, 1981, 1983).

A longer autoradiographic exposure of the experiment showing *copia*-free forms in embryos (Fig. 4(a)[1]) reveals a number of additional extrachromosomal species. We have occasionally observed similar forms of various elements in cultured cells (unpublished results). These may represent concatemeric forms, but have not been investigated further.

(f) *Variations in the abundance of circular DNA forms*

(i) *Comparison of cell lines*

Our results with embryos suggested that different cells may contain different amounts of circular forms of certain elements. To examine this possibility more closely, we compared Hirt supernatant fractions from two cultured cell lines, K_{co} (Echalier & Ohannesian, 1969, 1970) and Schneider line 2 (Schneider, 1972), for free forms of several elements. These two cell lines were earlier shown to contain different amounts of integrated copies of *copia*, 297 and 412 (Potter *et al.*, 1979). Furthermore, while the origin of the Schneider line is clearly Oregon R (see Potter *et al.*, 1979), the origin of the K_{co} line is uncertain, and it is possible that these lines were derived from different *Drosophila* strains.

We find that some elements are abundantly represented in unintegrated DNA in one cell line but not the other, even when cells are grown to late stationary phase to minimize possible effects of

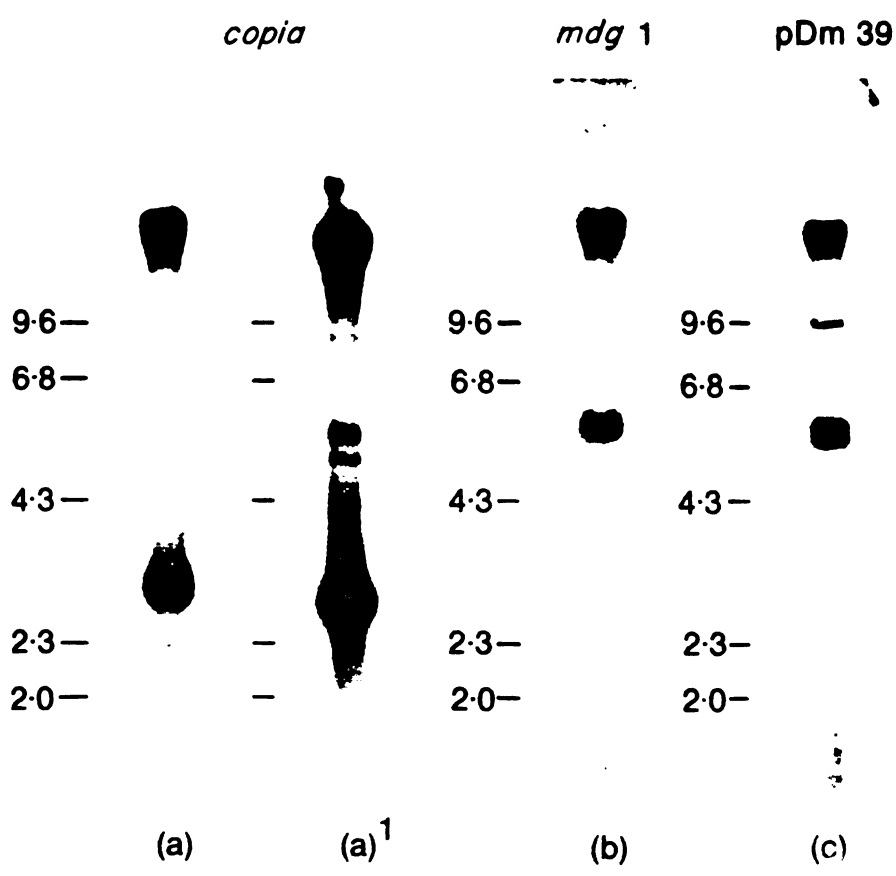


Figure 4. Analysis of unintegrated DNAs in embryos. Circular DNA was prepared from Oregon R embryos purified on CsCl/ethidium bromide gradients as described in Materials and Methods. DNA from approximately 10^8 cells was fractionated on agarose gels in 0·04 M-Tris-acetate, 0·005 M-EDTA, immobilized on nitrocellulose filters and annealed with ^{32}P-labeled probes containing: (a) *copia*, (b) *mdg*1 and (c) pDm39 sequences. To visualize slower migrating and less abundant species, a longer exposure of (a) is presented in (a)^1.

growth conditions, *copia* and 412 (pDm39) elements are abundant as free DNA forms in K_{co} cells but barely detectable in Schneider cells (Fig. 5(a),(b)), whereas other elements, e.g. pDm27 (Fig. 5(c)) are more abundant as free forms in Schneider than in K_{co} cells. Other elements, e.g. *mdg*1 and *mdg*3, appear to be present in similar amounts as free forms in both cell lines (data not shown).

The size difference observed between 412 free DNA forms in K_{co} cells (~7 kb) compared to those in Schneider cells (~5·5 kb) is reproducible. This apparent size difference is also observed when crude Hirt preparations are fractionated on agarose gels before purification on CsCl/ethidium bromide gradients (Fig. 6(c), compare lane 1 to lanes 2 and 3).

(ii) *Comparison between cell lines and between cell lines and embryos*

In order to examine the variation in copy number of free elements more closely, we compared Hirt supernatants from equivalent amounts of K_c, Schneider and embryo cells for element-specific free DNA forms. To rule out the possibility that the differences we have observed were due to losses during purification on CsCl/ethidium bromide gradients, DNA samples from Schneider and embryo cells were loaded directly on agarose gels, and co-fractionated with CsCl/ethidium bromide gradient-purified free DNA from K_{co} cells. Samples were fractionated and analyzed as described previously. Results of these analyses are shown in Figure 6 for (a) *copia*, (b) 297, (c) 412, (d) *gypsy*, (e) *mdg*3 and (f) the F-element-containing clone pDm151. Both *copia* and 297 form more abundant circular DNA species in embryos than in cell lines (Fig. 6(a) and (b)), while the 412 circles appear to be more abundant in K_{co} cells (Fig. 6(c)), *mdg*3 and *gypsy* also appear to be more abundant in embryos (Fig. 6(d) and (e), lane 3), while pDm151 is better represented as free DNA in Schneider cells (Fig. 6(f), lane 2). These differences in abundance of circular DNAs appear to be independent of the number of element-specific sequences integrated in the genome as is demonstrated by chromosomal dot-blots of equivalent amounts of DNA from the three cell types (Fig. 6).

(iii) *Transcription of elements in cell lines*

Reverse transcription is one of the mechanisms proposed to explain the generation of unintegrated forms of transposable elements (Flavell & Ish-

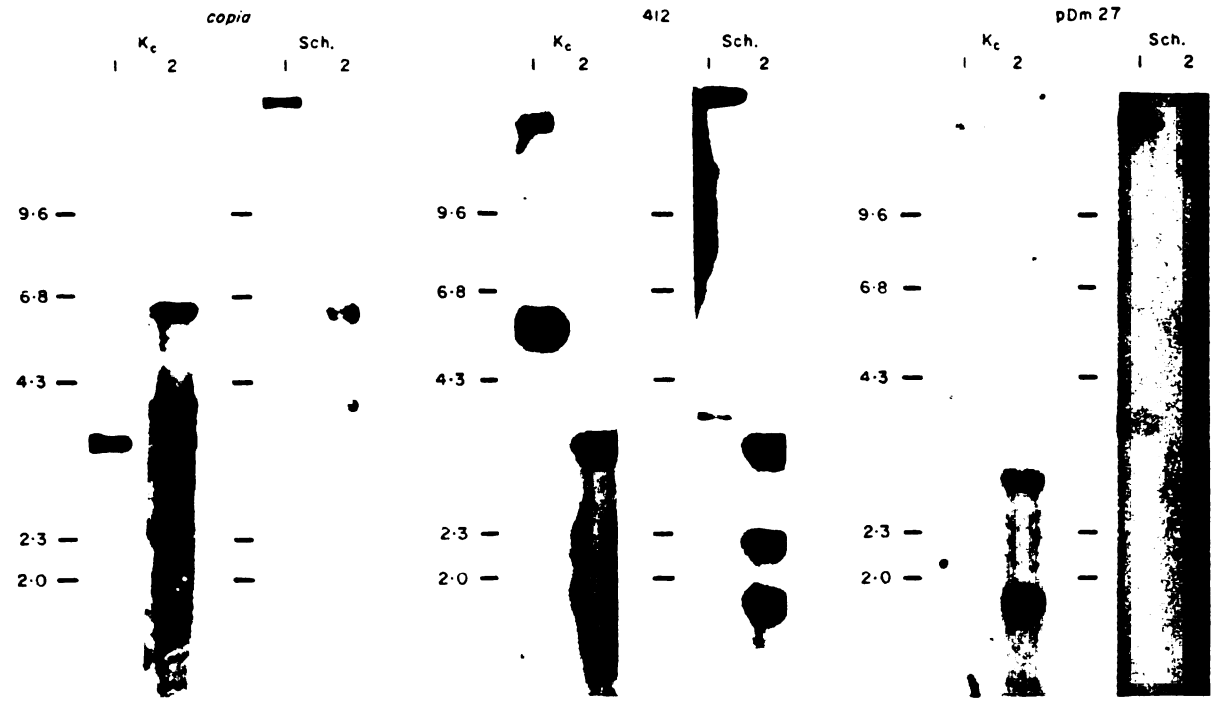


Figure 5. Cell line variation in the abundance of free DNA forms. Samples containing circular DNAs were prepared separately from K_{co} cells (K_c) and Schneider cells (Sch.) and analyzed as described previously. Filters were hybridized with ^{32}P-labeled probes prepared from: (a) *copia*, (b) 412 and (c) pDm27 sequences. Lane 1, free DNA material from approx. 5×10^7 cells; lane 2, 1 μg of *Hha*I-digested Oregon R chromosomal DNA.

Horowicz, 1981, 1983; Flavell, 1984; Shiba & Siago, 1983). This model postulates that genomic length RNA templates are copied to produce unintegrated DNA forms, implying that there might be a direct correlation between the concentration of element-specific RNA and covalently closed circular DNA. To examine this relationship, we have determined the size and amounts of element-specific transcripts in the two cell lines. Poly(A)$^+$ RNA was prepared from the two cell lines, fractionated on formaldehyde-containing gels, immobilized on nitrocellulose, and annealed with labeled element-specific probes. Examination of the different panels in Figure 7 shows no apparent correlation between the amounts of RNA and free DNA per cell. For example, not all the elements are represented by abundant poly(A)$^+$ RNA in a given cultured cell line even though they are present as free DNA forms, e.g. pDm27 and cDm2028 (Fig. 7(d) and (f); see Fig. 3(a) and (g)). In cases where abundant transcripts are detected, this lack of correlation still exists. A case in point is with the *copia* element (Fig. 7(a)). There is approximately twice as much poly(A)$^+$ *copia* RNA in Schneider cells (quantitated by dot-blot analysis) compared to K_{co} cells even though the levels of free DNA forms of *copia* are much lower in Schneider cells (see Fig. 5(a)). The element-specific size classes of RNA appear to be similar between the cell lines (Fig. 7(a) to (c) and (e)). These elements hybridize to apparently full-length transcripts in addition to a few smaller ones. From this we conclude that, at the level of our assays, we cannot find a correlation (either qualitatively or quantitatively) between the levels of poly(A)$^+$ RNA and free DNA per cell.

4. Discussion

(a) *Unintegrated DNAs homologous to other* copia*-like TEs*

There is a growing belief that the *copia*-like transposable elements of *Drosophila* are evolutionarily related to vertebrate retroviruses (Temin, 1980; Finnegan, 1983), and that they may transpose *via* a reverse-transcriptase-mediated pathway (Flavell & Ish-Horowicz, 1981, 1983; Flavell, 1984). According to this model, transposition would be effected by reverse-transcription of the full-length RNA template into full-length linear and circular double-stranded DNA forms, which would then integrate elsewhere in the genome (see Varmus, 1983). Such free DNA forms are presumed, therefore, to be intermediates in the transposition cycle.

We sought to investigate aspects of this model by asking whether the ability to produce unintegrated DNA forms is a general property of the *copia*-like elements. Our results indicate that most of the *copia*-like elements screened exist as extrachromosomal DNA forms in both *Drosophila* embryos and cultured cell lines (Figs 1, 4 and 6). One of the elements, *copia*, generates circles that have been characterized in cultured *Drosophila* cells by Flavell & Ish-Horowicz (1981, 1983). Free DNA forms annealing to the 412 element (Shepherd &

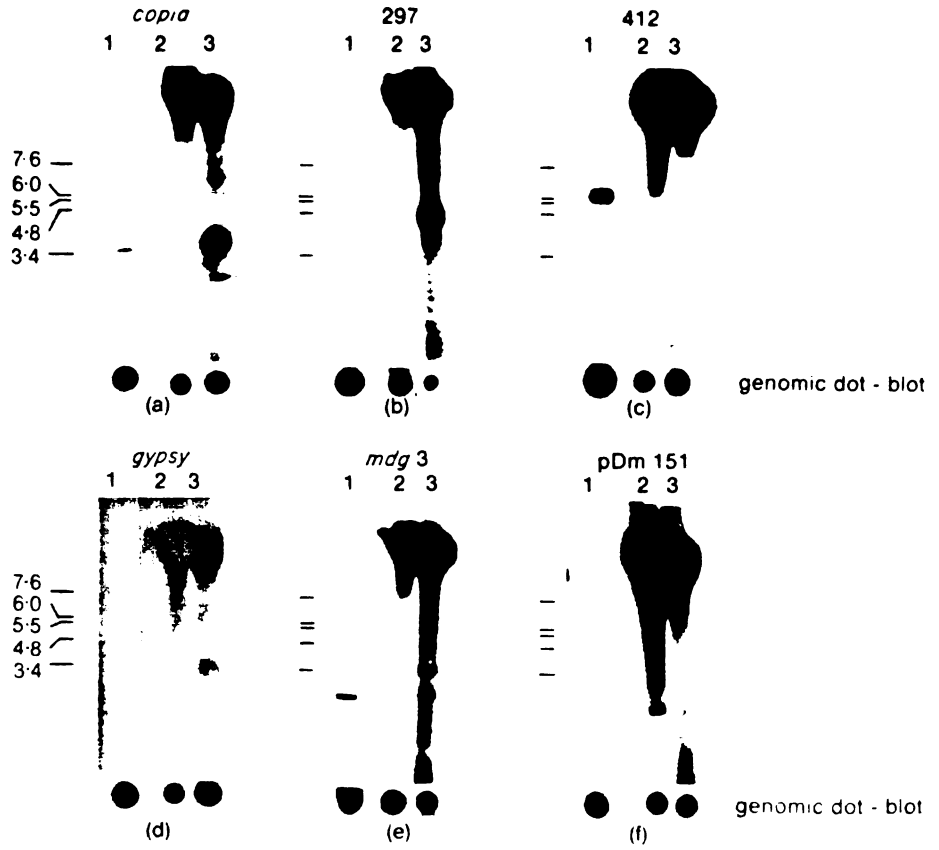


Figure 6. Comparing levels of unintegrated forms in cell lines and embryos. Samples containing free DNAs were analyzed on agarose gels in 0·04 M-Tris-acetate containing 0·005 M-EDTA before (lanes 2 and 3) or after CsCl/ethidium bromide gradient purification (lane 1). Lane 1, purified circular DNA from approx. 10^8 K_{co} cells; lane 2, crude circular DNA preparation from approx. 10^8 Schneider cells; lane 3, crude circular DNA preparation from approx. 10^8 Oregon R embryo cells. Chromosomal DNAs were prepared from cells as described in Materials and Methods. Amount of element-specific sequences in cells were determined in dot-blot analysis of equivalent amounts of thermally denatured DNAs as described by Thomas (1980).

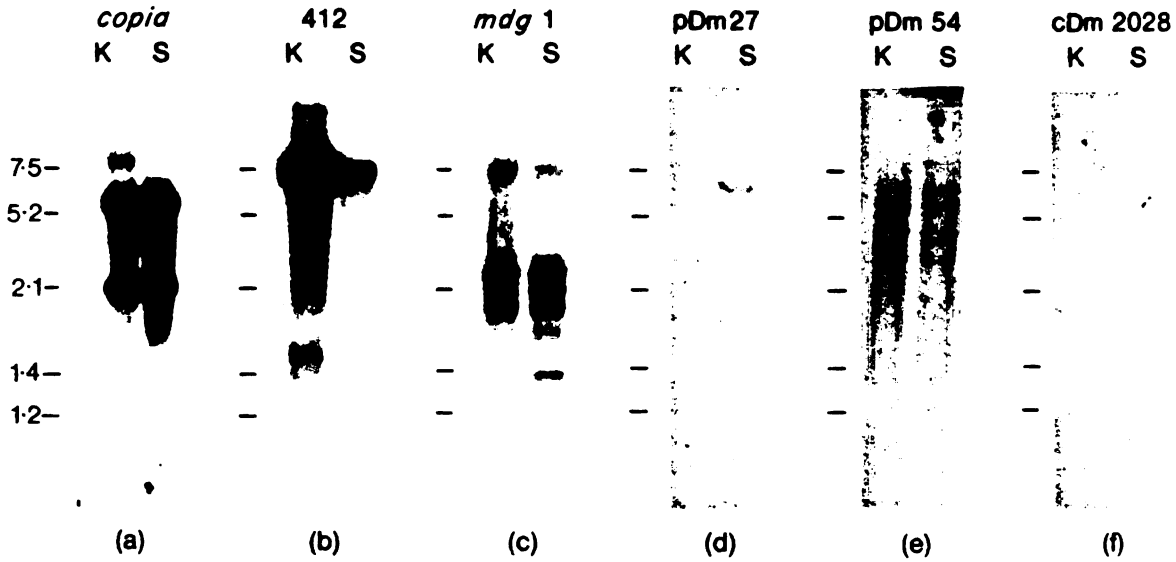


Figure 7. Analysis of RNA transcripts from K_{co} and Schneider cell lines. Approximately 2 μg of poly(A)$^+$ RNA prepared from cultured cells as described in Materials and Methods was fractionated on formaldehyde-containing gels, immobilized on nitrocellulose filters, and annealed with ^{32}P-labeled probes derived from elements shown in the Figure. K, RNA from K_{co} cells; S, RNA from Schneider cells.

Finnegan, 1984), and *mdg*1 and 297 elements (Junakovic & Ballario, 1984) have also been observed in cultured *Drosophila* cells. For the *copia* element, a rapidly migrating doublet of DNA was shown to consist of circles containing one and two LTRs. At least four additional elements tested here, 297 (Fig. 1(b)), *mdg*1, pDm39 (Fig. 4(b) and (c)) and *gypsy* (Fig. 6(d)) have unintegrated DNA forms migrating in a similar fashion. While the nature of these molecules has not been determined by molecular cloning and sequencing, they are likely to have similar structures. We are confident that the molecules we are examining are covalently closed circular DNA molecules because methods designed to enrich for circles (Devenish & Newlon, 1982) enrich for these free DNA forms, and the extrachromosomal DNAs migrate in the position of covalently closed circular DNA in CsCl/ethidium bromide gradients.

We have also observed more slowly migrating forms that hybridize to *copia* probes among circular DNAs from *Drosophila* embryos (Fig. 4(a)[1]). While these have not been studied further and their structure is unknown, they may be comparable to the slowly migrating DNA forms that have been observed in retrovirus-infected cells (Kung *et al.*, 1980). These turned out to be oligomers, consisting of monomeric units arranged in a "head-to-tail" orientation (Kung *et al.*, 1980).

The *mdg*3 transposable element also appears to have a *copia*-like structure. Nucleotide sequence analysis of two independent isolates from the genomic library prepared from the Canton S strain of *D. melanogaster* (Maniatis *et al.*, 1978) indicates that the reported 18 bp inverted repeat at the ends of the element (Bayev *et al.*, 1980) is not part of the *mdg*3 element. The data presented in Figure 2 suggest that the terminal nucleotides of the element are 5′ -T-T-T-A-G-C-C . . . A-C-T-A-C-A- 3′. *mdg*3 is proposed in this paper to be bounded by a 267 bp direct terminal repeat and to generate a 4 bp duplication of target sequence. With this assignment, *mdg*3 has terminal nucleotides that are related to those of other *copia*-like elements. Free DNA forms homologous to *mdg*3 are found in both cultured cell lines and in embryos.

(b) *Uncharacterized repetitive elements and F elements are present as free DNA forms in cultured* Drosophila *cells and embryos*

Most of the uncharacterized middle repetitive DNA elements tested (9/12) were found to be present as unintegrated DNA forms in the cells used (Fig. 3). pDm39 is homologous to the 412 element. The pDm39 clone exists as an extrachromosomal doublet in embryo cells (Fig. 4(c)), suggesting the existence of circles with one or two LTRs. Analysis of 412-specific circles from K_{co} cells showed that molecules with two LTRs are rare (Shepherd & Finnegan, 1984). This is interesting in the light of the fact that K_{co} cells have 412 circles with an apparently larger size than those in Schneider cells (Fig. 5(b)). Experiments are in progress to determine the nature of 412 circles in Schneider cells.

One other element, pDm151, was shown to be related to the F elements in that it hybridizes to the oligo-containing 3′ end of a cloned F element (data not shown). Subsequent experiments showed that F elements also hybridize to free DNA forms (Fig. 3(i) and (j)). The general class of elements to which F elements belong (DiNocera *et al.*, 1983) has been proposed to transpose by a mechanism that postulates an extrachromosomal DNA intermediate (Van Arsdell *et al.*, 1983; Sharp, 1983). The detailed structure of F-specific circles is currently under investigation.

