

Translation and developmental regulation of RNA encoded by the eukaryotic transposable element *copia*

(*Drosophila melanogaster*/hybrid-arrested translation/diazobenzyloxymethyl-cellulose)

ANDREW J. FLAVELL*†, STEPHANIE W. RUBY‡, JOHN J. TOOLE*§, BRYAN E. ROBERTS‡, AND GERALD M. RUBIN*¶||

*Sidney Farber Cancer Institute, Boston, Massachusetts 02115; and †Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115

Communicated by Mark Ptashne, August 26, 1980

ABSTRACT *copia*-specific RNA was isolated from *Drosophila melanogaster* tissue culture cells by hybridization of cytoplasmic polyadenylated RNA to *copia* DNA immobilized on cellulose. The purified RNA was translated in reticulocyte lysates. One major polypeptide of approximately 51,000 daltons was synthesized in addition to several others between 18,000 and 38,000 daltons. The 51,000-dalton polypeptide and several of the others are encoded by mRNAs of about 2000 nucleotides. The approximate locations on the *copia* element of the coding sequences for the 51,000-dalton polypeptide and several other proteins were determined by hybrid-arrested translation with *copia* restriction fragments. The relative abundance of *copia*-specific RNA was determined at various stages of the *Drosophila* life cycle. The level of *copia*-specific RNA is modulated during development of the organism, with the highest level occurring during the larval stages.

Transposable elements have been described for several eukaryotic organisms such as maize, *Drosophila*, and yeast (1-4), and the evolutionary diversity of these species suggests that transposition of DNA segments occurs in all eukaryotes. Transposable elements promote a wide variety of genetic effects such as unstable mutations, deletion formation, and other chromosomal rearrangements (1-4).

copia is a DNA sequence present in about 50 closely conserved copies in the *Drosophila melanogaster* embryo genome (5, 6) and is transposable as evidenced by striking differences in both the number and chromosomal locations of *copia* elements between strains and also between tissue culture cells and embryos (5, 7). *copia* consists of a DNA sequence of approximately 4.5 kilobases (kb) flanked at each end by identical 276-nucleotide-long sequences (8) that are arranged in the same orientation with respect to the element (6, 8). *copia* is transcribed in tissue culture cells into two major RNA species approximately 5 kb and 2 kb long (refs. 9 and 10; this study) which together constitute about 3% of the cell polyadenylated RNA (6). The structural features of *copia* and its insertion sites (7, 11) are closely paralleled by the yeast transposable element, Ty1 (12, 13), by the integrated proviruses of certain retroviruses (14, 15), and by several prokaryotic transposons (1, 16).

The mechanism of transposition is best understood for prokaryotic transposons (for review, see ref. 1). Proteins encoded by the transposable element Tn3 mediate or repress transposition (17) and a "transposase" encoded by another element, Tn5, also has been reported (18). In contrast, no proteins encoded by eukaryotic transposable elements have yet been described despite several reports of RNAs encoded by these elements (1, 6). We report here that the *Drosophila* transposable

element *copia* encodes mRNA species that are translated *in vitro* into several polypeptides. We also present data identifying the RNA species responsible for the synthesis of these polypeptides and the approximate locations within the element of the coding regions for some of the polypeptides. Furthermore, we show that the expression of *copia* RNAs and, therefore, presumably of the proteins encoded by them is modulated during the development of the organism.

MATERIALS AND METHODS

Cells, DNA, and RNA. Eschaler's Kc₀ line of *Drosophila* cells was propagated by the method of Rubin and Hogness (19). Cytoplasmic polyadenylated RNA was isolated from these cells by the method of Flavell *et al.* (20). All experiments involving the growth of recombinant plasmids were performed in a P2 category laboratory under containment conditions specified by the National Institutes of Health guidelines. Methods for the isolation of plasmid DNAs were described by Wensink *et al.* (21). Purification of DNA restriction fragments was as described by Levis *et al.* (8). Total cell RNA was isolated from *Drosophila melanogaster* (Oregon R) embryos, larvae, and adults by the method of Fyrberg *et al.* (22).

Purification of *copia*-Specific RNA. Cytoplasmic polyadenylated RNA (200-400 µg) from Kc₀ cells was hybridized with either the *copia*-containing plasmids cDm2056 and cDm5002 or with restriction fragments thereof covalently bound to cellulose (23) as described by Paucha and Smith (24) with the following modifications. High-stringency purification of *copia*-specific RNA was accomplished by two sequential hybridizations to the same sample of DNA cellulose followed by exhaustive washing with 50 mM Tris-HCl, pH 7.5/50% (vol/vol) formamide at 37°C. For optimal RNA integrity but lower purity, only one hybridization was used. During all purifications, a significant amount of DNA was released from the cellulose. This was removed from the mRNA preparations by oligo(dT)-cellulose chromatography. Sucrose gradient ultracentrifugation in 50% formamide was as described by Flavell *et al.* (20).

In Vitro Translation and Polypeptide Analysis. RNA was translated in a micrococcal nuclease-treated, rabbit reticulocyte

Abbreviations: kb, kilobases; P-51, P-38, P-31, P-28, P-19.5, P-19, and P-18, *copia*-specific polypeptides approximately 51,000-18,000 daltons, respectively.

† Present address: Imperial Cancer Research Fund, Mill Hill Laboratories, London, England.

§ Present address: Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139.

¶ Present address: Carnegie Institution of Washington, 115 West University Parkway, Baltimore, MD 21210.

|| To whom reprint requests should be addressed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

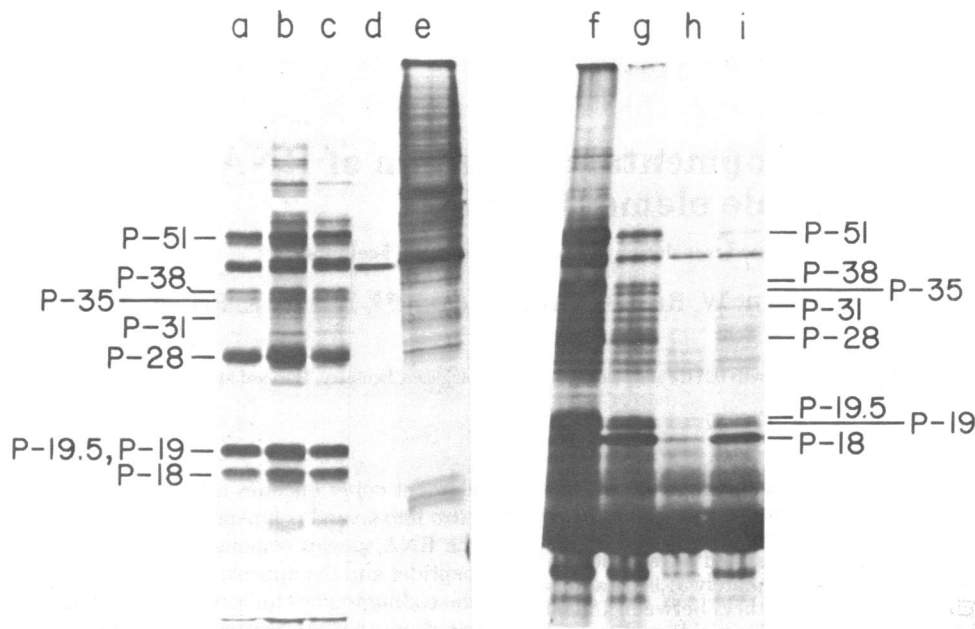


FIG. 1. *In vitro* translation of *copia*-specific RNAs. Lane a: *copia*-specific RNA was purified from 5 μ g of Kc₀ cell polyadenylated RNA by hybridization to the 5.2-kb *Hpa* II restriction fragment of cDm2056 (containing the entire *copia* element plus 0.2-kb flanking DNA) covalently bound to cellulose. *Copia*-specific mRNA was eluted from the cellulose and translated in a reticulocyte lysate system with [³⁵S]methionine, and the products were sized on a NaDodSO₄/polyacrylamide gel. The synthesized polypeptides were visualized by fluorography. Lanes a–e: a, *copia*-specific polypeptides of molecular sizes between 51,000 and 18,000 daltons (P-51 to P-18, respectively); b and c, a similar experiment with DNA from the *copia*-containing plasmids cDm5002 and cDm2056, respectively; d, no addition of mRNA; e, with 0.16 μ g of total Kc₀ cell polyadenylated cytoplasmic RNA. Lanes f–i, hybrid-arrested translation experiment, in which cDm2056-specific mRNA (approximately 0.2 μ g per lane) was hybridized with excess cDm2056 DNA before translation: f, no hybridization; g, mock hybridization (no addition of DNA); h, with 6.8 μ g of cDm2056; i, with 6.8 μ g of cDm2056 DNA and hybrids that were heat-denatured prior to translation.

system using [³⁵S]methionine (New England Nuclear) label (25). Polypeptides were fractionated in NaDodSO₄/17.5% (wt/vol) polyacrylamide gels (26) and visualized by fluorography (27). Apparent molecular weights were determined by using adenovirus 2-specific proteins synthesized late in the infectious cycle as standards (28). Hybrid-arrested translation (29, 30) was performed with the following modifications. Before hybridization, the DNA, *copia*-selected RNA, and 25 μ g of calf liver tRNA (Boehringer) were denatured at 100°C for 2 min in 45 μ l of 89% (vol/vol) formamide/10 mM sodium 1,4-piperazinebis(ethanesulfonate), pH 6.4. After being rapidly chilled on ice, the sample was adjusted to 50 μ l of 80% (vol/vol) formamide/0.4 M NaCl/50 mM sodium 1,4-piperazinebis(ethanesulfonate), pH 6.4/1 mM EDTA and incubated at 37°C for 3 hr. After addition of 350 μ l of ice-cold H₂O, half of the sample was heated at 100°C for 2 min. Both sample halves were then treated with ethanol, and the precipitate was washed twice with 70% (vol/vol) ethanol, dried, and translated.