(c) *Variation of levels of free DNA forms*

(i) *Differences between cell lines*

We were surprised to observe a difference in the levels of free DNA forms between the cell lines and between cell lines and embryos (Figs 5 and 6). For example, *copia*, 297, *mdg*3 and *gypsy* are more abundantly represented as free DNA forms in embryos than in the cell lines (Fig. 6). On the other hand, 412 is more abundant as free DNA forms in K_{co} cells than in embryos or the Schneider cell line (Fig. 6(b)). F elements seem to exist as detectable free DNA forms only in Schneider cells (Fig. 6(f)). These differences are not due simply to growth of cells in culture; K_{co} and Schneider cells show differences as great as each cell line shows to embryos (Figs 5 and 6; Table 2). It is equally hard to invoke strain difference as a possible cause of this difference; Schneider cells were derived from Oregon R (see Potter *et al.*, 1979), yet the levels in these cells compared to Oregon R embryos are drastically different (Fig. 6; Table 2).

(ii) *Lack of correlation between free and integrated DNA sequences in cell lines*

There is also no apparent correlation between free DNA in cell lines and embryos and the amount of element-specific sequences in the genome. For example, embryos contain less *copia* and 297 sequences in their genome compared to corresponding cultured cells (Potter *et al.*, 1979; unpublished results), and yet still contain 10 to 20-fold more free DNA. Similar observations have recently been reported (Junakovic & Ballario, 1984) for the *copia* 297, 412 and *mdg*1 elements in K_{co} cells.

(iii) *Lack of correlation between free DNA and poly(A)$^+$ RNA*

Transcription studies of the cultured lines (Fig. 7) also fail to reveal any clear correlation between RNA present per cell and free DNA. Furthermore, some elements (e.g. cDm2028, pDm151) can be detected as free DNA forms in some cells although RNA transcripts are not detectable. It should be noted that the levels of free DNA per cell for an average element is very low, about 0·05 to 0·15

copies per cell, and we have not measured the rates of synthesis and disappearance of the free forms. It is conceivable that very low, even undetectable, amounts of RNA may be sufficient to produce the free DNA observed. Moreover, the actual template for reverse transcriptase-mediated synthesis might be located within virus-like particles (Shiba & Siago, 1983) rather than being free in the cytoplasm or nucleus. Other factors might also influence free DNA synthesis, assuming reverse transcription is the mechanism used. For example, the enzyme reverse transcriptase could be limiting or its activity different between cell lines and embryos.

(d) *Potential significance*

Is there biological significance in the differences we have observed in the amounts of free DNA of various elements in the different contexts? There is evidence that the transcription of some of the *copia*-like elements is regulated during development (Flavell *et al.*, 1980; Young & Schwartz, 1980; Scherer *et al.*, 1981) and it is possible that there may be corresponding changes in amounts of free DNA during development, despite failure to observe correlations between RNA and free DNA levels in this survey. Therefore, if at least some of the free DNA serves as intermediates in transposition, it is possible that we are indirectly measuring mobilization of these elements and that the frequencies are determined by or influencing developmental stages. In addition, the discovery of solo LTRs for the 412, *B104*, *gypsy* and *copia* elements indicates that these elements may be excised from the genome leaving one LTR in the chromosome and generating free circular DNA with a single LTR. In the case of *copia* (P. M. Bingham, personal communication) and *gypsy* (W. Bender, personal communication), such events are associated with phenotypic changes, suggesting a role for these elements in the regulation of gene expression.

We thank John Majors for advice and discussion and colleagues listed in Table 1 for DNA clones. This work was supported by grants from the National Institutes of Health and the American Cancer Society. K.G.M. was supported by a Fulbright Fellowship. H.E.V. is an American Cancer Society Professor of Molecular Virology. We are indebted to J. Marinos for preparing the manuscript.

References

Bayev, A. A., Krayev, A. S., Lyubomirskaya, Ilyin, Y. V., Skryabin, K. G. & Georgiev, G. P. (1980). *Nucl. Acids Res.* **8**, 3263–3273.

Benton, W. D. & Davis, R. W. (1977). *Science*, **196**, 180–182.

Bingham, P. M. (1980). *Cold Spring Harbor Symp. Quant. Biol.* **45**, 519–525.

Cameron, J. R., Loh, E. Y. & Davis, R. W. (1979). *Cell*, **16**, 739–751.

Casse, F., Boucher, C., Julliot, J. S., Michel, M. & Denarie, J. (1979). *J. Gen. Microbiol.* **113**, 229–242.

Dawid, I. B., Long, E. O., DiNocera, P. P. & Pardue, M. L. (1981). *Cell*, **25**, 399–408.

Denhardt, D. J. (1966). *Biochem. Biophys. Res. Commun.* **23**, 641–646.

Devenish, R. J. & Newlon, C. S. (1982). *Gene*, **18**, 277–288.

DiNocera, P. P., Digan, M. E. & Dawid, I. B. (1983). *J. Mol. Biol.* **168**, 715–727.

Dowsett, A. P. & Young, M. W. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 4570–4574.

Dunsmuir, P., Broreia, R. J., Simon, M. A. & Rubin, G. M. (1980). *Cell*, **21**, 575–579.

Echalier, G. & Ohanessian, A. (1969). *C.R. Acad. Sci. (Paris)*, **261**, 1771–1773.

Echalier, G. & Ohanessian, A. (1970). *In Vitro*, **6**, 162–172.

Flavell, A. J. (1984). *Nature (London)*, **310**, 514–516.

Flavell, A. J. & Ish-Horowicz, D. (1981). *Nature (London)*, **292**, 591–595.

Flavell, A. J. & Ish-Horowicz, D. (1983). *Cell*, **34**, 415–419.

Flavell, A. J., Ruby, S. W., Toole, J. J., Roberts, B. E. & Rubin, G. M. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 7107–7111.

Georgiev, G. P., Ilyin, Y. V., Chmeliauskaite, Y. G., Ryskov, A. P., Kramekov, D. A., Skryabin, K. G., Krayev, A. S., Lukanidia, E. M. & Grigoryan, M. S. (1980). *Cold Spring Harbor Symp. Quant. Biol.* **45**, 641–653.

Heine, C. W., Kelly, D. C. & Avery, R. J. J. (1980). *J. Gen. Virol.* **49**, 385–395.

Junakovic, N. & Ballario, P. (1984). *Plasmid*, **11**, 109–115.

Kleckner, N. (1981). *Annu. Rev. Genet.* **15**, 341–404.

Kole, L. B., Haynes, S. R. & Jelinek, W. R. (1983). *J. Mol. Biol.* **165**, 257–286.

Kung, H. J., Shank, P. R., Bishop, J. M. & Varmus, H. E. (1980). *Virology*, **103**, 425–433.

Levis, R., Dunsmuir, P. & Rubin, G. M. (1980). *Cell*, **21**, 581–588.

Maniatis, T., Hardison, R. C., Lacy, E., Laver, J., O'Connell, C. O., Quon, D., Sim, G. K. & Efstratiatis, A. (1978). *Cell*, **15**, 687–701.

Maxam, A. & Gilbert, W. (1980). *Methods Enzymol.* **65**, 499–560.

McGinnis, W. & Beckendorf, S. K. (1983). *Nucl. Acids Res.* **11**, 737–751.

Meyerowitz, E. M. & Hogness, D. S. (1982). *Cell*, **28**, 165–176.

Nusse, R. & Varmus, H. E. (1982). *Cell*, **31**, 99–109.

O'Hare, K. & Rubin, G. M. (1983). *Cell*, **34**, 25–34.

Potter, S. S., Brorein, W. J., Dunsmuir, P. & Rubin, G. M. (1979). *Cell*, **17**, 415–427.

Potter, S. S., Truett, M., Phillips, M. & Maher, A. (1980). *Cell*, **20**, 639–647.

Radloff, R., Baver, W. & Vinograd, J. (1967). *Proc. Nat. Acad. Sci., U.S.A.* **57**, 1514–1521.

Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977). *J. Mol. Biol.* **113**, 237–251.

Rogers, J. (1983). *Nature (London)*, **306**, 113–114.

Rubin, G. M. (1983). In *Mobile Genetic Elements* (Shapiro, J. A., ed.), pp. 329–361. Academic Press, New York.

Scherer, G., Telford, J., Baldari, C. & Pirrotta, V. (1981). *Develop. Biol.* **86**, 438–447.

Scherer, G., Tschudi, C., Perera, J., Delius, H. & Pirrota, V. (1982). *J. Mol. Biol.* **157**, 435–451.

Schneider, I. (1972). *J. Embryol. Exp. Morph.* **27**, 353–365.

Shafit-Zagardo, B., Brown, F. L., Zavodyn, P. J. & Maio, J. J. (1983). *Nature (London)*, **304**, 277–280.
Sharp, P. A. (1983). *Nature (London)*, **301**, 471–472.
Shepherd, B. M. & Finnegan, D. J. (1984). *J. Mol. Biol.*, in the press.
Shiba, T. & Siago, K. (1983). *Nature (London)*, **302**, 119–124.
Smith, D. R. & Calvo, J. M. (1980). *Nucl. Acids Res.* **8**, 2255–2274.
Southern, E. M. (1975). *J. Mol. Biol.* **98**, 503–517.
Spradling, A. C. & Rubin, G. M. (1981). *Annu. Rev. Genet.* 219–264.
Stanfield, S. & Helinski, D. R. (1976). *Cell*, **9**, 333–345.
Tchurikov, N. A., Ilyin, Y. V., Skryabin, K. G., Ananiev, E. V., Bayev, A. A., Krayev, A. S., Zelentsova, E. S., Kulgushkin, V. V., Lyubomirskaya, N. V. & Georgiev, G. P. (1980). *Cold Spring Harbor Symp. Quant. Biol.* **45**, 655–665.
Temin, H. M. (1981). *Cell*, **21**, 599–600.
Thomas, P. S. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 5201–5205.
Truett, M. A., Jones, R. S. & Potter, S. S. (1981). *Cell*, **24**, 753–763.
Van Arsdell, S. W., Denison, R. A., Bernstein, L. B., Weiner, A. W., Manser, T. & Gesteland, R. F. (1981). *Cell*, **26**, 11–17.
Varmus, H. E. (1983). In *Mobile Genetic Elements* (Shapiro, J. A., ed.), pp. 411–503, Academic Press, New York.
Varmus, H. E. & Swanstrom, R. (1982). In *The Molecular Biology of Tumor Viruses*, part III, *RNA Tumor Viruses* (Weiss, R. A., Teich, N., Varmus, H. E., & Coffin, J., eds), pp. 369–512, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
Varmus, H. E., Quintrell, N. & Ortiz, S. (1981). *Cell*, **25**, 23–36.
Vieira, J. & Messing, J. (1982). *Gene*, **19**, 259–268.
Will, B. M., Bayev, A. A. & Finnegan, D. J. (1981). *J. Mol. Biol.* **153**, 897–915.
Young, M. W. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 6274–6278.
Young, M. W. & Schwartz, H. E. (1980). *Cold Spring Harbor Symp. Quant. Biol.* **45**, 629–640.

Edited by M. Gottesman

CHAPTER III

THE ORIGIN OF SOME ABERRANT TY STRUCTURES IN SACCHAROMYCES CEREVICIAE

Structures and possible origins of delta-containing sequences in the genome of Saccharomyces cereviciae.

Kevin G.Mossie1 and Harold E.Varmus1,2

Departments of Biochemistry and Biophysics[1],

and

Microbiology and Immunology[1,2].

University of California, San Francisco. San Francisco, CA 94143.

ABSTRACT :

We have analysed the structure of a previously cloned, permuted Ty element (Ty S13 - (1)), and found that it contains a single intact delta flanked by epsilon sequences. This unusual structure may have been formed by a homologous recombination event between a free Ty circular transposition intermediate and integrated Ty or delta sequences in the Saccharomyces cerevisiae genome. To determine if other aberrant Ty sequences, which could have been formed via this pathway involving Ty circles containing two deltas in tandem, exist in the yeast genome, we have screened the yeast genome for delta-delta structures. The results of this screen has uncovered possible Ty integration into both unique and repeated (sigma) DNA. Delta-delta structures obtained suggest that, among the other possibilities, these structures may originate from a homologous recombination event between unintegrated circular Ty DNAs with Ty (or delta) sequences integrated in the genome. We have used oligonucleotides specific to one of the delta-delta junction sequences to isolate flanking Ty sequences. The structure of this clone is consistent with notion that a fraction of abherrant Ty or delta elements in the yeast genome may be formed via this pathway.

INTRODUCTION:

The dispersed, moderately repeated Ty family of transposable elements in Saccharomyces cereviciae (1,2) share distinct structural and functional similarities with proviral forms of vertebrate retroviral proviruses (3). The elements are long, generate short duplications of target sequence on integration, and are bounded by direct repeats (called long terminal repeats or LTRs in retroviruses, and deltas in Ty). These direct repeats flank long open reading frames that encode gene products that are required for viral DNA replication (3), and are presumed to be required for Ty transposition (4,5). Furthermore, the expression of Ty gene products (6), and some retroviral genes utilise a novel ribosomal frameshifting mechanism (7). Both Ty and retroviruses are transcribed to yield genomic-length RNAs that are terminally redundant (3,8). Finally, Ty elements, like retroviral proviruses, have sequences analogous to those used in priming retroviral DNA synthesis : a tRNA primer binding site at one internal boundary of the terminal repeat (4,8,9) and a polypurine stretch at the other internal boundary (3-5).

Boeke et al. (10) demonstrated directly, by showing the loss of marked intron fragments in a newly-transposed Ty, that Ty elements transpose via an RNA intermediate. Futhermore, they provided evidence for the direction of the transposition pathway, by demonstrating that a delta

polymorphism in the U_5 region of the left delta, was transferred to the U_5 region in the right delta in newly transposed elements (10). Subsequent studies by Garfinkel et.al. (11) and Mellor et.al. (12) indicated that the induction of Ty element transposition correlated with the appearance in the cytoplasm, of intracellular virus-like particles containing Ty RNA and having associated reverse transcriptase activity (11,12). Ty elements also resemble the different families of copia-like transposable elements in Drosophila in structure (13), transcriptional properties (14-16), and gene organisation (15,17-20). Furthermore, virus-like particles with associated reverse transcriptase activity and copia RNA have been detected in Drosophila (21).

The evidence that the Drosophila elements transpose via an RNA intermediate has emerged through analysis of putative unintegrated transposition intermediates (22-28). Analyses of these structures have enabled investigators to conclude that such elements transpose in in a pathway that is similar to retroviral DNA replication pathway (3,28). Although the evidence for Ty transposition via an RNA intermediate is overwhelming (10-12), details of this pathway are still unclear. This is largely due to the difficulties encountered in detecting putative unintegrated transposition intermediates (unpublished observations) even in cells where high frequency Ty transposition has been induced (Boeke -

pers. comm.). Our inability to detect unintegrated Ty forms may be due to the fact that Ty movement by homologous (i.e. recombinational) events (10^{-4}-10^{-5}) is considerably higher than Ty movement by non-homologous (i.e. transpositional) events (10^{-7}-10^{-8}) (2,10).

To determine if a fraction of putative Ty transpositional intermediates, similar to those observed for retroviruses (3), and the Drosophila elements (21-25), is integrated in the yeast genome by homologous recombination we analysed an aberrant Ty molecule, and constructed a scheme for the isolation of other abherrant Ty elements that were likely to have been formed via this pathway. The rationale for our approach was that since we could not detect unintegrated transposition intermediates, it could be possible to detect such molecules "trapped" by a homologous integration event as shown in Figure 1. We describe the results of this screen and present the strucure of an aberrant Ty element that is likely to have originated by a homologous recombination event between an unintegrated circular DNA Ty element and either an intact Ty or a solo delta in the yeast genome.

MATERIALS and METHODS :

a) Bacterial and Yeast Strains:

Bacterial strains LE 392, c600, and MA 150 (Hfl-1) were provided by C.Nottenberg. Yeast strains DC021/022, and XG 1 (29) were supplied by

C.Newlon.

b) Preparation and analysis of DNA:

Yeast genomic DNA was prepared as described by Davis et.al. (30). DNA was further purified by banding in CsCl gradients as described by Shank and Varmus (31). Plasmid and bacteriophage DNA were prepared and purified using standard methods (32). M13 mp18 replicative (RF) and single-stranded (ss) DNAs were prepared as described by Messing (33).

c) Molecular Cloning:

Small Xho1 fragments were prepared by digesting genomic DNA with Xho1, fractionation on agarose gels, followed by electroelutrion. These were cloned into the Xho1 site of a modified lambda gt10 vector. Recombinants were screened using delta-specific probes as described previously (25). Large (10-15kb) fragments were cloned from partial digest with restriction endonuclease Mbo1 into the Bam H1 site of the EMBL 4 vector (34). Recombinants were screened using labelled oligonucleotides as described by Zoller & Smith (35). Restriction fragments were subcloned into SP6 or M13 vector systems for further analysis.

d) Nucleotide sequence analyses:

DNA segments were labelled by the chemical degradation method as desribed previously (25). DNA fragments, subcloned in M13 RF vectors, were sequenced using chain terminators as described by Sanger et.al. (36).

RESULTS:

a)Analysis of the Structure of Ty S13:

The Ty S13 clone, isolated and described by Cameron et.al. (1), has a structure which suggested that it may have arisen from a Ty circle. We were interested in the structure of Ty S13 around the delta sequences. The restriction map of Ty S13 is shown in Figure 2, and is compared to the restriction map of a linear, nonpermuted Ty 912 (4). The nucleotide sequence of the Hpa 1 delta-containing fragment was determined and is represented in Figure 2C. The nucleotide sequence of the delta fragment was almost identical to the delta sequence in Ty 912 (4), with minor changes (data not shown). The apparently intact delta is flanked on the left by the polypurine tract and on the right by the tRNA primer binding site. An obvious feature from these analyses, therefore, is that Ty S13 is a permuted Ty element, with an intact delta located between the epsilon sequences. There are a number of ways in which the S13 clone could have been generated (Figure 3) : A) S13 could be derived from tandemly duplicated Ty forms integrated in the genome. These Ty forms could have arisen from a number of recombinational events, including unequal crossing over, or homologous recombination between a Ty circle with one delta, and an intact Ty integrated in the genome. B) S13 could be cloned from a free Ty circle with one delta, and C) S13 could have originated from a Ty circle which had integrated non-homologously with non Ty

chromosomal DNA sequences. It is not possible to distinguish among these events from analysis of this clone.

b) Strategy for cloning delta-delta junction fragments:

One possible way to explain the formation of Ty S13 is by a homologous integration event involving a Ty circle containg one delta (analogous with retroviruses (3) and the Drosophila elements (25)). If there are in fact free Ty circles containing two deltas in tandem, integration of such circles would generate Ty similar structures with the two deltas in tandem forming a junction between epsilon sequences (see Figure 1). To screen yeast genomic DNA for the presence of these structures, we took advantage of the observation that most Ty elements have a unique Xho1 site in each of the deltas (2). If Ty circles containing two deltas in tandem are integrated in the yeast genome by the homologous recombination event, then digestion of yeast genomic DNA with Xho1 should generate an Xho1 fragment whose length would be the size of the delta sequence (330bp). The junction sequence between the two deltas should therefore represent the junction sequence in free Ty circular DNA prior to integration (see Figure 1). We prepared and cloned Xho1 fragments into a modified lambda gt10 vector as described in Materials and Methods. Recombinant clones were screened using labelled delta fragments, and positive clones were subcloned into M13 vectors and sequenced as described in Materials and Methods. Results of these analyses are presented in Figure 4. The

structure of the Ty delta represented in terms of its U_3-R-U_5 domains is shown in Figure 4A.

a) Ty integration in cellular DNA : Integration of a Ty element close to an Xho1 site should generate Xho1 fragments that are partly delta and partly flanking cellular sequence. Clone SC.24a (Fig. (3B)) is an example of such clones picked up in this screen. The DNA sequence upstream of the start of delta sequence is non-Ty sequence and probably represents flanking unique cellular DNA. Clone SC.4a (Fig. 3(C)) is an example of a possible Ty integration into another repetitive element, the sigma element(37). Integration of a Ty (or delta) sequences into sigma sequences have been observed previously (38), and analyses of flanking sigma sequences indicate that this clone probably represents an independent integration event (Fig. 3(C),(38)).

b) Delta-delta junctions : Two different clones were isolated in which delta sequences were found in tandem. The delta-delta junction was determined by mapping the end of one delta and aligning adjoining sequences with known delta sequence. Clones SC.7a and SC.11a are identical up to the 5' end of Ty (or delta) sequence. The selected region of the nucleotide sequence of these clones is shown in Fig. 4 and schematically depicted in terms of the U_3-R-U_5 components of the delta structure (Fig. 4(D-E)). In clone SC.7a, the delta sequences between the Xho1 sites are in the same

orientation. The upstream delta, however, is missing 66 nucleotides, which include all of U_5 and about 20 nucleotides of R (Figure 4)D)). Clone SC.11a also has two deltas in tandem. As in the previous clone, sequences from the Xho1 site indicate that the right delta is intact. The left delta sequences are in an inverted orientation relative to the right delta, and all of U_5 sequences are missing.

There are a number of ways to explain how the Xho1 delta-delta structures described above could be generated. The simplest explanation would be the integration of Ty element non-homologously into delta sequences in either orientation to generate structures resembling clones SC.7a and SC.11a. Such events have been observed in the SUP4 locus (39) and the region near the LEU2 gene on the left arm of chromosome III (40). The second possibility, is that the Xho1 delta-delta fragments were part of a Ty circle containing tandem deltas which integrated by recombining homologously with Ty or delta sequences in the genome. Clone SC.11a could have resulted from a homologous integration event involving a Ty circle that had undergone autointegration (23,23,41) prior to recombining with Ty or delta sequences in the genome. Defective circular DNA molecules with similar structures have been described in retroviruses (41) and Drosophila retrotransposons (23,24). The last possibility, is that these fragments were not cloned from genomic DNA, but from free Ty DNAs.

In order to determine the origin of the Xho1 delta-delta fragments analysed above, we used labelled oligonucleotides TAATCCCAACATATACAGAAT and TTATCCCAACAGGTTCCATTA, corresponding to the delta junctions in clones SC.7a and SC.11a, respectively, to screen yeast genome for these structures. Southern blot analyses using these oligomers as probes indicate there were at least two copies for each clone in the yeast genome (Figure 5). We presume that these are genomic rather than contaminating free circular Ty DNAs since some of the bands were larger than than the size of the full-length Ty molecule (Figure 5). To determine the structure of these molecules, we prepared genomic libraries and screened these with labelled oligomers as described in Materials and Methods. To eliminate artifacts, we rescreened positive clones with labelled delta probes before further characterisation. Using the probe designed to clone sequences flanking clone SC.7a, we isolated a lambda clone, lambda SC.7a (0-1b), and the structure is presented in Figure 5. Restriction enzyme analysis using labelled Ty or delta probes indicate that sequences flanking the Ty element contain neither Ty nor delta sequences. The delta-delta junction has been mapped to be contained entirely within a 700bp Bgl II fragment. The presence of the Bgl II site so close to delta sequences suggests that it may be the same Bgl II site that marks the beginning of the polypurine tract in Ty 912 (4). We have therefore tentatively assigned the direction of Ty transcription as

shown in Figure 6.

Other possible ways to explain the origin of the structure shown in Figure 6. (a), Integration of a Ty element 66 nucleotides from the 3' end of a delta sequence, or (b), homologous recombination of a Ty circle with a defective circle junction and either an intact Ty or delta sequence (Figure 7). Determination of the nucleotide sequence of the right delta (in progress) should allow us to distinguish between these possibilities (see below).

DISCUSSION :

Although the evidence for Ty element transposition via an RNA intermediate is overwhelming, details of this pathway are not clear. From the work published by Boeke et.al. (10), we know that the strategy for DNA synthesis during transposition closely resembles that of retrovirus DNA since Boeke et al.(10) showed that it is U_5 information from the 5' delta, and U_3 information, from the 3' delta that are used to provide delta sequences in newly transposed elements. Since Ty and the Drosophila retrotransposons appear to be functionally equivalent, it seems plausible to assume that they would transpose in a similar pathway, with similar intermediates. We have been unable to detect free Ty DNA forms from Hirt supernatant preparations using different yeast strains. We concluded that if these exist, they are present at levels lower than one copy per 10^3

cells. Other investgators were also unable to detect free Ty DNA forms even in cells where Ty transposition is induced (10) (Boeke - pers.comm.). Our inability, however, to detect unintegrated Ty forms may suggest a basic difference between Ty and the Drosophila elements and retroviruses. Some of the clones we have analysed are potential candidates for having arised by integration of free Ty DNA intermediates into genomic Ty forms. Although the notion that Ty S13 had arisen from an integration event involving an unintegrated Ty circle containing one delta with a Ty element integrated in the genome is possible, S13 could be derived from tandemly repeated Ty elements that arose via an unequal crossover event between two Ty elements, located at different positions in the yeast genome. The reason for isolating and cloning small (300-350bp Xho1) fragments was to selectively clone potential circle junction fragments. Our screen effectively picked up Ty integrations near Xho1 sites in the yeast genome, one in unknown yeast DNA, and the other in the repeated element sigma. Ty integration into the sigma element has been reported (37,38). We do not know if sigma elements are hot spots for Ty integration, but they do provide a large target since they are repeated in the genome. The clones we isolated where the deltas are in tandem could have arisen by Ty integration into other Ty or delta sequences. Clone SC.7a is a candidate clone, however, for having arisen from a recombination event involving a Ty circle with a defective circle junction and a delta sequence in the yeast

genome. It is interesting to notethat for both retroviruses and the Drosophila retrotransposons (3,22-28), most of the unintegrated circular DNA molecules analysed are defective, suggesting that good integration substrates are efficiently integrated by the element-encoded endonuclease. Using that analogy, therefore, it is likely that in the case of Ty intermedites most of the putative DNA intermediates with defective 'circle junctions', will be introduced into the yeast genome via a homologous recombination pathway.

REFERENCES:

1. Cameron, J.R., Loh, E.Y., and Davis, R.W. (1979). Evidence for transposition of dispersed repetitive DNA families in yeast. *Cell 16*, 739-751.

2. Roeder, G.S., and Fink, G.R. (1983). Transposable elements in yeast. In Mobile Genetic Elements, J.A. Shapiro, ed. (New York : Academic Press), pp. 299-326.

3. Varmus, H.E. (1983). Retroviruses. In Mobile genetic elements, J.A. Shapiro, ed. (New York : Academic Press), pp. 411-502.

4. Clare, J., and Farabaugh, P.J. (1985). Nucleotide sequence of a Ty1 element : evidence for a novel mechanism of gene expresssion. *Proc.Natl.Acad.Sci.U.S.A.* 82, 2829-2833.

5. Hauber, J., Nelbock-Hochstetter, P., and Feldmann, H. (1985). Nucleotide sequence of a Ty element from yeast. *Nucleic Acids Res.* 13, 2745-2758.

6. Mellor, J.S., Fulton, S.M., Dobson, M.J., Wilson, W.,Kingsman, S.M., and Kingsman, A.J. (1985). A retrovirus-like strategy for expression of a fusion protein encoded by yeast transposon Ty1. *Nature (London)* 313, 243-246.

7. Jacks, T., and Varmus, H.E. (1985). Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* 230, 1237-1242.

8. Elder, R.T., Loh, E.Y., and Davis, R.W. (1983). RNA from the yeast transposable element Ty1 has both ends in the direct repeats, a structure similar to retrovirus RNA. *Proc.Natl.Acad.Sci.U.S.A.* 80, 2432-2436.

9. Taylor, J.M. (1977). An analysis of the role of tRNA species as primers for the transcription into DNA of RNA tumor virus genomes. *Biochim.Biophys.Acta* 473, 57-71.

10. Boeke, J.D., Garfinkel, D.J., Stiles, C.A., and Fink, G.R. (1985). Ty elements transpose through an RNA intermediate. *Cell* 40, 491-500.

11. Garfinkel, D.J., Boeke, J.D., and Fink, G.R. (1985). Ty element transposition : reverse transcriptase and virus-like particles. *Cell* 42, 507-517.

12. Mellor, J., Malim, M.H., Gull, K., Tuite, M.F., McCready, S., Dibbayawan, T., Kingsman, S., and Kingsman, A.J. (1985). Reverse transcriptase activity and Ty RNA are associated with virus-like particles in yeast. *Nature (London)* 316, 583-586.

13. Rubin, G.M. (1983). Dispersed repetitive DNAs in Drosophila. In Mobile Genetic Elements, J.A. Shapiro, ed. (New York : Academic Press), pp 329-361.

14. Flavell, A.J., Levis, R., Simon, M.A., and Rubin, G.M. (1982). The termini of RNAs encoded by the transposable element copia. *Nucleic Acids Res.* 9, 6279-6291.

15. Emori, Y., Shiba, T., Kanaya.S., Inouye, S., Yuki, S., and Saigo, K. (1985). The nucleotide sequence of copia and copia-related RNA in Drosophila virus-like particles. *Nature (London)* 315, 773-776.

16. Scherer, G., Tschudi, G., Perera, J., Delius, H., and Pirritta, V. (1982). B104, a new dispersed repeated gene family in Drosophila melanogaster and its analogies with retroviruses. *J.Mol.Biol.* 157, 435-451.

17. Mount, S.M., and Rubin, G.M. (1985). Complete sequence of the Drosophila transposable element copia : homology between copia and retroviral proteins.

18. Marlor, R.L., Parkhurst, S.M., and Corces, V.G. (1986). The Drosophila melanogaster element gypsy transposable element encodes putative gene products homologous to retroviral proteins. *Mol. Cell Biol.* 6,

1129-1134.

19. Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K., and Yuki, S. (1984). Identification of the coding sequence for reverse transcriptase-like enzyme in a transposable genetic element 17.6, in Drosophila melanogaster. *Nature (London) 312,* 659-661.

20. Yuki, S., Ishimaru, S., Inouye, S., and Saigo, K. (1986). Identification of genes for reverse transcriptase-like enzymes in two Drosophila retrotransposons, 412 and gypsy : a rapid detection method of reverse transcriptase genes using YXDD box probes. *Nucleic Acids Res. 14,* 3017-3029.

21. Shiba, T., and Saigo, K. (1983). Retrovirus-like particles containing RNA homologous to the transposable element copia in Drosophila melanogaster. *Nature (London) 302,* 119-123.

22. Flavell, A.J., and Ish-Horowicz, D. (1981). Extrachromosomal circular copies of the eukaryotic transposable element copia in cultured Drosophila cells. *Nature (London) 292,* 591-595.

23. Flavell, A.J., and Ish-Horowicz, D. (1983). The origin of circular copia elements. *Cell 34,* 591-595.

24. Shepard, B.M., and Finnegan, D.J. (1985). Structure of circular copies of the 412 transposable element present in Drosophila melanogaster tissue culture cells and isolation of free 412 long terminal repeat. *J.Mol.Biol.*

25. Mossie, K.G., Young, M.W., and Varmus, H.E. (1985). Extrachromosomal DNA forms of copia-like transposable elements, F elements, and middle repetitive DNA sequences in Drosophila melanogaster : variation in cultured cells and embryos. *J.Mol.Biol.* *182*, 31-43.

26. Flavell, A.J. (1984). Role of reverse transcription in the generation of extrachromosomal copia mobile elements. *Nature (London)* *310*, 514-516.

27. Flavell, A.J., and Brierley, C. (1986). The termini of extrachromosomal linear copia elements. *Nucleic Acids Res.* *14*, 3659-3669.

28. Arkhipova, I.R., Nazo, A.M., Cherkasova, V.A., Gorelova, T.V., Schuppe, N.G., and Ilyin, Y.V. (1986). The steps of reverse transcription of Drosophila mobile dispersed genetic elements and U_3-R-U_5 structure of their LTRs. *Cell* *44*, 555-563.

29. Devenish, R.J., and Newlon, C.S. (1982). Isolation and characterisation of yeast ring chromosome by a method applicable to other circular DNAs. *Gene* *18*, 277-288.

30. Davis, R.W., Thomas, M., Cameron, J.R., St. John, T.P., Scherer, S., and Padgett, R.A. (1980). Rapid DNA isolation for enzymatic and hybridisation analysis. In *Methods in Enzymology* *65*, 404-411.

31. Shank, P., and Varmus, H.E. (1976).

32. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning - a Laboratory Manual. Cold Spring Harbor Laboratory

33. Messing, J. (1983). New M13 vectors for cloning. In *Methods in Enzymology* 101, 20-78.

34.

35. Zoller, M.J., and Smith, M. (1983). Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors. In *Methods in Enzymology* 100, 468-500.

36. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proc.Natl.Acad.Sci.U.S.A.* 74, 5463-5467.

37. Del Rey, F.J., Donahue, T.F., and Fink, G.R. (1982). *Sigma*, a repetitive element found adjacent to tRNA genes in yeast. *Proc.Natl.Acad.Sci.U.S.A.* 79, 4138-4142.

38. Chisholm, G.E., Genbauffe, F.S., and Cooper, T.G. (1984). *Tau*, a repeated DNA in yeast. *Proc.Natl.Acad.Sci.U.S.A.* 81, 2965-2969.

39. Rothstein, R., Helms, C., Rosenberg, N. (1987). Concerted deletions and inversions caused by mitotic recombination between the yeast delta sequences. *Mol.Cell.Biol. (in press).*

40. Warmington, J.R., Anwar, R., Newlon, C.S., Waring, R.B., Davies, W.R., Indge, K.J., and Oliver, S.G. (1986). A hot spot for Ty transpositionm on the left arm of yeast chromosome III. *Nucleic Acids Res.*

14, 3475-3485.

FIGURE LEGENDS :

Figure 1: Homologous recombinational model for the generation of Ty aberrant Ty structures. An integration event involving a Ty circle containing one delta would generate a structure similar to that observed for Ty S13 if an intact Ty is invoved and the structure in Figure 1(a), if a circle recombines with a solo delta. A Ty circle containing two deltas would recombine with a solo delta to produce a similarly structured element with the exception that one of the termini would contain the two deltas in tandem. Functionally, the junction between the delta is identical to the one in free Ty circles. The position of unique Xho 1 sites in delta sequences is also shown.

Figure 2 : Structural analysis of Ty S13. The restriction map of Ty S13 (b) is shown in relation to the partial restriction map of linear, non-permuted Ty 912 (a) (4). The nucleotide sequence of the 1200bp Hpa1 fragment shown above was determined, and the sequence around the delta is depicted in Figure 1(c). The nucleotide sequence is nearly identical to the published Ty 912 sequence (4), with few base changes (data not shown).

Figure 3 : Possible mechanisms for generating the Ty S13 clone : (A) The

5.4kb Eco R1 fragment have been cloned from Ty elements that are in tandem, or (B) from a free Ty circle, or (C) a Ty molecule that had integrated abherrantly in a nonhomologous fashion in chromosomal DNA.

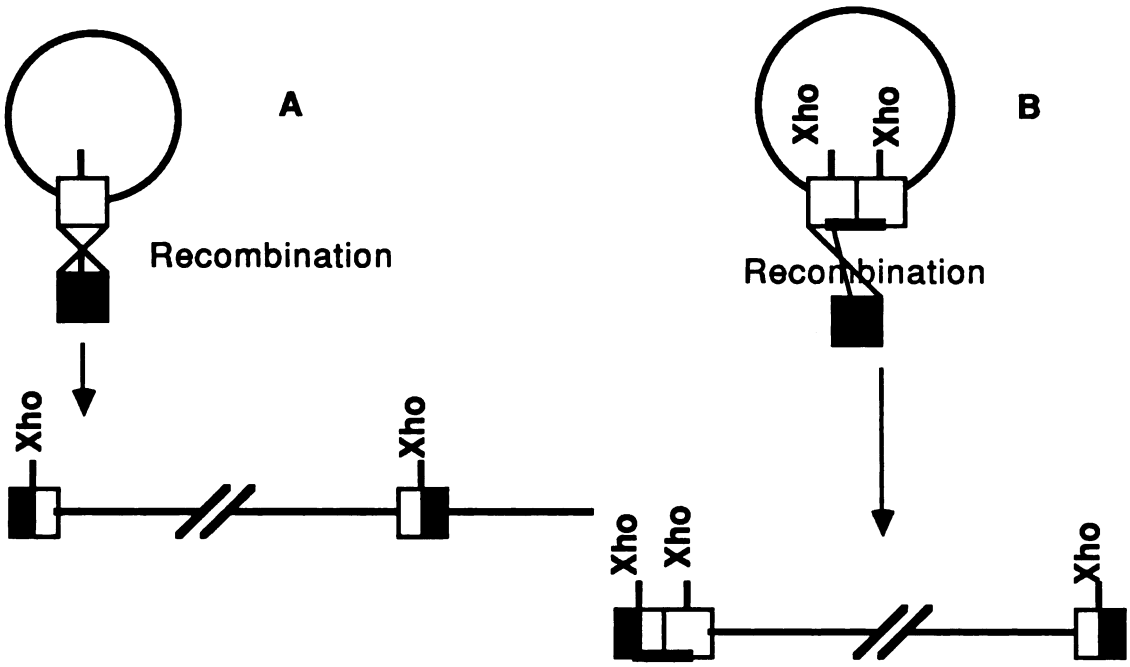
Figure 4 : Structural analysis of 300-350bp Xho1 fragments. (A) The structure of the Ty delta is presented, showing the U_3-R-U_5 domains. (B), (C) Nucleotide sequence of possible integration of Ty or delta sequences in non-Ty sequences. The position and direction of delta is as shown in Fig. 3(A). (D), (E) Association and possible integration of Ty (or delta) sequences in other Ty (or delta) sites. The position of the sequences in the overall delta sequence is as shown. The direction of the delta sequence is shown by the arrows.

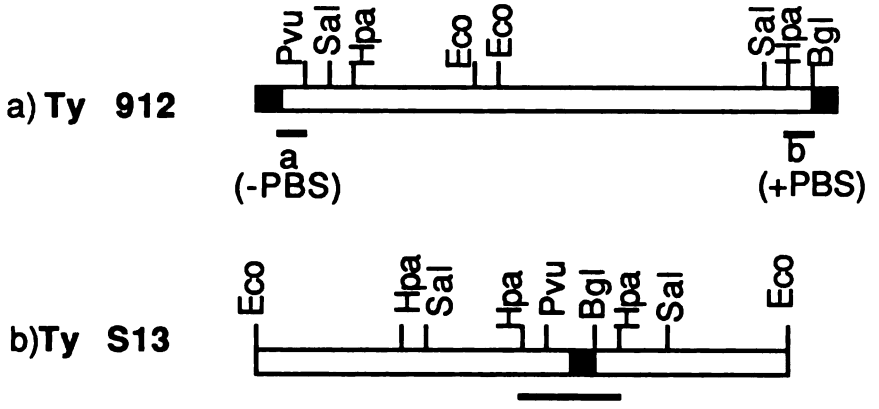
Figure 5 : Southern blot analysis of yeast genomic DNA using labelled oligonucleotides specific to delta-delta junction sequences. 3-5micrograms of yeast DNA was digested with either Xba1 or Xho II and fractionated on agarose gels as described in Materials & Methods. DNA immobilised on nitrocellulose filters was hybridised with labelled oligomers as described (35).

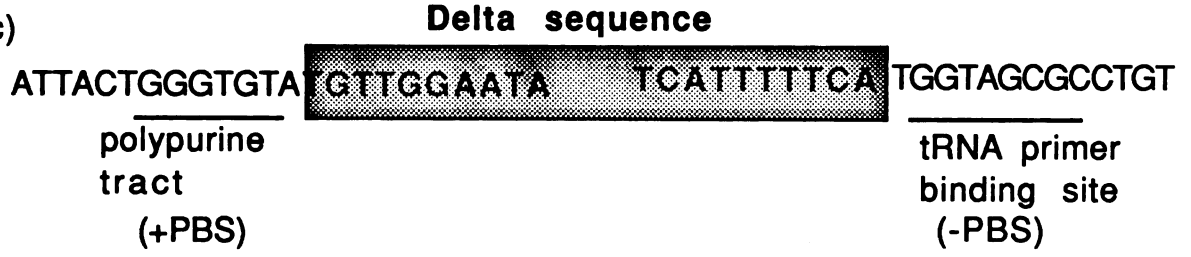
Figure 6 : Restriction map of lambda clone SC.7a (0-1b) containing

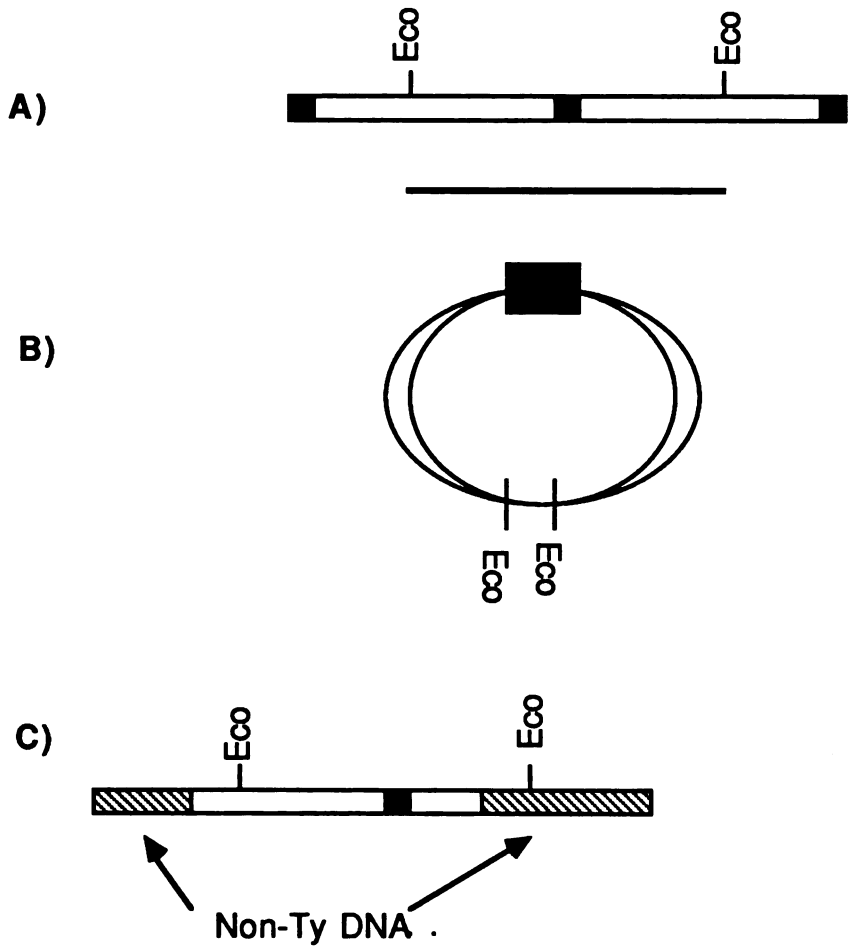
delta-delta junction in clone SC.7a. The delta -delta junction is contained within the approximateky 700bp Bgl II fragment at the 3' end of the Ty element as shown. Delta sequences are shown as filled squares.

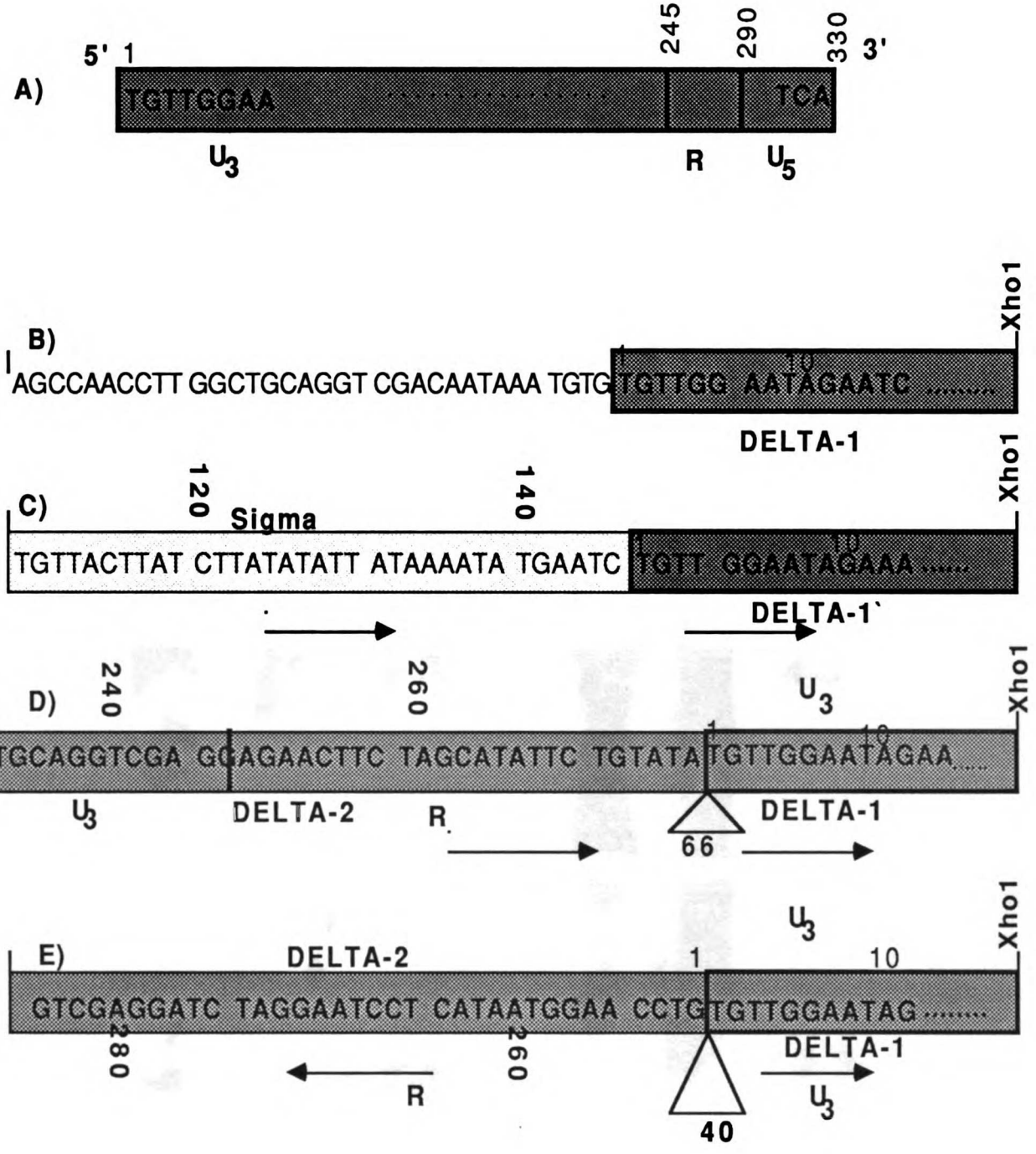
Figure 7 : Possible mechanisms to explain the formation of lambda SC.7a (0-1b) clone. (a) Integration of a Ty element into a solo delta in the yeast genome, 66 nucleotides from the 3 ' end of the delta. (b) homologous recombination between a Ty circle with a defective circle junction, and a free delta in the genome. The difference between these events can be resolved by analysis of the structure of the downstream delta. In the case of event (a), the 66nucleotides missing at the delta-delta junction should be present 3' to the end of the downstream delta (i.e. the downstream delta should contain an extra 66 nucleotides of delta sequence). If (b) is true, the downstream delta should not contain any extra delta sequence.

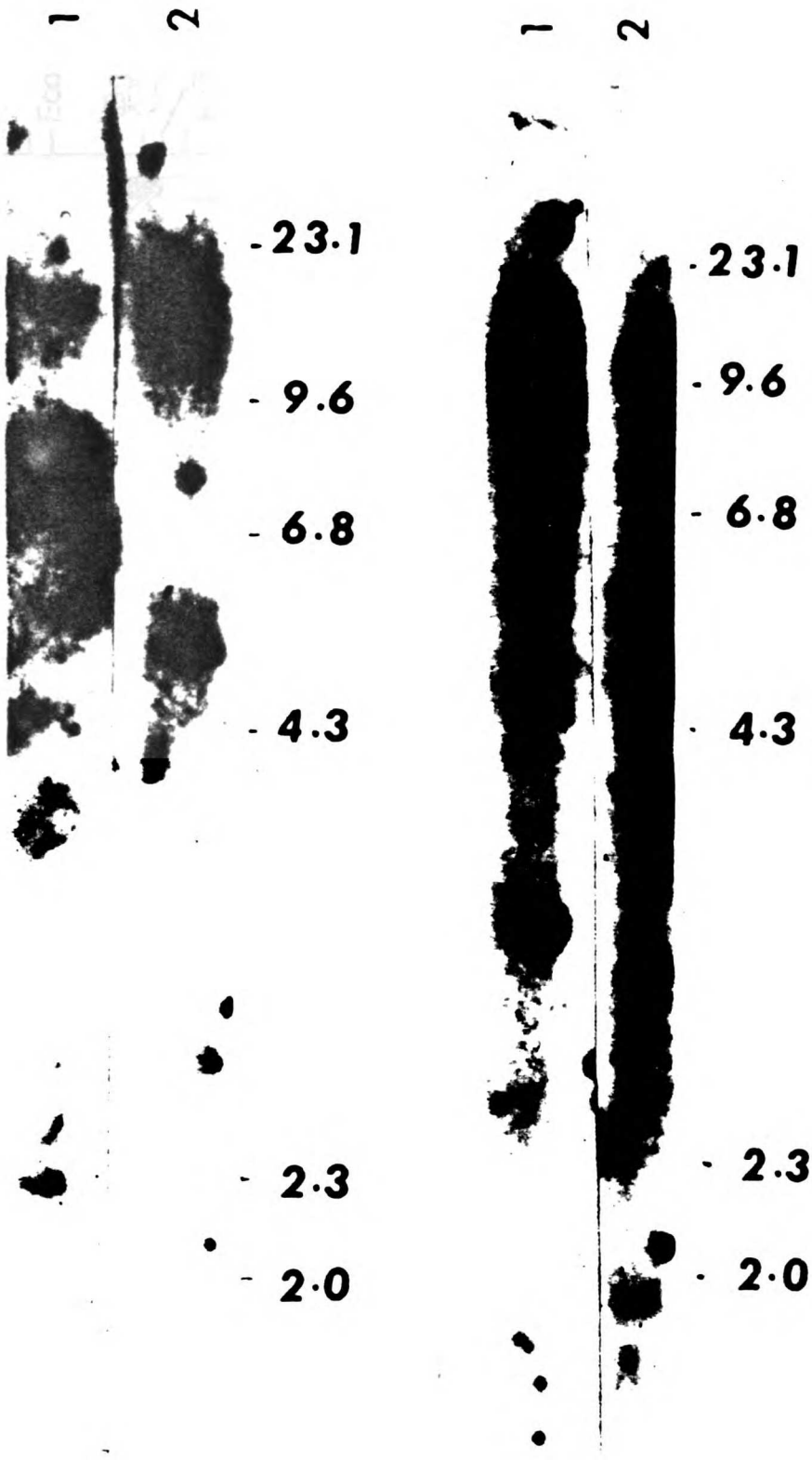

A
Recombination
Xho
Xho
B
Xho
Xho
Recombination
Xho
Xho
Xho

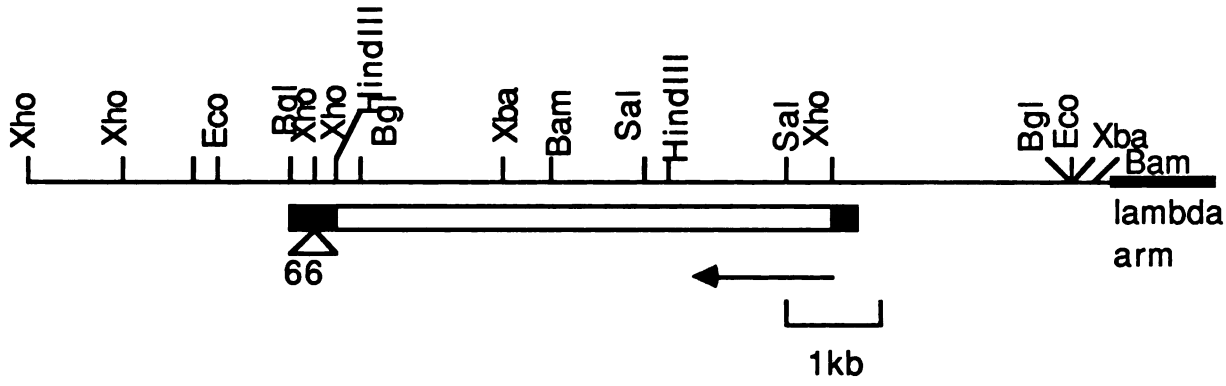

a) Ty 912
Pvu
Sal
Hpa
Eco
Eco
Sal
Hpa
Bgl
a
b
(-PBS)
(+PBS)
b)Ty S13
Eco
Hpa
Sal
Hpa
Pvu
Bgl
Hpa
Sal
Eco

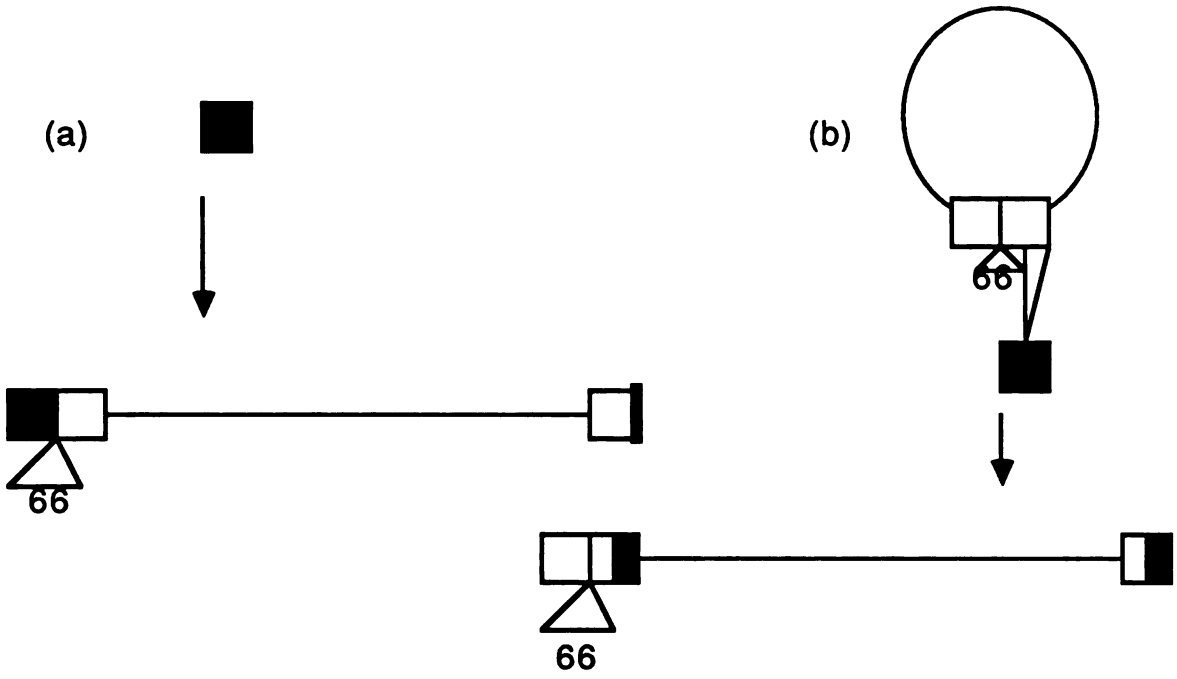

c)
Delta sequence
ATTACTGGGTGTA
TGTTGGAATA
TCATTTTTCA
TGGTAGCGCCTGT
polypurine tract (+PBS)
tRNA primer binding site (-PBS)


A)
Eco
Eco
B)
Eco
Eco
C)
Eco
Eco
Non-Ty DNA


A)
5' 1
245
290
330
3'
TGTTGGAA
TCA
U3
R
U5
B)
Xho1
AGCCAACCTT GGCTGCAGGT CGACAATAAA TGTG
TGTTGG AATAGAATC .....
DELTA-1
C)
120
Sigma
140
Xho1
TGTTACTTAT CTTATATATT ATAAAATA TGAATC
TGTT GGAATAGAAA .....
DELTA-1`
D)
240
260
U3
Xho1
TGCAGGTCGA GCAGAACTTC TAGCATATTC TGTATA TGTTGGAATAGAA.....
U3
DELTA-2
R
DELTA-1
66
E)
U3
Xho1
DELTA-2
1
10
GTCGAGGATC TAGGAATCCT CATAATGGAA CCTG TGTTGGAATAG .....
280
260
DELTA-1
R
U3
40


1
2
23.1
9.6
6.8
4.3
2.3
2.0
1
2
23.1
9.6
6.8
4.3
2.3
2.0


Xho
Xho
Eco
Bgl
Xho
Xho
HindIII
Bgl
Xba
Bam
Sal
HindIII
Sal
Xho
Bgl
Eco
Xba
Bam
lambda
arm
66
1kb


(a)
66
(b)
66
66

Structural Organisation, Transcriptional Properties, and Stage-Specific Expression of F-Elements in Drosophila melanogaster.

Kevin G.Mossie[1] and Harold E.Varmus[1,2]

Departments of Biochemistry and Biophysics[1],and Microbiology and Immunology

University of California San Francisco,San Francisco,CA 94143.

ABSTRACT:

We have detected F-specific transcripts in both cultured Drosophila cells and in flies. Analysis of these transcripts indicates that in cultured cells, F-specific transcripts are heterogeneous, non-polyadenylated, and predominantly located in the nucleus. In addition, discrete transcripts have been detected in both cytoplasmic RNA, and poly $(A)^+$ fractions prepared from whole cells. Analysis of F transcripts in flies indicate heterogeneous transcripts from both strands are present at various times during development; however, temporal regulation affects only the strand the strand that contains the oligoadenylate terminated end. We also describe the unusual association of F-element sequences with F-like ribosomal Type 1 insertion sequences.

INTRODUCTION:

The dispersed moderately repeated F- elements make up a distinct structural class of transposable elements in Drosophila (1,2). Present in approximately 30 copies per genome in flies, they differ from other well-characterised transposable elements including those belonging to the fold-back (3,4), P (5), and the copia families (6,7) in that their ends lack terminally repeated sequences characteristic of most transposable elements (see ref.8). Instead, the left or 3' ends of these elements are characterised by dAMP-rich stretches that have no homology with sequences at their 5' ends (1,2). In addition, most F-family members have heterogeneous and truncated 5' ends. Characterised F-family members were found to be flanked by 8-13 bp of host target sequence duplicated after integration. In this regard, F-elements resemble a class of mammalian transposable elements, termed the long interspersed elements, or LINES. These elements

are collectively designated the L-1 family, or LINE-1 (9,10).

In the L-1 family of transposable elements, of which the Bam or R type are rodent elements (11), and the Kpn elements are the human analogs (12), the hallmarks are truncated 5' ends and oligoadenylated 3' ends. These elements are flanked by 7-20 bp of target-site sequences that are duplicated on integration (12,13). These properties are also shared by mammalian short interspersed repeats (SINES), of which the human Alu is the prototype (14,15), and processed pseudogenes (reviewed in ref.16). It has been suggested, but not proven, that these elements transpose via an RNA intermediate(14-16). Recently, some L-1 elements in the mouse genome have been reported to have sequences with homology to several RNA-dependent DNA polymerases of viral and transposable element origin(17,18). Recently, structural analysis of the Drosophila I element (45) has revealed amino acid homologies between the I element and the mammalian LINE 1 elements, including the domain that has homology with the reverse transcriptase genes (45).

In addition to the F-elements and the recently-characterised I elements (45), the Drosophila genome has also been found to contain two major classes of ribosomal insertion-like elements which, though mainly associated with the ribosomal locus, also have properties of transposable elements (19-23). Like F-elements, these classes of non-homologous ribosomal insertions are long and lack terminal repeats at their ends. Type 1 ribosomal insertions (5kb) interrupt about 60% of rDNA units on the X-chromosome and are also found in chromocentric heterochromatin and in unique (euchromatic) DNA (22-24). They were found to generate 7-15bp duplications of host sequence (22-23), but unlike F-elements do not have an A-rich stretch at either end. Type 2 insertions (3.4kb), on the other hand,

interrupt approximately 75% of rDNA units on both X-and Y-chromosomes (19-22). It is believed that Type 2 insertions are restricted mostly, if not entirely, to the ribosomal locus. Although members have truncated left ends and A-rich right ends, they were found not to generate any target-site duplications (22,25). The two ribosomal insertions are also transcribed differently;while the predominant Type 2 transcript is of genomic length (3.4kb), and exclusively nuclear RNA (26),Type 1 elements are less abundantly transcribed. The transcripts were predominantly nuclear and heterogeneous (27), and some of these were reported to be derived from flanking rDNA. Furthermore, discrete 1kb cytoplasmic transcripts which also contained flanking rDNA sequences were observed (27).

Recently, another F-like element has been reported in Drosophila (28,29). This new class of transposable elements, termed the G-elements, was initially identified as an insertion of a repeated sequence into an F-element that had integrated into a Type 1 ribosomal insertion. Although G-elements resemble F-elements in that they lack terminal repeats, generate target-site duplications at the integration site and have one end that is oligoadenylated (29), they nonetheless appear to resemble the ribosomal insertion-like elements in the following properties: 1). they are predominantly found associated with ribosomal DNA and rDNA insertions, 2). they are found as tandem arrays interspersed with rDNA sequences, 3). they share homology in the site of integration in rDNA with ribosomal insertion-like elements, suggesting that insertion into rDNA may have occurred partly through a common pathway (29) and 4). they are restricted to the ribosomal locus and chromocentric heterochromatin.

In this report, we describe the association of F element sequences with the F-like ribosomal Type 1 insertion-like element. We also report for the the

transcriptional properties of F elements. The results of our analyses indicate that F elements are transcribed into heterogeneous, poly $(A)^-$ transcripts that are located predominantly in the nucleus in cultured Drosophila cells. In addition, we have observed smaller cytoplasmic, polyA- transcripts that hybridise with F element probes in these cells. Furthermore, we also present evidence that suggests that, although both strands of F elements are transcribed in flies, only the oligoadenylate-terminated strand is synthesised in a developmentally-regulated manner. We propose that this is the coding strand and discuss these observations in terms of parallels with the mammalian elements.

MATERIALS & METHODS:

a) DNA clones and subclones:

Clones containing dispersed moderately repetitive DNAs from D.melanogaster were a gift from M.Young (Rockefeller Univ.) and have been described (30,31). Lambda clone a19, containing a complete (4.7kb) F element (2) was supplied by I.Dawid. DNA clones containing Types 1 and 2 ribosomal insertion like elements, as well as the G element, were also kindly provided by I.Dawid. Additional genomic clones described in this paper were isolated from a library of D.melanogaster Canton S DNA in lambda Charon 4A (32) as described previously (33). Restriction fragments were subcloned in SP6 vectors (34).

b)Preparation of nucleic acids:

Genomic DNA was prepared from cultured cells and embryos as described previously (33). Stage-specific total RNA was prepared by the modification of the guanidium-thiocyante-CsCl method as described (36). Cytoplasmic, nuclear, and poly A^+ RNA fractions were prepared from cultured cells as

desribed (32).

c)Analysis of nucleic acids:

Nucleic acid preparations were fractionated on agarose gels and analysed as described previously (33). Strand-specific RNA probes were were prepared from restriction fragments subcloned into SP6 vector systems and labelled complementary RNAs prepared as recomended by the manufactures. Nick-translated probes were prepared as described previously (33).

RESULTS:

a) Association of the oligoadenylated (3') end of F-elements with sequences homologous to the Type 1 ribosomal insertion-like element:

One of the plasmid DNA clones we had analysed previously for the presence of unintegrated circular DNAs in cultured Drosophila cells and embryos contained F element sequences (33). This clone, pDm 151, orginally isolated by Dowsett & Young (31) had been shown to to be moderately repeated and dispersed in the Drosophila genome by in situ hybridisation (31). Examination of this clone, however, revealed that F-sequences were tightly associated with another moderately repeated sequence, which we tentatively termed the X repeat. The X repeat, which we later identified as a Type 1 ribosomal insertion-like element, was also present in other uncharacterised dispersed, moderately repeated DNA clones that were also described by Dowsett & Young (31), notably pDm 54, pDm 185, and pDm 298, even though these clones lacked any F-sequences (see Table 1). The structure and domains of F-element and the ribosomal insertion in pDm 151 is shown in Figure 1(b).

To determine if the association of F-element sequences with the F-like

Type 1 insertion were co-incidental, we used labelled pDm 151 to isolate additional genomic clones from the D.melanogaster Canton S library and screened these for hybridisation to both F-element sequences and the ribosomal Type 1 insertion sequences. The results of this screen are presented in Table 1.Three classes of clones were found: 1). Class 1 (7/18) are clones which contain ribosomal Type 1 insertion-like element sequences not associated with F-element sequences. 2). Class 2 clones (9/18) have F-element sequences only, and some of these show little or no hybridisation to what we tentatively designated as the 5' F-element probe (pl 27-see Fig.1(a)),suggesting that some of these clones contained truncated F-elements. 3). Class 3 clones contain F-element sequences which are linked to Type 1 insertion sequences. It should be pointed out though, that these numbers may not be statistically significant, since some of the clones may be identical, and we made no effort to characterise all of the clones in detail. We did, however, test some of these clones for the presence of rDNA and Type 2 insertion sequences, but the results were negative (see footnote to Table 1).

To determine if these elements were tightly linked as in pDm 151, we characterised one of these clones, lambda Dm F 0.12, in detail. The structure of this clone is shown in Figure 1(b). An approximately 1.8kb F-element has inserted into a 5kb ribosomal Type 1 insertion-like element, probably in the ribosomal locus, since the rest of flanking sequences were rDNA sequences. The site of insertion of the F-element into the Type 1 insertion is mapped near the left end of the element.

b)Distribution of F-specific transcripts in Cultured Drosophila Cells:

To determine the distribution of F-specific transcripts in Drosophila cells, we prepared RNA from cultured Drosophila cells as described in Materials &

Methods. Equivalent amounts of RNA prepared from cytoplasmic, nuclear, as well as polyA$^+$RNA were fractionated on formaldehyde-containing agarose gels, immobilised on nitrocellulose filters, and hybridised with F-specific probes. Figure 2A shows a result of these analyses. The majority of F-specific transcripts are predominantly located in the nucleus in these cells (lane 1). These nuclear transcripts are heterogeneous in nature, with sizes ranging from approximately 500bp to well over 5kb, the size of the longest F-element characterised to date (2). This heterogeneity is also observed when the polyA$^+$ fraction is analysed (lane 3). Although the amounts of F-transcripts are much reduced in this fraction (see legend to Figure 3A), evidently present within this fraction are discrete transcripts which are superimposed over heterogeneous F-transcripts. The sizes of some of these transcripts are also larger than the presumably full length 5kb F-element. Analysis of the cytoplasmic fraction reveals heterogeneous transcripts (lane 2), with strongest hybridisation at a position slighty below that of the 2kb 18S ribosomal RNA band. The position of the 18S RNA can be visualised in this and subsequent autoradiographs as a blank area due to competition for binding to the filter at this position. Similar observations have been noted with another F-like element (1).

d) F-transcripts during development:

For one F-like element, Dawid et al.(1), reported that element-specific transcripts were most abundant in Drosophila embryos. They also noted that both strands of that clone hybridised to RNA. To determine if the F-transcripts were limited to any one strand, we analysed stage-specific RNAs from D.melanogaster using strand specific RNA probes (34). To limit detection of transcripts from flanking unique sequences, we used an internal fragment (probe c-see Fig.1(a)), which had been subcloned into SP6 vector systems, as a probe. As can be noted in Figure 3, both strands of

probe c hybridise with F-specific sequences. The pattern of hybridisation, however, is different for both strands. Figure 3A shows the pattern of hybridisation when a strand that is the same as the oligoadenylate terminated F-strand (minus strand) is used as a probe. F-specific transcripts are not limited to any particular stage in development. The apparent peak at 12 hours is due to excess RNA loaded in this lane. Furthermore, high molecular weight F transcripts were observed with certain preparations when this strand is used as a probe (e.g. lanes 1,8). Figure 3B, however, suggests that transcription seen by this strand of F-elements may in fact be regulated. Using the strand that hybridises to the oligoadenylate-terminated F-strand (plus strand) as a probe, we noticed a different pattern. The levels of transcripts detected with this probe is lower than in the provious panel (Fig.3B). Detectable F transcripts do not accumulate until late in embryogenesis, with maximal transcription in mid-larval stages. The levels of RNA decrease in pupae and are barely detectable in adults. For this reason we propose that this is the F coding strand, and that F-specific transcripts have the same polarity as the oligoadenylate-terminating plus strand. Another point worth mentioning from the panels above is that the minus strand RNA is more abundant and predominantly larger. The high molecular weight transcripts seen at 0-3hr preparations, however, was not reproduced in subsequent preparations of RNA at this stage and may be artifactual (data not shown). Analysis of these RNAs with RNase A at low (0.1 X SSC) or high (2 X SSC), suggest that while high molecular weight RNAs are essentially single stranded, the majority of lower molecular weight RNAs exist in a double stranded form (data not shown),

DISCUSSION:

a) Scrambled arrangement of F and F-like elements :

The results of our analyses bring out few parallels between the properties of the Drosophila F elements and the mammalian sequences. We found that the 3' end of F elements is found associated with ribosomal Type 1 sequences (Figure 1 and also Table 1). The one clone we describe in detail here appears to be a case of an F element which had integrated into a ribosomal insertion within rDNA sequences. This association of F- and ribosomal Type 1 insertion is similar to that reported by Dawid et al.(1). The data presented in Table 1 have been confirmed and extended to include the G-elements by DiNocera et.al.(29). This scrambled arrangement of F-and F-like sequences in Drosophila, albeit restricted to the ribosomal locus, is similar to that observed by Wensik et.al.(37) and has also been observed for the human Kpn1 elements (38-41). Common sequences within some shorter Kpn 1 family members are not colinear with the longest Kpn 1 element, but are either reordered or inverted relative to each other (see Grimaldi et al.(13)). It is interesting to note that the w^{IR2} mutation in Drosophila arose on a chromosome that already carried the white mutation w^1 (46). This mutation is associated with an insertion of an F-like element (47-48). In the w^{IR2} mutation, the F-like I factor was also found inserted within the w^1 element (45). Whether F sequences co-transpose with F-like sequences can only be determined by careful analysis of putative cointegrate structures that are found not to associated with rDNA sequences. Finally, the scrambled arrangement of F and F-like elements is reminiscent of the scrambled arrangement observed with the mammalian L1 family members, which may be an intrinsic property of these sequences.

b)Transcription of F-elements and identification of the putative coding strand:

We have detected F-specific transcripts in cultured Drosophila cells. Our

results indicate that the majority of F-specific transcripts are not polyadenylated, are heterogeneous and are located predominantly in the nucleus (Figure 2A). We have also detected smaller (ca.1.8kb) cytoplasmic transcripts,which we do not believe are polyadenylated since we do not detect hybridisation at this position in the poly A^+ RNA fraction. Transcripts of the human Kpn 1 sequence were also reported to be heterogeneous, poly $(A)^-$, and predominantly nuclear (42,43). In addition, discrete, cytoplasmic transcripts that were smaller than the full-length element have also been observed (43). Similar properties are also exhibited by the Drosophila Type 1 ribosomal insertion element (27).

Our analyses have also revealed that although both strands of F-elements were transcribed, only one strand appeared to be transcriptionally-regulated during development (Figures 3A & 3B).From the orientation of our RNA probes we conclude that regulated transcripts had oligoadenylated ends (carried the A-rich stretch characteristic of F-elements). In this regard, therefore, F-elements closely resemble the mammalian L1 sequences, since DiGiovanni et.al.(41) also found that the dAMP-rich sequence in Kpn 1 cDNAs was on the same strand as the poly(A) sequence. Dawid et.al.(1) could not detect F-specific transcripts in Drosophila embryos.Our results, however, are not necessarily consistent with their observations since, in our hands F-transcription is maximal in late larval stages and minimal in embryos (Figure 3B). In this regard, F-elements are similar to other characterised retrotransposons in Drosophila that have been shown to behave as independently regulated transcriptional units and expressed at different times during development (see Spradling & Rubin,(44)).

The observation that the primate L1 family encodes proteins that have limited homology to reverse transcriptases has strengthened the belief that

these elements transpose via an RNA intermediate (17,18). Since human and primate LINES are reported to be 60-70% homologous (see refs.13,35), it is likely that homology in the putative pol region will extend to the human elements. Nucleotide sequence analysis of the I factor (45), is the first real indication that these elements also exist in Drosophila. Most significant were the reported homologies between I factors and the mammalian L1 elements (45). The results of our analyses clearly strengthen the parallels between F-elements and the mammalian L1 elements. Both sequences are transcribed into heterogeneous, poly(A)$^-$ and predominantly nuclear RNAs. Both elements also hybridise to smaller discrete cytoplasmic transcripts, a fraction of which is not polyadenylated. For Kpn 1 cytoplasmic transcripts, it was shown that none of these is associated with polyribosomes (43). The significance of these similarities is not clear.

REFERENCES:

1. Dawid, I.B., Long, E.O., DiNocera, P.P., and Pardue, M.L.(1981). Ribosomal insertion-like elements in Drosophila melanogaster are interspersed with mobile sequences. *Cell* 25, 399-408.

2. DiNocera, P.P., Digan, M.E., and Dawid,I.B.(1983). A family of oligoadenylate-terminated transposable elements in Drosophila melanogaster. *J.Mol.Biol.* 168, 715-727.

3. Potter, S.S., Truett, M., Phillips, M., and Maher, A.(1980). Eukaryotic transposable elements with inverted terminal repeats.*Cell* 20, 639-647.

4. Truett, M.A., Jones, R.S., and Potter, S.S.(1981). Unusual structure of the FB family of transposable elements in Drosophila

melanogaster.*Cell* 24,753-763.

5. O'Hare, K., and Rubin, G.M.(1984). Structures of P transposable elements and their sites of insertions and excision in the Drosophila melanogaster genome. *Cell* 34, 25-35.

6. Ilyin, Y.V., Chmeliauskaite, V.G., Ananiev, E.V., Lyubomirskaya, N., Kulguskin, V.V., Bayev, A.A., and Georgiev, G.P.(1980). Mobile dispersed genetic element MDG 1 of Drosophila mellanogaster: structural organisation. *Nucl.Acids Res.* 8, 5333-5346

7. Levis, R., Dunsmuir, P., and Rubin, G.M.(1980). Terminal repeats of the Drosophila transposable element copia : nucleotide sequence and genomic organisation. *Cell* 21, 581-588.

8. Rubin, G.M.(1983). Dispersed repetitive DNAs in Drosophila. In Mobile Genetic Elements (J.E.Shapiro, ed) pp.329-361.

9. Singer, M.F., Skowronski, J.(1985). Making sense out of LINES : Long interspersed repeat sequences in mammalian genomes. *Trends Biochem.Sci.* 10, 119-121.

10. Singer, M.F.(1982). Highly repeated sequences in mammalian genomes. *Int.Rev.Cytol.* 76, 67-112.

11. Soriano, P., Meunier-Rotival, M., and Bernardi, G.(1985). The distribution of interspersed repeats is nonuniform and conserved in mouse and human genomes. *Proc.Natl.Acad.Sci.U.S.A.* 80, 1816-1820.

12. Adams, J.W., Kaufman, R.E., Kretchmer, P.J., Harrison, M., and Nienhuis, A.W.(1980). A family of long reiterated DNA sequences,one copy of which is next to the human beta globin gene. *Nucl.Acids Res.* 8, 6113-6128.

13. Grimaldi, G., Skowronski, J., and Singer,M.F.(1984). Defining the beginning and the end of the Kpn 1 family segments. *Embo J.* 3, 1753-1759.

14. Jagadeswaren, P., Forget, B.G., and Weissman,S.M.(1981).

Short interspersed repetitive DNA elements in eukaryotes:transposable elements generated by reverse transcription of RNA polymerase III transcription? *Cell* 26, 141-142.

15. Van Arsdell, S.W., Denison, R.A., Bernstein, L.B., Weiner, A.M., Mancer, T., and Gesteland, R.F.(1981). Direct repeats flank three small nuclear RNA pseudogenes in the human genome. *Cell* 26, 11-17.

16. Singer, M.F.(1982). SINES and LINES : Highly repeated short and long interspersed sequences in mammalian genomes. *Cell* 28, 433-434.

17. Hattori, M., Kuhara, S., Takenaka, O., and Sakaki, Y.(1986). L1 family of repetitive DNA in primates may be derived from a sequence encoding a reverse transcriptase-related protein. *Nature (London)* 321, 624-628.

18. Loeb, D.D., Padgett, R.W., Hardies.S.C., Shehee, W.R., Comer, M.B., Edgell, M.H., and Hutchison, C.A.III(1986). The sequence of a large L1Md reveals a tandemly repeated 5' end and several features found in retrotransposons. *Mol.Cell.Biol.* 6, 168-182.

19. Wellauer, P.K., Dawid, I.B., and Tartoff, K.D.(1978). X and Y chromosomal ribosomal DNA segments of Drosophila: comparison of spacers and insertions. *Cell* 14, 269-278.

20. Kidd, S.J., and Glover, D.M.(1980). A DNA segment from D.melanogaster contains five tandemly repeating units homologous to the major rDNA insertion. *Cell* 19, 103-119.

21. Dawid, I.B., Wellauer, P.K., and Long, E.O.(1978). Ribosomal DNA in Drosophila melanogaster.I.Isolation and characterisation of cloned fragments.*J.Mol.Biol.* 126, 749-768.

22. Roiha, H., Miller, J.R., Woods, L.C.,and Glover, D.M.(1981). Arrangements and rearrangements of sequences flanking the two types of rDNA insertions in D.melanogaster. *Nature (London)* 290, 749-753.

23. Rae, P.M., Kohorn, B.D., and Wade, R.P.(1980). The 10kb Drosophila virilis 28S rDNA intervening sequence is flanked by a direct repeat of 14bp of coding sequence. *Nucl.Acids Res.* 8, 3491-3505.

24. Dawid, I.B., and Botchan, P.(1977). Sequences homologous to ribosomal insertions occur in the Drosophila genome outside the nucleolus organiser. *Proc.Natl.Acad.Sci.U.S.A.* 74, 4233-4237.

25. Roiha, H., and Glover, D.M.(1980). Characterisation of a complete Type II insertions in cloned DNA segments from Drosophila melanogaster. *J.Mol.Biol.* 140, 341-355.

26. Kidd, S.J., and Glover, D.M.(1981). Drosophila melanogaster DNA containing Type II insertions is variably expressed in different strains and tissues. *J.Mol.Biol.* 151, 645-662.

27. Long, E.O., and Dawid, I.B.(1979). Expression of ribosomal insertions in Drosophila melanogaster. *Cell* 18, 1185-1196.

28. DiNocera, P.P., and Dawid, I.B.(1983). *Nucl.Acids Res.*11,5475-5482.

29. DiNocera, P.P., Graziani, F., and Lavorgna, G.(1986). Genomic and structural organisation of D.melanogaster G elements. *Nucl.Acids Res.* 14, 675-691.

30. Young, M.W.(1979). Middle repetitive DNA:A fluid component of the Drosophila genome. *Proc.Natl.Acad.Sci.U.S.A.* 76, 6274-6278.31

31. Dowsett, A.P., and Young, M.W.(1982). Differing levels of dispersed repetitive DNA among closely-related species of Drosophila. *Proc.Natl.Acad.Sci.U.S.A.* 79, 4570-4574.

32. Maniatis, T., Fritsch, E.F., and Sambrook, J.(1982). Molecular cloning-A laboratory manual. Cold Spring Harbor Laboratory.

33. Mossie, K.G., Young, M.W., and Varmus, H.E.(1985). Extrachromosomal DNA forms of copia-like transposable elements, F-elements, and middle repetitive DNA sequences in Drosophila

melenogaster. *J.Mol.Biol.* 182, 31-43.

34. Green, M.R., Maniatis, T., and Melton, D.A.(1983). Human beta-globin pre-mRNA synthesised in vitro is accurately spliced in Xenopus oocyte nuclei. *Cell* 32, 681-694.

35. Rogers, J.H.(1985). The structure and evolution of retroposons. *Int.Rev.Cytol.* 93 , 231-279.

36. Wensik, P., Tabata, S., and Pachi, C.(1979). The clustered and scrambled arrangement of moderately repetitive elements in Drosophila DNA. *Cell* 18, 1231-1246.

37. Potter, S.S.(1984). Rearranged sequences of a human Kpn 1 element. *Proc.Natl.Acad.Sci.U.S.A.* 81, 1012-1016.

38. Gerbhard, W., and Zachau, H.G.(1983). Organisation of the R family and other interspersed repetitive DNA sequences in the mouse genome. *J.Mol.Biol.* 170, 255-270.

39. Thayer, R.E., and Singer, M.F.(1983). Interruption of an alpha-satellite array by a short member of the Kpn 1 family of interspersed,highly repeated monkey DNA sequences. *Mol.Cell.Biol.* 3, 967-973.

40. Lerman, M.I., Thayer, R.E., and Singer, M.F.(1983). Kpn 1 family of long interspersed repeated DNA sequences in primates: polymorphism of family members and evidence for transcription. *Proc.Natl.Acad.Sci.U.S.A.* 80 , 3966-3970.

41. DiGiovanni, L., Haynes, S.R., Misra, R., and Jelinek, W.R.(1983). Kpn 1 family of long repeated DNA of man:evidence for entry into genomic DNA of DNA copies of poly(A)-terminated Kpn 1 RNAs.*Proc.Natl.Acad.Sci.U.S.A.* 80 ,6533-6537.

42. Shafit-Zagardo, B., Brown, F.L., Zavoyn, P.J., and Maio, J.J.(1983). Transcription of the Kpn 1 families oflong interspersed DNAs in human cells. *Nature (London)* 307 , 277-280.

43. Kole, L.B., Haynes, S.R., and Jelinek, W.R.(1983). Discrete and heterogenous high molecular weight RNAs complementary to a long dispersed repeat family (a possible transposon) of human DNA. *J.Mol.Biol. 165* , 257-286.

44. Spradling, A.C, and Rubin, G.M.(1981). Drosophila genome organisation:conserved and dynamic aspects. *Annu.Rev.Genet. 15* , 219-264.

45. Fawcett, D.H., Lister,C.K., Kellet, E., and Finnegan, D.H.(1986). Transposable elements controlling I-R hybrid dysgenesis in D.melanogaster are similar to mammalian LINES. *Cell 47* , 1007-1015.

46. Sang, H.M., Pelisson, A., and Finnegan, D.H.(1984). Molecular lesions associated with white gene mutations induced by I-R hybrid dysgenesis in Drosophila melanogaster. *EMBO J. 3,* 3079-3085.

47. Zachar, Z., and Bingham, P.(1982). Regulation of white locus expression : the structure of mutant alleles at the white locus of Drosophila melanogaster. *Cell 30*, 529-541.

48. O'Hare, K., Levis, R., and Rubin, G.M.(1983). Transcription of the white locus in Drosophila melanogaster. *Proc.Natl.Acad.Sci.U.S.A. 80,* 6917-6921

FIGURE LEGENDS:

Figure 1: Physical maps of cloned DNAs described in this study. (A). A partial restriction map of a bacteriophage clone, lambda a19, containing the longest characterised F-element (DiNocera et.al. (J.Mol.Biol.168,715(1983)). (a) A putative F-transcript showing the position of the dAMP-rich end. (b) F subclones pl 25, and pl 27 which contain approximately the right and left half of F element sequences, plus flanking DNA, respectively. (c) The EcoR1/Bgl II fragment which was used as an F-specific probe (probe c). (B). Plasmid clone pDm 151, and phage clone lambda DmF0.12 containing both F and Type 1 ribosomal insertion sequences. The domains of repetitive DNA sequences are shown in the lay 1.Restriction sites are as follows : H, Hind III; R, Eco R1; B; Bam H1; Bg, Bgl II; X, Xho 1; S, Sac 1; P, Pst 1; S; Sal 1; Hh, Hha 1.

Figure 2: Distribition of F-transcripts in cultured Drosophila cells. Total cytoplasmic, nuclear, as well as total poly A^+RNA were prepared from Kc cells as described in Materials & Methods. The amounts of RNA loaded per lane were 50 micrograms for nuclear RNA and 30 micrograms each for both the cytoplasmic and poly A^+ RNA lanes. The size markers were derived from migration of ribosomal and copia RNAs that were used in these analyses as internal controls. Nick-translated probe were prepared using 3'F-(pl 25) sequences (see Figure 1A). Lane 1, total nuclear RNA; lane 2, total cytoplasmic RNA; lane 3, total poly $(A)^+$ RNA.

Figure 3: Stage- and strand-specific expressionof F-elements. RNA was prepared from 0.5-3 grams of tissue collected at different times during development. Total RNA was prepared from tissue powder by a modification of the guanidium thiocyanate-CsCl method and analysed as

described in Materials and Methods. Each lane contains approximately 100-150 micrograms of total RNA. Size markers were derived from migration positionsof ribosomal RNA,which can also be visualised as a blank area at the position of the 2.0kb RNA. The 2.3kb Eco/Bg fragment (probe c) subcloned in SP6 vector systems was used to generate labelled RNA probes. The RNA probes are (A) sense-RNA probe is the same strand as the oligoadenylate-terminated strand, and (B) the RNA probe is complementary to the oligoadenylate-terminated strand. (C) control hybridisation using the ribosomal protein (rp41) gene. Nick-translated rp41 sequences were used to hybridise to the filters shown above. The lanes are (1)0-3hrs; (2)3-6hrs; (3)6-12hrs; (4)12-20hrs; (5)18-24; (6)20-26hrs; (7)26-56hrs; (8)56-72hrs; (9)5days(larvae); (10)6-8days(late larvae-early pupae); (11)12-14days (adults).

Table 1: Analysis of genomic clones homologous to pDm 151

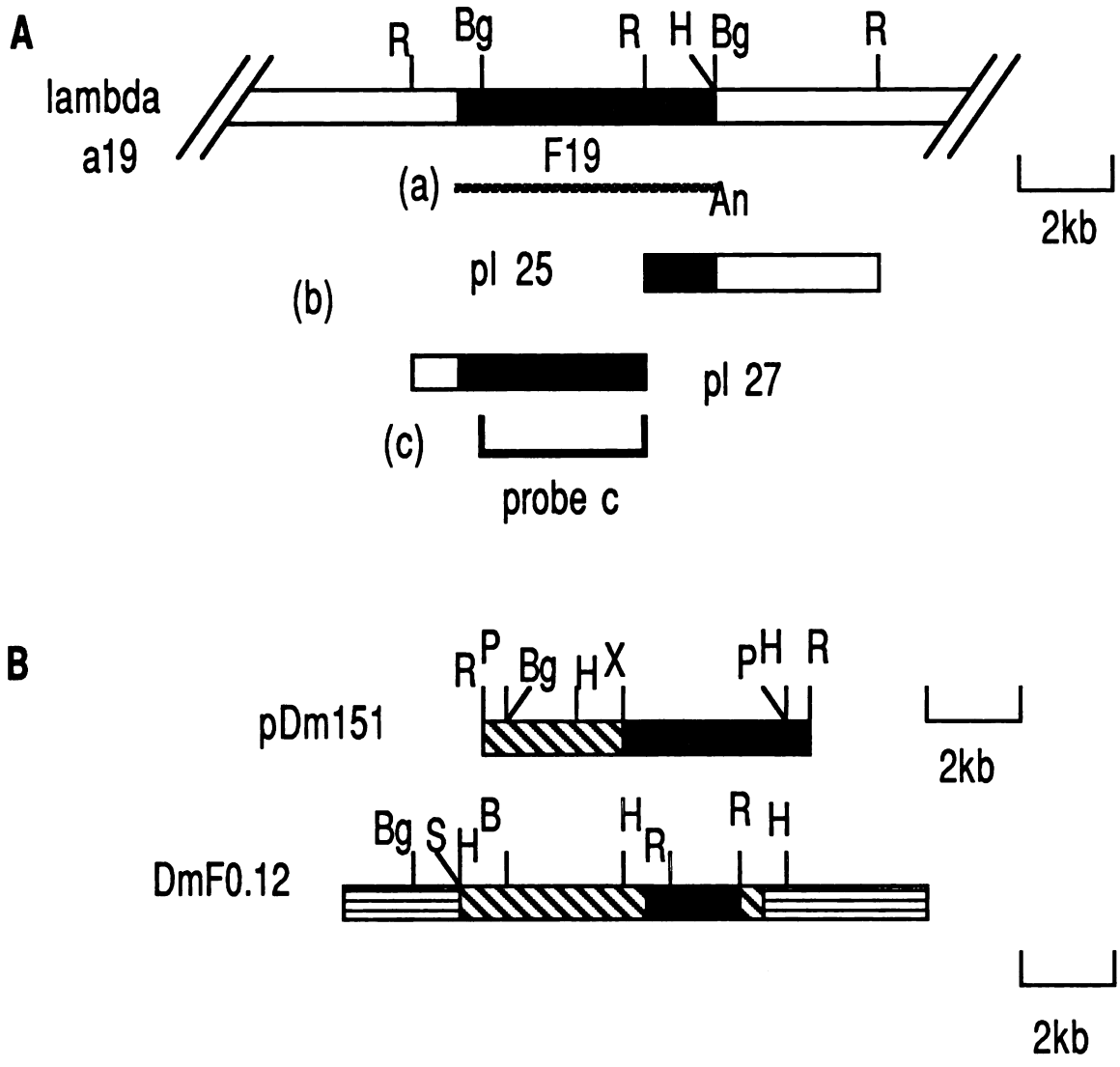
Isolate #	Homology to 5' F-(pl 27)	3'F-(pl 25)	pDm 54(Type 1)	Class
0.02	-	-	++	1
0.03	-	-	++	1
0.05	-	-	+	1[a]
0.06	-	-	+	1
0.07	-	[+	++]	2[a]
0.09	-	+	-	3[a]
0.10	-	+	-	3
0.11	++	+++	-	3[a]
0.12	+	[+++	++]	2[b]
0.13	+	+++	-	3
0.14	++	+++	-	3
0.15	-	[+	++]	2[a]
0.16	-	[+	++]	2
0.17	-	+	-	3
0.18	-	+	-	3
0.19	-	[++	++]	2[a]
0.20	-	[++	+]	2
0.21	-	[++	+]	2

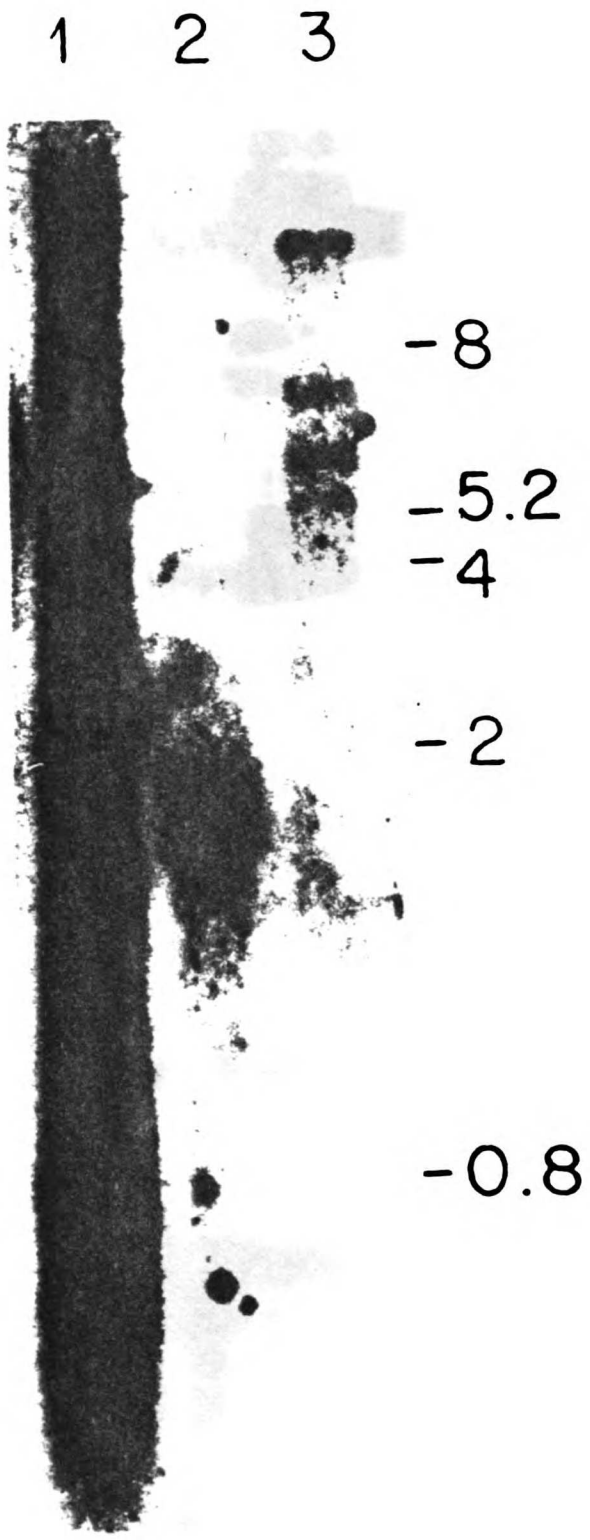
Genomic clones 0.02-0.21 were isolated from the Canton S library as described in Materials and Methods. Replica filters were initially

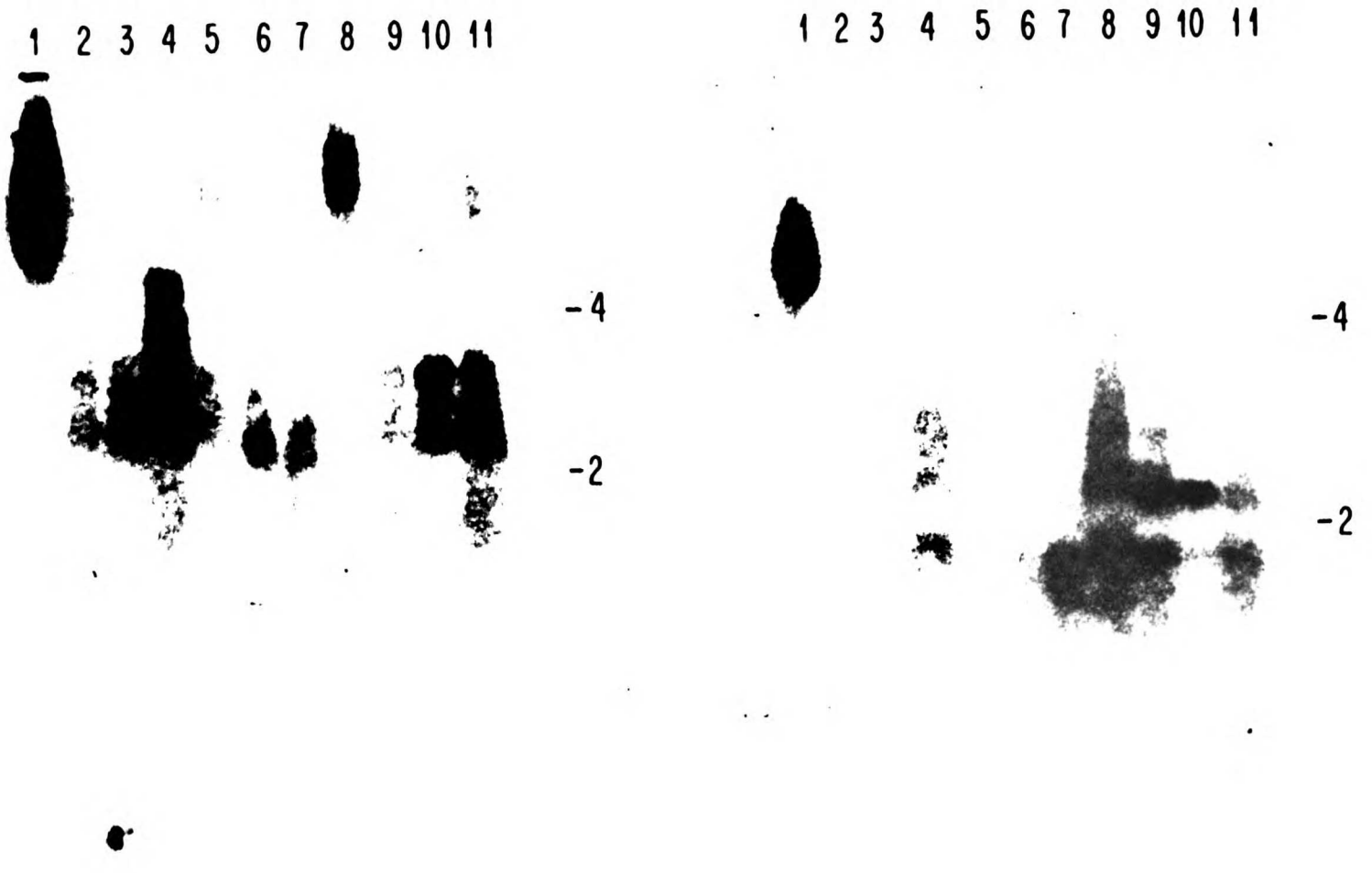
hybridised with nick-translated pDm 151 probes as described previously (33), and the position of positive plaques were marked. Filters were then melted, and each filter rehybridised with nick-translated pl 25, pl 27, and pDm 54. Signals: (+) = hybridisation, and (-) = no hybridisation.

Footnote: a. Tested for cross-reactivity with with rDNA and Type 2 insertion probes

b. Tested for reactivity with Type 2 insertion probes


A
R
Bg
R
H
Bg
R
lambda
a19
F19
(a)
An
2kb
pl 25
(b)
pl 27
(c)
probe c
B
R
P
Bg
H
X
pH
R
pDm151
2kb
Bg
S
H
B
H
R
R
H
DmF0.12
2kb
Flanking cellular DNA
F-Element sequences
Type 1 ribosomal
insertion
rDNA


1
2
3
-8
-5.2
-4
-2
-0.8


1 2 3 4 5 6 7 8 9 10 11
-4
-2
1 2 3 4 5 6 7 8 9 10 11
-4
-2
97
1 2 3 4 5 6 7 8 9 10 11

APPENDIX

Construction of a *copia*-based vector to monitor tranposition in *Drosophila*

ABSTRACT :

We describe the partial construction and characterisation of a copia-based 'hopping' vector using as a selectable marker, the neomycin phosphotransferase gene. The theoretical basis for the construction is the observation that newly transposed retrotransposons lose introduced intron s (Boeke et.al. 1985). We have engineered a selection for transposition by introducing an intron within the structural gene of the selectable marker. After transfection into cells. Newly-transposed copia elements, which had lost the intervening intron sequences during reverse transcription, could then be selected for by plating cells on media containing G418.

INTRODUCTION :

There is now substantial evidence to suggest that copia and other retroviral-like transposable elements in Drosophila transpose via an RNA intermediate (see Chapter 1- Introduction). The transposition pathway has been inferred from the identification of putative tranposition intermediates (see also Chapters 2 & 3), and is generally believed to be as shown in Figure 1. Details of this pathway are not as well-defined as with the case of retroviral replication, and no direct evidence yet exists for this pathway. The main problem facing investigators studying the transposition pathway

of these elements in Drosophila is the low transposition frequency. Furthermore, since these elements are present in multiple copies per genome in flies, it is also desirable to be able to mark a particular element, and then follow its movement in the genome by following the marker. Even with these modifications, the low transposition frequency of these elements still makes the experiment difficult unless a selection can somehow be engineered to mark the movement of the element, e.g., drug resistance.

A direct assay for transposition has recently been described in yeast by Boeke and his co-workers (1). These investigators took advantage of observations by Scherer et.al.(2) that a high proportion (up to 80%) of His^+ revertants obtained from a promoter deletion mutant of the HIS 3 gene (his3-4) were Ty insertions in the upstream region of the gene. By simply selecting for His^+ revertants, therefore, they could then follow the movements of Ty elements into the upstream region of his3-4. Using this transposition assay, they were able to demonstrate by marking a particular Ty element with intron fragments and mutations in the U3 and U_5 regions, that transposition was not only dependent on the transcription of that particular element, but also that it occurred via an RNA intermediate (1). Garfinkel et.al. (3), and Mellor et.al. (4), subsequently showed that Ty

transposition, after induction, was closely associated with the appearance of intracellular particles in yeast. The success of these experiments in yeast was therefore largely due to the availability of two important tools essential in studying transposition : ability to regulate transposition and a reasonably good selection scheme to monitor the transposition event.

Motivated by the success of this approach in yeast, we sought to attempt similar experiments in Drosophila, using as a choice retrotransposon, the copia transposable element. The copia clone used in these constructs had recently (within the last 50 years) integrated into the white locus causing the white-apricot (w^a) phenotype, and has been recently sequenced (5). Nucleotide sequence analysis indicate that the element is probably functional, since it has a conserved long open reading frame capable of encoding a protein of approximately 1400 amino acids (5, 6) (see Figure 2A). Furthermore, S1 mapping studies of the white-apricot mutation indicated that the element was also transcribed (7,8).

Our construction scheme is shown in Figure 2. Since no hot-spots for copia insertions have been conclusively identified, we chose to take advantage of the observation that intron segments placed into Ty elements are efficiently spliced during the transposition process (1). We have inserted

within the copia element, a neomycin phosphotransferase gene, which is driven by the Drosophila heat-shock promoter (hsp-neo) and has recently been shown to work well in Drosophila (9) Figure (2B). The orientation of the neo gene in the test construct is opposite to that of the copia major transcript (Figure 2C). The neo gene itself is interrupted by a short cuticle protein intron (10) whose orientation is the same as the copia mRNA, but opposite relative to the neo transcript. In this configuration, therefore, the neo gene is inactive and can only be functional if the cuticle protein intron is removed by a splicing event, followed by re-integration into the genome of a spliced cDNA copy of the copia/neo construct (Figures 2C-2G).

MATERIALS & METHODS:

Bacteial Strains:

E.coli HB101 (11) and JM101(12) were a gift from C.Nottenberg.

DNA Clones:

Phage and plasmid clones used in this study are described in the text. The 21-mer oligonucleotide used for mutagenesis were purchased from the facility, UCSF, San Francisco, CA.

Materials:

E.coli DNA Polymerase (large fragment), T4 DNA ligase, exonuclease III, and all the restriction endonucleases used were purchased from New England Biolabs. Labelled nucleotides were obtained from Armesham. Cold nucleotides were obtained from P-L Biochemicals. Nitrocellulose filters were purchased from Schleicher & Schuell.

Media:

The following media were used: 2X YT: 16g tryptone, 5g yeast extract, 5g NaCl per liter. YT plates: 8g tryptone, 5g yeast extract, 5g NaCl, and 15g agarose per liter. YT soft agar: same as YT agar except that the concentration of agarose is 0.8%. M9-supplemented with B1: 1X M9 salts (12), plus 0.2% glucose,1mM $MgSO_4$, 0.0001% thiamin-HCl, 0.4% casamino acids, and 0.1mM CaCl.

Transformation of JM101:

Competent JM101 cells were prepared by a modification of a protocol developed by Hanahan. 0.5ml of overnight culture was used to inoculate 100mls of YT media. Cells were grown at 37^{o}C to mid-log phase (approximately 2hours). Cells were then harvested and resuspended in 20mls Tbf I (30mM potassium acetate, 50mM $MnCl_2$, 100mM KCl, 10mM

$CaCl_2$,15%(w/v) glycerol, adjusted to pH 5.8 with 0.2M acetic acid), and left on ice for 60-90 minutes. Cells were harvested at low speed, and resuspended in 5mls Tbf II (10mM Na-MOPS, pH 7.0, 75mM CaCl, 10mM KCl, and 15%(w/v) glycerol). 200 microliter aliquots were dispensed into pre-cooled eppendorf tubes on ice, and stored at -70°C until use. To transform competent cells, 10-20 microliters of DNA was added to the cells, and the mixture kept on ice for at least 30 minutes. Cells were then heat shocked at 42°C for 2 minutes,and added to 3mls of top agar, and the mixture used to overlay YT agar plates. Plates were incubated overnight at 37°C overnight. Competent HB101 cells were also similarly prepared. Competent cells were transformed by adding 20-30 microliter of cells to 10 microliter of DNA, samples kept on ice for 15-20 minutes, and heat-shocked at 37° for 2 minutes, or at 42° for 1 minute, before being kept on ice for 1-2 minutes.100-200 microliter of YT broth are added before plating cells on antibiotic-containing plates.

5'Phosphorylation of Mutagenic Oligonucleotides:

A. **For mutagenesis:** Approximately 200 picomoles of mutagenic oligonucleotides were mixed with 2 microliters of 10X kinase buffer (0.5M Tris.Cl (pH 7.6), 0.1M $MgCl_2$, 50mM dithiothreitol, 1mM spermidine, 1mM EDTA), 1 microliter 10mM rATP, and 4 units of T4 polynucleotide kinase, in

a total volume of 20 microliters. The reaction was carried out at 37°C for 1 hr and terminated by heating at 65°C for 10 minutes.

B. <u>For Hybridisation:</u> The mutagenic oligonicleotide was phosphorylated as in A above with 50 microcurries of gamma-^{32}P rATP (3000Ci/mmol) as sole source of ATP. Labelled oligonucleotide was seperated from unreacted ^{32}P-rATP by chromatography on Whatman DE-52 cellulose. After the kinase labelling reaction, the reaction mixture was diluted with 0.5mls 10mM Tris.Cl (pH 7.5), 1mM EDTA, 100mM NaCl, and applied to a small DE-52 column equilibrated with the same buffer. The column was washed with 10 volumes of the same buffer, and the oligonucleotide was eluted with 3 X 0.5mls 10mM Tris.Cl, 1mM EDTA, and 800mM NaCl. The labelled oligonucleotide was used directly for hybridisation.

<u>Preparation of DNAs:</u>

Plasmid DNAs were prepared using standard procedures (11). M13 single-stranded and RF DNAs prepared as described by Zoller & Smith (13).

<u>Oligonucleotide-Directed Mutagenesis:</u>

Plasmid DNA was linearised by didestion with appropriate restriction enzymes (see text). Linear DNA was purified by electroelution from

agarose gels, phenol extracted once, and ethanol precipitated. DNA was resuspended in 10mM Tris.Cl (ph 7.5), 1mM dithiothreitol, 2mM $MgCl_2$, at approximately 100 nanograms per ml (total volume 0.1mls). 60-100 units of exonuclease III were added and the mixture incubated at room temperature for 60 minutes. The reaction was stopped by the addition of a drop of chloroform, and the mixture heated at 65°C for 15 minutes to evaporate the chloroform. 20 microliters of the reaction was removed, and to it added 2.2 microliters 1M Tris.Cl (pH 7.5),and 2.2 microliters of 0.1M $MgCl_2$. 50-70 picomoles of the 5' phosphorylated oligonucleotide were added. The samples were then incubated at 55°C for 10 minutes, and then at room temperature for 5 minutes to anneal the oligomer. 10 microliters of the reaction was added to 10 microliters of 100 micromolar solution containing all four deoxyribonucleotides,0.4mM rATP, 2.5 microliters 10 X ligase salts, and 15 units of the Klenow fragment of E.coli DNA Polymerase, and 1000 units of T4 DNA ligase. The reaction was allowed to incubate at 14°C overnight.

Transformation and Screening for Mutant Clones:

The DNA from the oligonucleotide-primed reactions was used to transform competent HB101 cells as described earlier. Transformed bacteria were plated onto nitrocellulose filters placed on 100mM YT-agar plates

containing 50 micrograms per milliter ampicillin, and incubated at 37°C overnight. Colonies were replica-plated onto a second set of nitrocellulose filters and the replica filters incubated on ampicillin-containing plates overnight. Filters were prepared for hybridisation as described by Hanahan & Meselson (15), and prehybridised in 6 X SSC (1 X SSC is 0.15M NaCl, 0.015M sodium citrate) containing 0.2% SDS and 10 X Denhardts mixture (0.2% bovine serum albumin, 0.2% polyvinyl pyrrollidone, 0.2% Ficoll), at 65°C for 1 hour. Filters were then rinsed with 6 X SSC, and then hybridised with labelled oligonucleotides in 6 X SSC, 10 X Denhardts at room temperature overnight. Filters were then washed in 3 changes of 6 X SSC at room temperature for a total of 20 minutes, air dried, and exposed to Kodak XR-5 X-ray film at -70°C for three hours. Sucessive washes in 6 X SSC at 37, 42, and 50°C, respectively, for 5 minutes were performed, and each time the filters were exposed to X-ray film as described previously.

Plasmid DNA was prepared from putative mutant colonies, and this was used to transform fresh HB101 cells. Colonies were re-screened with labelled oligomers as described previously, and positive colonies were prepared for subsequent cloning steps.

Plasmid DNA Transfection:

Copia-based vectors containing the neo (neomycin phosphotransferase) gene were used to transfect Drosophila Schneider cells, using the calcium-phosphate precipitation method (14). Transfectants were selected by plating the cells in Schneider medium , containing microgram per milliliter G418, and cells were grown at room temperature.

RESULTS:

a) Generation of p^{a8}-neo constructs:

The scheme used to clone the neo gene into the copia element is shown in Figure 3. Plasmid p^{a8} is linearised by digesting with Hpa 1 and the ends were dephosphorylated by treating with Bacterial Alkaline Phosphatase (BAP). The neo fragment was prepared by digesting plasmid hsp-neo with Eco R1 and Sma 1. The Eco R1 ends were repaired with Klenow and dATP and dTTP. The 1.3kb blunt ended fragment was then cloned into the Hpa 1 site in p^{a8} by ligating with T4 DNA ligase, generating $p^{a8neo\#3}$ and $p^{a8\#4}$, repectively. The orientation of the neo insert was determined by digesting recombinant clones with Pst 1 (data not shown).

b) DNA Transformation with p^{a8}-neo Constructs:

To determine if the neo gene was functional within the copia element in

either orientation, we used $p^{a8\#3}$ and $p^{a8\#4}$ DNAs to transfect Drosophila Schneider line 2 cells. Poly $(A)^+$ RNAs were prepared from a pool of G418-resistant transformants and analysed for both copia and neo sequences as described in Materials and Methods. The results of these analyses is shown in Figure 4. Using nick-translated copia as a probe, we detected the expected 5.2 and 2.0kb copia RNAs (Figure 4A). We could not detect, in RNA from the transformants the approximately 1.3kb neo transcript, or the larger (6.5kb) copia/neo transcript when copia is used as a probe. This was to be expected since there are 100 or so copies of the copia elements in cultured Drosophila cells, which are highly transcribed, making the detection of the putative p^{a8} transcript difficult with a copia probe. Using strand-specific probes specific for the neo we could detect both neo and copia/neo transcripts. When the strand complementary to the copia RNA strand was used as a probe, we saw the expected sized species in $p^{a8\#3}$-transformed cells; bands representing the large (6.5kb) copia/neo transcript, a transcript migrating at approximately 3.3kb, and the 1.3kb neo transcript (see arrows, Figure 4B). The extent of hybridisation of the 6.5kb and 3.3kb transcripts suggests that they may be related to the 5.2kb and 2.0kb copia transcripts in Figure 4A. We also observed an unexpectedly strong transcript migrating at the position of the 2.0kb RNA (Fig.4B). This transcript is seen in both constructs when the neo

gene is used as a probe. The nature of this transcript is unknown, although it is possible that this may be due to background hybridisation on the filter from copia probes which were initially used. No hybridisation is observed when the opposite strand was used as probe (Figure 4C). In $p^{a8\#4}$-transformed cells, on the other hand, we saw both the large (6.5kb) copia/neo transcript, and the 2kb copia RNA species. When the opposite strand is used as a probe, we only saw the expected 1.3kb neo transcript. Taken together, these observations indicate that the neo gene is functional within the copia element in an orientation-independent manner.

c) **Preparation of Intron Fragment for cloning:**

We chose to use the second cuticle protein (CP II) intron (10) to clone into the neo gene in p^{a8} vectors. This short (60bp) intron is contained within a 600bp Bam H1/Sal 1 fragment in pCP II-7 (10) as shown in Figure 5A. The nucleotide sequence at the intron-exon junctions is also shown (Figure 5B). Note that the strand presented in Fig. 5B is the antisense strand (as judged by sequences complementary to the consensus GT...AG motif at the internal intron boundaries. There is a fortuitous Mnl 1 site that cuts this sequence at the left intron-exon boundary, as shown in Fig. 5B. To generate an intron fragment, we chose to generate an Mnl 1 site at the right intron exon boundary. To achieve this, we synthesised a 22bp mutagenic

oligonucleotide (CACAAACTTGGAGGTGTTGGCT) that is homologous to sequences at the right intron-exon boundary (Fig. 5B). The Mnl 1 recognition sequence to be generated and the nucleotides changed are illustrated in Fig. 5B.

The Bam/Sal fragment from pCP II-7 was subloned in pBR322. To generate the 60bp Mnl 1 intron fragment, we linearised pBR322(CP II-Bam/Sal) by digestion with Sal 1. Single-stranded regions were generated by limited digestion with exonuclease III. The mutagenic oligonucleotide was annealed to exposed complementary single stranded regions, and gaps were repaired with Klenow fragment of DNA polymerase 1, and blunt-ended ends ligated with T4 DNA ligase, as shown in Figure 4C. Plasmid DNA was used to transform competent HB101 cells, and recombinants were screened as described in Materials & Methods. Four recombinant clones (pBR322(CP II-Bam/Sal -Mnl1 #^{s}1-4)) were picked and plasmid DNAs prepared from them for the next cloning step.

d) <u>Cloning the intron fragment in p^{a8}-neo constructs</u>:

Since the intron fragment is generated as a blunt-ended fragment, we chose to clone it into the Pvu II site in p^{a8}-neo constructs. Pvu II offers a number of advantages : a) it is located fairly early into the <u>neo</u> gene such

that insertions here should inactivate the gene (see Figure 6A). b) it offers an advantage in screening the orientation of the intron insert in recombinant clones, since the Pvu II site is regenerated in both orientations. By simply looking at the size of the 400bp Pvu II fragment, we could determine the orientation of the intron insert (see Figure 6B). This was a useful screening procedure since we had chosen to shot-gun clone the insert from a slightly larger 900bp Bam/Nru 1 fragment, which had been digested with Mnl 1, and only intron-containing recombinants generated the diagnostic 400bp PVu II band.

The scheme for cloning the intron fragment is shown in Figure 6C. A good fraction of the clones analysed had lost the diagnostic 400bp Pvu II band, indicating that we were indeed selecting for insertions within the Pvu II site in the neo gene. In all cases where this fragment was lost, it was accompanied by change in the size of fragments linked to the 400bp Pvu II fragment, further confirming that we were looking at inserts into the desired site (data not shown). A number of clones still retained the 400bp Pvu II band, and these were also tested for other abherrancies by digesting with Pst1, and comparing the patterns obtained with that of parent vector construct. Clones p^{a8}neo#3-21, &#26 had Pst1 and Pvu II digest patterns that were similar to those of parental p^{a8}neo#3. Clones

$p^{a8\underline{neo}\#4\text{-}4,4\text{-}16,\&4\text{-}17}$ were likewise indistinguishable to the parental clone $p^{a8\underline{neo}\#4}$ (data not shown). These clones were re-checked for sensitivity to added kanamycin before further analysis. We were surprised to observe that none of these, or clones we subsequently isolated had the expected larger 460bp Pvu II fragment (data not shown). Analysis of the intron sequence indicated that there were 3 stop codons when the intron was in the orientation that corresponded to the larger (460bp) fragment, of which one would be in frame with the neo reading frame. Conversely, there were six stop codons on the opposite strand, corresponding to the observed (410bp) fragment, of which two would be in the same frame as the neo gene. It is not clear if these observations are related.

We have confirmed by nucleotide sequence analysis, that at least in clones $p^{a8\underline{neo}\#4\text{-}4,4\text{-}16,\&4\text{-}17}$ that there are no insertions in the upstream Pvu II site (i.e.5' to the heat shock promoter) (Fig.5B). This was important to know since reversion of the kan^S or $G418^S$ phenotype, in this case would not be related to the transposition assay developed. Furthermore, preliminary nucleotide analysis, indicate that $p^{a8\underline{neo}}$ constructs #4-16,and 4-17 have inserts whose size is consistent with the CPII intron. Complete nucleotide analysis of this region is in progress.

DISCUSSION :

The partial construct we have described is the initial step in building what we hope will be a useful tool in studying, not only transpositional aspects of the copia retrotransposon, but also as a probe in studying other interesting genetic systems in Drosophila. Although the ideal construct is not yet complete (see below), we already know through both genetic (G418 selection) and biochemical (RNA analyses) that the neo gene is functional within the copia element in an orientation independent manner (Figure 4). From examination of the levels of the p^{a8neo} constructs, the introduced copia/neo vectors are transcribed at fairly low levels in cells. It is, therefore, desirable to be able to increase the amount of transcription of the vector sequences in cells. The way we plan to improve on this construct is to replace the copia promoter with the Drosophila heat-shock promoter, in a way that is analogous with Ty constructs (1). Using the regulatable heat shock promoter would allow us to monitor transposition events.

Analyses of transcriptional pattern observed in Figure 4 are interesting in terms of the speculation over the nature of the 2kb copia RNA (5). The Hpa1 site which has been used to clone the neo in copia is at position 4777 on the copia map. The 2kb copia RNA does not hybridise to restriction fragments between the EcoR1 sit at position 2300 and the Hinf site at

position 4556 (22). It has been proposed, from analysis of the copia sequence, that the 2kb RNA initiates at the amino end of the long open reading frame, and is probably spliced at position 1605 to an acceptor site at position 4760 (5). Splicing the 5kb copia RNA at these sites would result in an RNA that is roughly 2kb in size, which would encode a protein of approximately 50,000 daltons (23), the last 34 amino acids being encoded by a distict ORF lying between 4760-4860 (5). If this hypothesis is true, one would therefore expect the neo sequences to be part of this RNA, and that in both p^{a8}neo constructs, the traditional 2kb RNA should be replaced by a larger 3.3kb (2kb copia +1.3kb neo) transcript. Our results indicate that this may in fact be the case. The levels of the 6.5kb and 3.3kb transcripts correspond to the levels observed for the 5.2kb and 2.0kb RNAs (Figure 4A), indicating that these are related. Left unresolved, however, is the nature of the abundant 2.0kb RNA observed with neo probe in these transformants. It is possible that this transcript is not neo-specific, but background 2.0kb copia transcripts which remained when the filter was melted.

REFERENCES:

1. Boeke, J.D., Garkinkel, D.J., Styles,C.A., and Fink, G.R.(1985). Ty elements transpose via an RNA intermediate. *Cell 40* ,491-500.

2. Scherer, S., Mann, C., and Davis, R.W.(1982). Reversion of a promoter deletion in yeast. *Science 298* ,815-819.

3. Garfinkel, D.J., Boeke, J.D., and Fink, G.R.(1985). Ty element transposition : Reverse transcriptase and virus-like particles. *Cell 42* ,507-517.

4. Mellor, J., Malim, M.H., Gull,K., Tuite, M.F., McCready, S., Dibbayawan, T., Kingsman,S., and Kingsman, A.J.(1985). Reverse transcriptase activity and Ty RNA are associated with virus-like particles in yeast. *Nature (London) 318* ,583-586.

5. Mount, S.M., and Rubin, G.M.(1985). Complete nucleotide sequence of the Drosophila transposable element copia : Homology between copia and retroviral proteins. *Mol.Cell.Biol. 5* ,1630-1638.

6. Emori, Y., Shiba, T., Kanaya, S., Inouye, S., Yuki, S., and Saigo, K.(1985). Nucleotide sequence of copia and copia-related RNA in Drosophila virus-like particles. *Nature (London) 315* ,773-776.

7. Levis, R., Dunsmuir, P., and Rubin, G.M.(1984). Effects of

transposable element insertions on RNA encoded by the white gene of Drosophila. *Cell 38* ,471-481.

8. Pirrotta, V., and Brockl, C.(1984). Transcription of the Drosophila white locus and some of its mutants. *EMBO J. 3* ,563-568.

9. Steller, H., and Pirrotta, V.(1986). A transposable P vector that confers G-418 resistance to Drosophila larvae. *EMBO J. 6,*

10. Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristrom, J., and Davidson, N.(1982). Cuticle protein genes of Drosophila : Structure, organisation, and evolution of four clustered genes. *Cell 29* ,1027-1040.

11. Maniatis, T., Fritsch, E.F., and Sambrook, J.(1982). Molecular cloning - a laboratory manual. Cold Spring Harbor Laboratory.

12. Messing, J.(1983). New M13 vectors for cloning. In Methods in Enzymology.(R. Wu, ed.) Vol 101, pp.20-78.

13. Zoller, M.J., and Smith, M.(1984). Oligonucleotide - directed mutagenesis : A simple method using two oligonucleotide primers and a single - stranded DNA template. *DNA 3* ,479-488.

14. Wigler,M., Silverstein, S., Lee, L.S., Pellicer, A., Cheng, Y.C., and Axel, R.(1977). Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells. *Cell 11* ,223-232.

15. Elder, R.T., Loh, E.Y., and Davis, R.W. (1983). RNA from the yeast transposable element Ty1 has both ends in the direct repeats, a

structure similar to retrovirus RNA. *Proc.Natl.Acad.Sci.U.S.A.* 80, 2432-2436.

16. Shiba, T., and Saigo, K. (1983). Retrovirus-like particles containing RNA homologous to the transposable element copia in Drosophila melanogaster. *Nature (London)* 302, 119-123.17.

17. Flavell, A.J., and Brierley, C. (1986). The termini of extrachromosomal linear copia elements. *Nucleic Acids Res.* 14, 3659-3669.

18. Flavell, A.J., and Ish-Horowicz, D. (1981). Extrachromosomal circular copies of the eukaryotic transposable element copia in cultured Drosophila cells. *Nature (London)* 292, 591-595.

19. Flavell, A.J., and Ish-Horowicz, D. (1983). The origin of circular copia elements. *Cell* 34, 415-419.

20. Mossie, K.G., Young, M.W., and Varmus, H.E. (1985). Extrachromosomal DNA forms of copia-like transposable, F-elements, and middle repetitive DNA sequences in Drosophila melanogaster. *J.Mol.Biol.* 182, 31-43.

21. Panganiban, A.T., and Temin, H.M. (1984). Circles with two LTRs are precursors to integrated retrovirus DNA. *Cell* 36, 673-679.

22. Schwartz, H.E., Lockett, T.J., and Young, M.W. (1982). Analysis of transcripts from two families of nomadic DNA. *J.Mol.Biol.* 157, 49-68.

23. Flavell, A.J., Ruby, S.W., Toole, J.T., Roberts, B.E., and Rubin, G.M. (1980). Translation and developmental regulation of RNA encoded by the eukaryotic transposable element copia. *Proc.Natl.Acad.Sci.U.S.A.* 77, 7107-7111.

FIGURE LEGENDS :

Figure 1 : General model for the transposition pathway of the Drosophila and yeast retrotransposons. Elements integrated in the genome are transcribed to yield genomic length transcripts that are terminally redundant (6,15). RNA templates are packaged into predominantly cytoplasmic (Ty-3,4), or nuclear (copia-16) particles where reverse transcription probably occurs. The product of this reaction is a free double stranded linear DNA molecule that is colinear with integrated forms (17). Linear molecules are then presumed to be circularised to yield free circular DNA forms containing one or two LTRs (18-20). Circles containing two LTRs are presumed, in analogy to retroviruses (21), to be acted on by an element-encoded endonuclease and integrated in the genome.

Figure 2 : Theoretical basis for the construction of a copia-based 'hopping' vector. (a) The structure of the copia element in clone p^{a8}

showing the position of the open reading frames (5). The unique Hpa1 site in copia is positioned between these open reading frames. In the experimental construct, $p^{a8\#4}$ (b), the orientation of the heat-shock promoter-driven neo gene (hsp-neo) gene (9) is opposite to that of the 5.2kb copia RNA. The 60bp cuticle protein II (CPII) intron is cloned into the neo gene such that it can only be spliced from the copia (but not neo) transcript. The constructs are used to transform Drosophila Schneider cells and transpotion events are monitored by selecting for growth in the presence of G418 (d)-(h). Note that this selection scheme demends that spliced copia transcripts be reverse transcribed, and cDNAs reintegrate into the genome to produce stable G418 cells.

Figure 3 : The neo gene is prepared from plasmid pUC-hsneo (9), by digestion with EcoR1 and Sma1. The ends are repaired by treatment with the Klenow fragment of E.coli DNA polymerase, and the 1.3kb neo fragment isolated by electroelution. This is then ligated to the copia vector, p^{a8} which had been opened at the Hpa1 site, to generate plasmid constructs, $p^{a8\#3}$ and $p^{a8\#4}$ respectively.

Figure 4 : Northern blot analyses of Drosophila cells transfected with p^{a8neo} vectors. Schneider cells were transfected with the p^{a8neo} constructs as described in Materials and Methods. Pools of G418 resistant cells were propagated, and poly $(A)^+$ RNAs prepared as described

previously (20). Approximately 10 micrograms of poly $(A)^+$ RNA were loaded per lane, and samples were fractionated on formaldehyde-containing agarose gels and analysed as described previously (20). The probes are A) nick-translated copia (cDm 5002), and B) strand-specific RNA probes derived from the construct Eco/Sma hspneo in pGEM vectors. The lanes are 3, $p^{a8\#3}$ poly$(A)^+$ RNA; C, Control Schneider poly$(A)^+$ RNA; and 4, $p^{a8\#4}$ poly$(A)^+$ RNA.

Figure 5 : Scheme for generating CPII intron fragments. A). Restriction map of pCPII-7 (10) containing the intron fragment. Exon sequences are represented by closed areas, and the intron sequences by the darkly-shaded areas. B). Nucleotide sequence showing intron/exon boundaries in pCPII. Exon sequences are presented in broad letters. C). Nucleotide sequence of the synthesised mutagenic oligonucleotide designed to generate an Mnl1 site at the right intron/exon boundary. The position of Mnl1 sites is as shown above.

Figure 6 : Using the Pvu II recognition sequence to determine the orientation of intron sequences. A). Intron sequences at the intron/exon boundary. We have reconstructed the nature of recombinants when these sequences are cloned at the Pvu II site. B). Effect of intron orientation on the size of the marker 400bp Pvu II fragment. Note that non-intron fragments cloned at this site will not generate this marker Pvu II fragment.

Neo sequences are represented as shaded areas, and the heat-shock promoter sequences as the open areas. The intervening intron fragment is depicted as a thin line.

Figure 7 : Scheme for cloning intron fragment into p^{a8neo} vector constructs. The source of intron fragment to be cloned, pBR322(Cp II Bam/Sal-Mnl1 #1), is digested with Bam H1 and Nru1 to generate the 900bp fragment. This fragment is electroeluted, and digested with Mnl1, before ligation to Pvu II linearised p^{a8neo} vectors. Vector DNA is prepared by isolating full-length Pvu II linears by partial digestion with Pvu II. Linear DNA is then purified by electroelution, and dephosphorylated by treament with the Bacterial Alkaline Phosphatase. Transformants are screened by replica-plating on L-plates containing kanamycin. Kan^S colonies are screened for the presence of the marker Pvu II fragment.

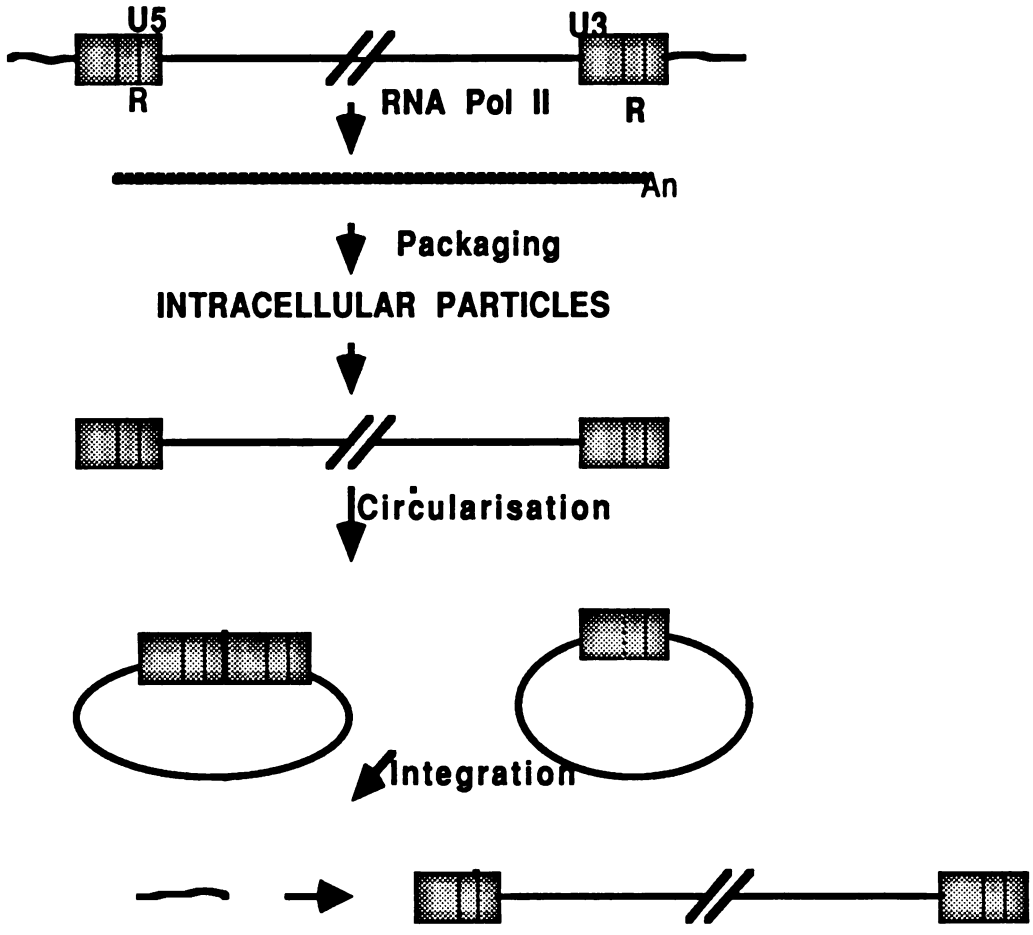
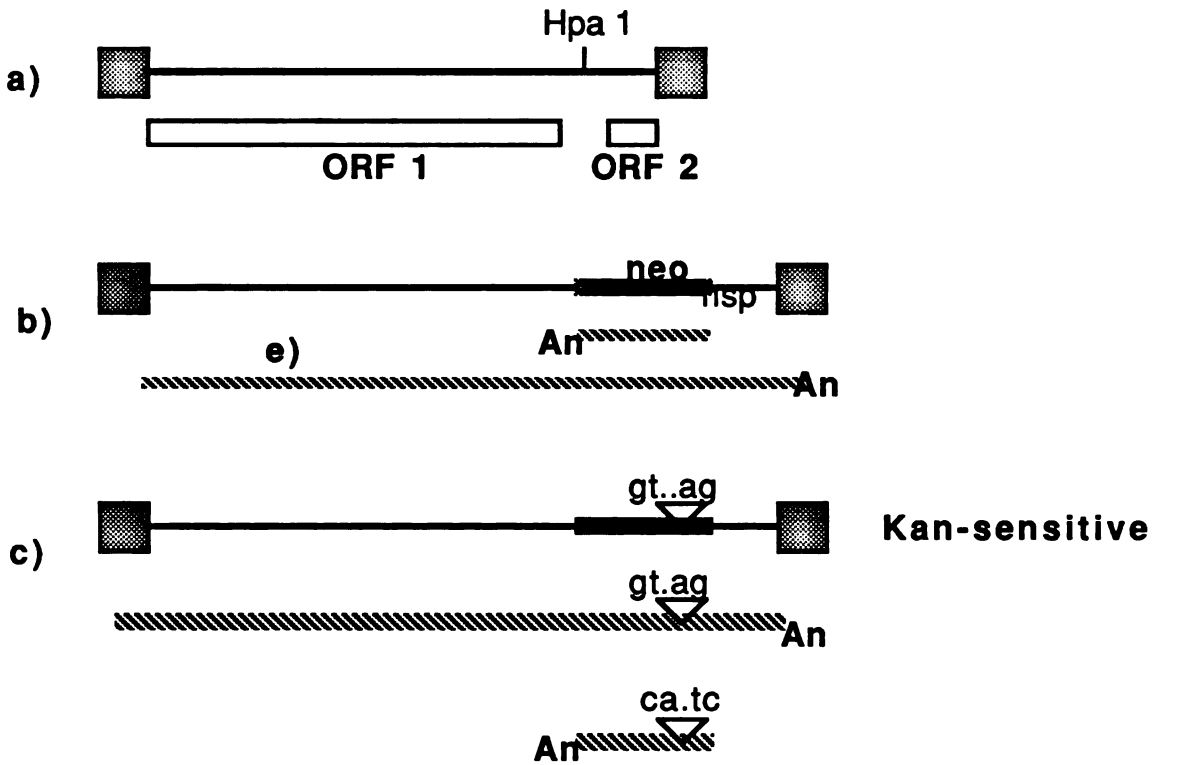
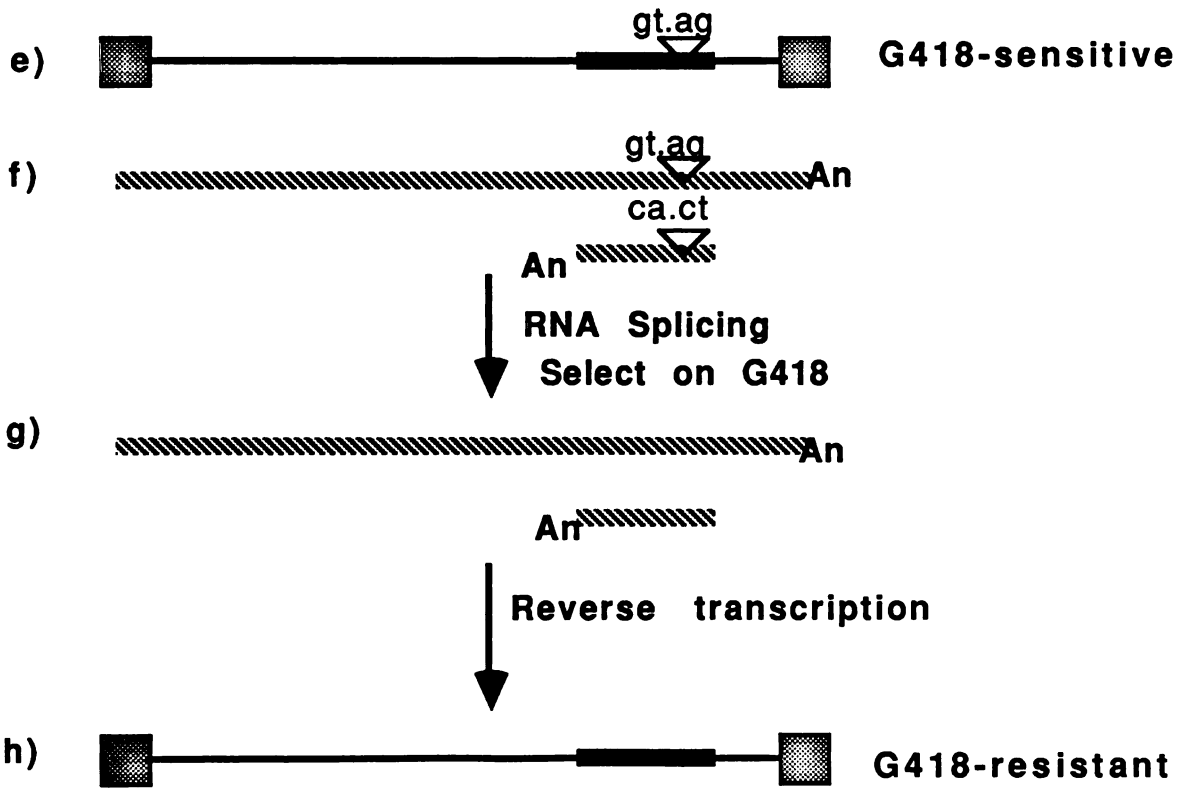
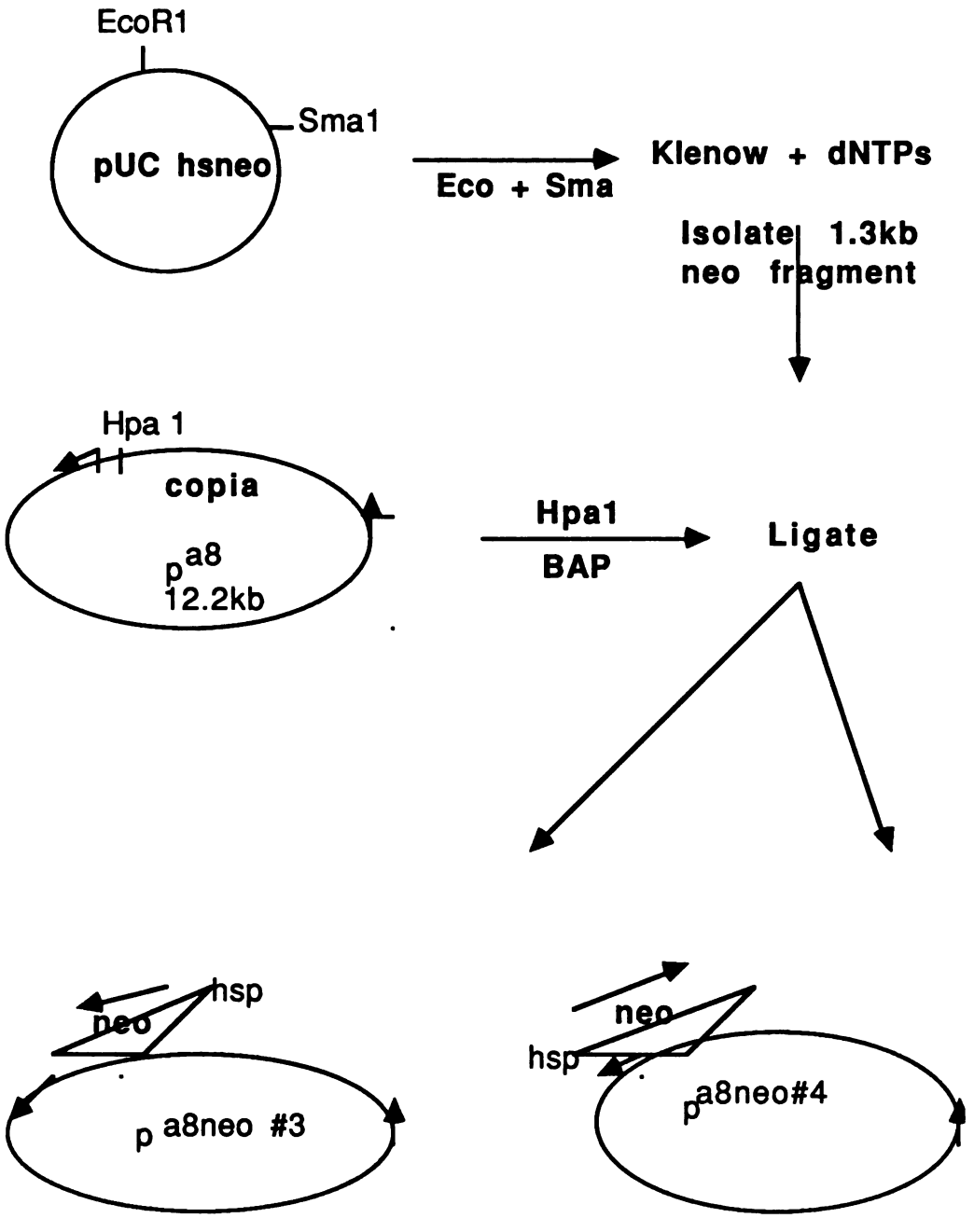


Figure 1: General pathway for the transposition of retrotransposons in Drosophila and yeast.



d) Transform cells. Co-selection to monitor DNA uptake




EcoR1
Sma1
pUC hsneo
Eco + Sma
Klenow + dNTPs
Isolate 1.3kb
neo fragment
Hpa 1
copia
p a8
12.2kb
Hpa1
BAP
Ligate
hsp
neo
p a8neo #3
neo
hsp
p a8neo#4

A Pvu II recognition site 5' CAGCTG 3'

CP II intron sequence 5' ctggat.....actcac 3'
3' gaccta.....tgagtg 5'

Observing the GT...AG rule, the correct orientation of intron is
5' gtgagt.....atccag 3'
3' cactca.....taggtc 5'

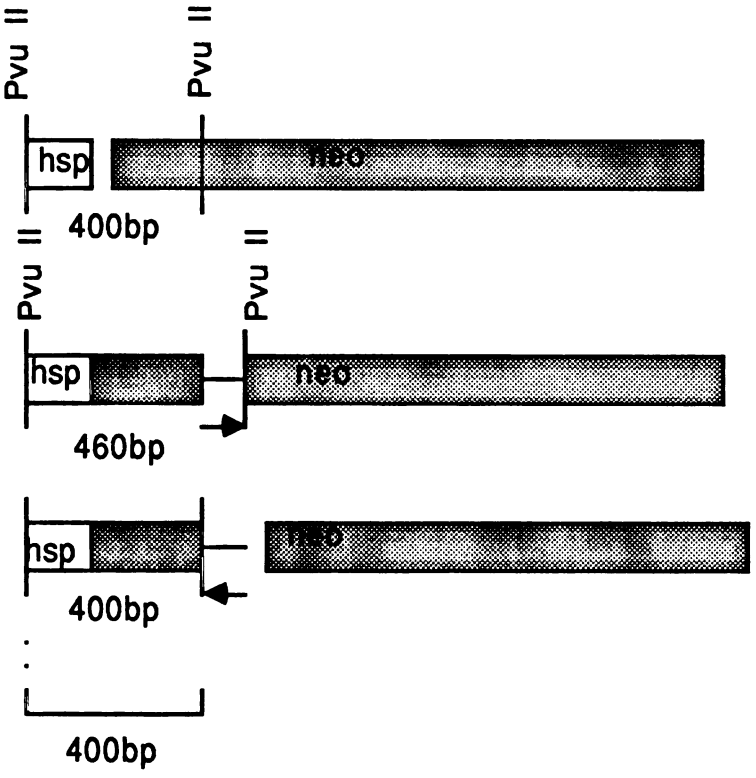
Insertion of intron sequences into the Pvu II site therefore generates

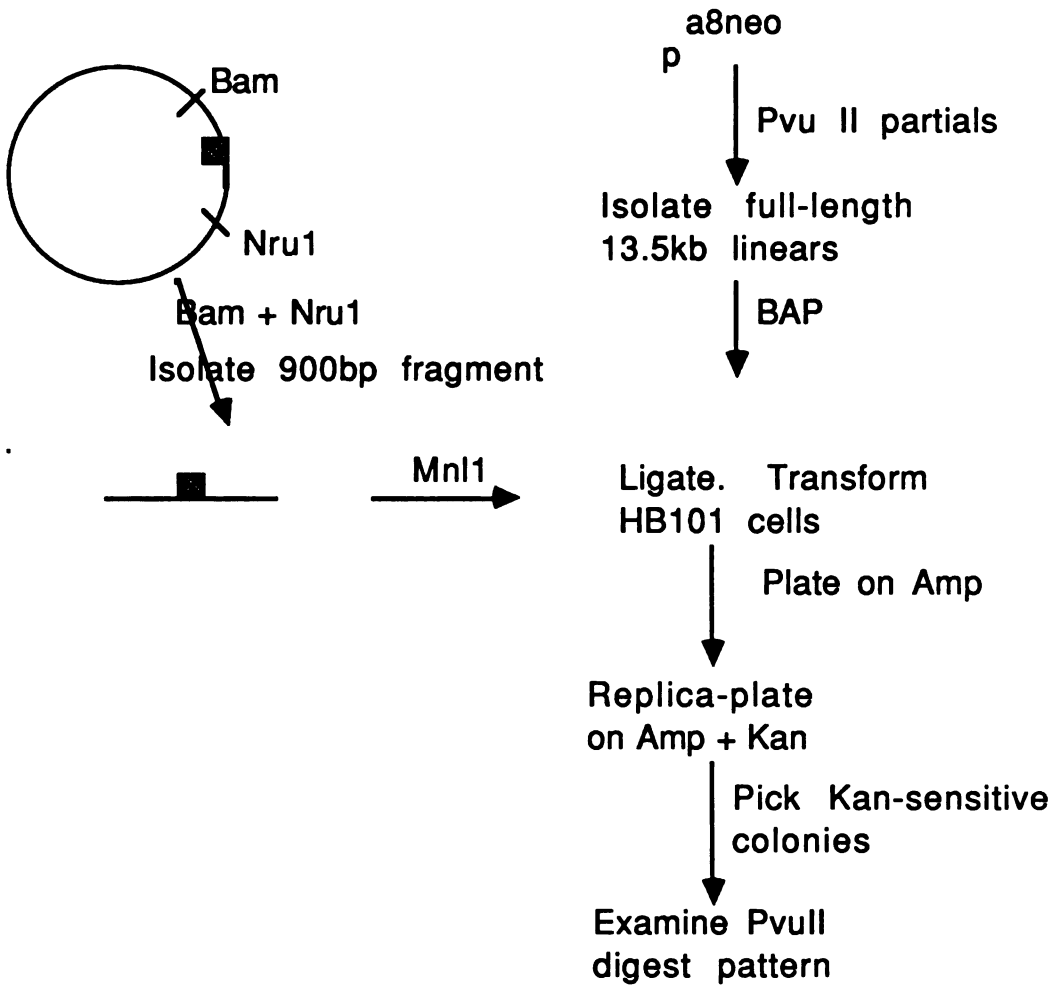
5' CAGgtgagt.....atccagCTG 3'
Pvu II

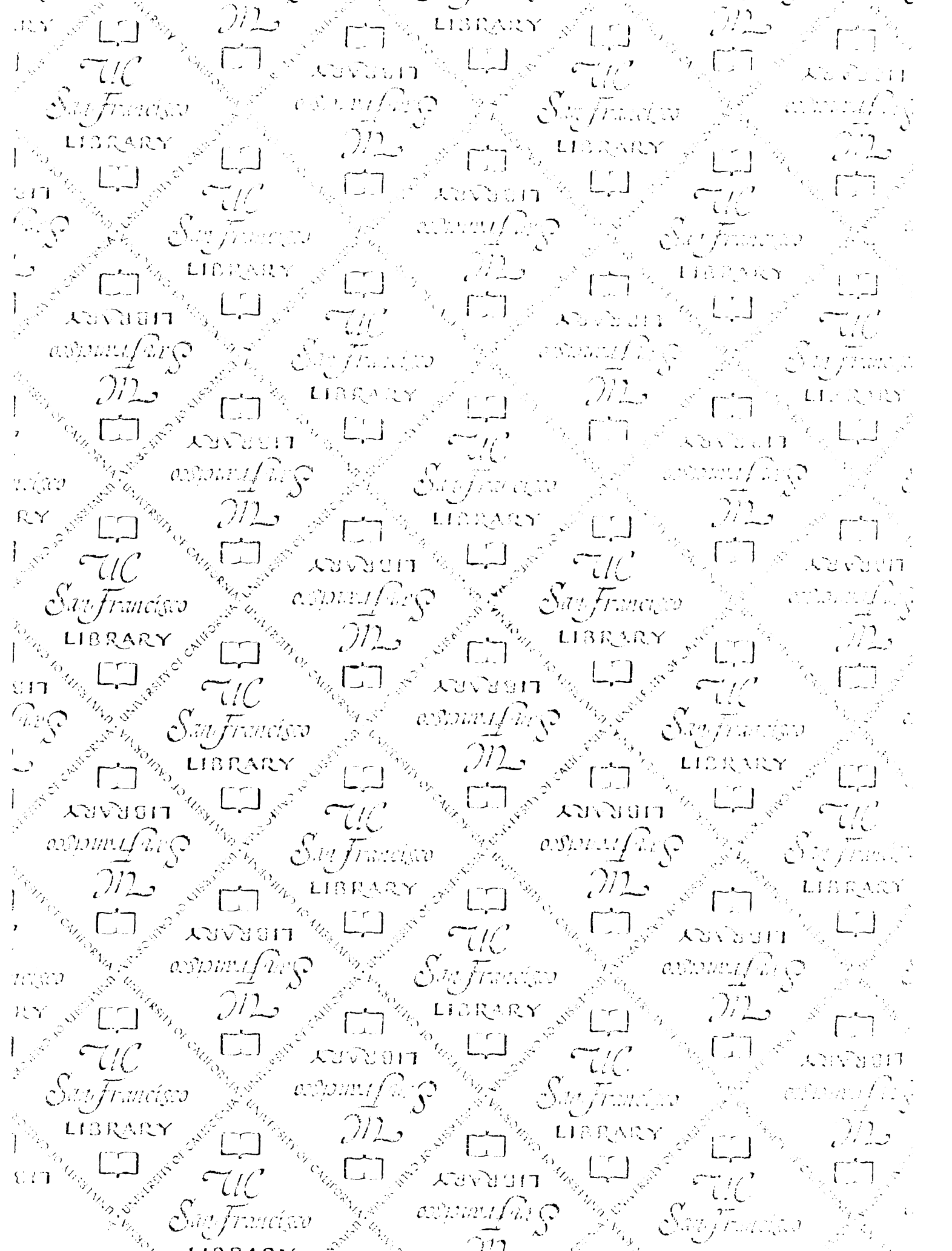
or

5' CAGctggat.....actcacCTG 3'
Pvu II

B.




a8neo
p
Bam
Nru1
Pvu II partials
Isolate full-length
13.5kb linears
BAP
Bam + Nru1
Isolate 900bp fragment
Mnl1
Ligate. Transform
HB101 cells
Plate on Amp
Replica-plate
on Amp + Kan
Pick Kan-sensitive
colonies
Examine PvuII
digest pattern


UC
San Francisco
LIBRARY