RESULTS

***copia*-Specific RNA Encodes Several Polypeptides.** *copia*-specific mRNA was purified by hybridization of polyadenylated cytoplasmic RNA extracted from tissue culture cells to a restriction fragment of the plasmid cDm2056 containing the *copia* element and 0.2 kb of flanking *Drosophila* DNA. For convenience the fragment was bound to cellulose prior to hybridization. The purified RNA preparation was used to direct the synthesis of proteins in an mRNA-dependent rabbit reticulocyte cell-free system. A number of polypeptides were synthesized, the largest and usually most abundant being approximately 51,000 daltons (P-51; Fig. 1). None of the synthesized proteins seemed to be major constituents of the translation products of total tissue culture cell polyadenylated RNA. A similar set of polypeptides was synthesized by using RNA se-

lected by hybridization to cDm2056 or to another *copia*-containing plasmid, cDm5002 (11), and for convenience the entire cDm2056 plasmid was used in all subsequent experiments.

The following experiment confirmed that these polypeptides were *copia*-specific and not encoded by contaminating mRNA. The RNA preparation was hybridized with excess *copia* DNA in solution, and the hybridized products were translated (29, 30). Under these conditions, if the DNA fragment used were homologous to the protein-coding regions of a mRNA, then the resulting hybrid would not be translated; this technique is termed hybrid-arrested translation. The synthesis of P-51, together with several other polypeptides, was arrested by hybridizing with cDm2056 DNA, and the inhibition was reversed by melting the hybrids (Fig. 1). Also, when hybrid-arrested translation was performed with purified *copia* DNA restriction fragments, the same inhibitory effect was observed (see below). Therefore, P-51 and several other polypeptides are encoded by the *copia* element itself and by neither the non-*copia* DNA also present in cDm2056 nor contaminating mRNA.

Location of the Protein Coding Region of P-51 on the *copia* Element. To map the locations of the protein-coding regions on the *copia* DNA sequence, *copia*-specific RNA was hybridized with 10-fold molar excesses of purified restriction fragments of *copia* DNA prior to translation of the mRNA preparation. We used fragments that together span about 90% of the *copia* DNA element. The synthesis of P-51 was inhibited only by fragments A, B, and C, which span half of the element (Fig. 2). These fragments encode the *copia* 2-kb RNA and the 5' half of the 5-kb RNA (unpublished results; 10). It is significant that restriction fragment E did not arrest the translation of P-51. Fragment E contains the 276-base-pair direct repeat that is present in fragment A (which did arrest translation). Therefore, unless the high A-T content (8) or some unforeseen feature of the direct repeat sequence prevented its stable hybridization to

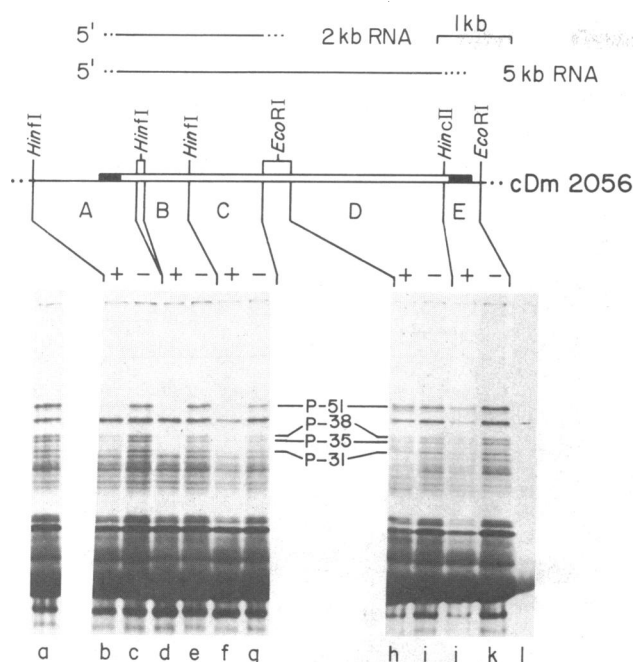


FIG. 2. Mapping *copia*-encoded proteins on the DNA genome by hybrid-arrested translation. Samples of *copia*-specific mRNA (approximately 0.2 μ g) were annealed with restriction fragments A-E of cDm2056 DNA. Each sample was split into two halves, one of which was translated (+ lanes) and the other was heat-denatured prior to translation (- lanes). Radiolabeled products were identified by electrophoresis on a NaDodSO₄/polyacrylamide gel and by fluorography for 48 hr. Lanes: a, mock hybrid-arrested translation with no addition of DNA; b and c, with 0.6 μ g of fragment A; d and e, with 0.6 μ g of fragment B; f and g, with 0.4 μ g of fragment C; h and i, with 0.9 μ g of fragment D; j and k, with 0.2 μ g of fragment E; l, no addition of mRNA. *copia*-specific polypeptides are indicated as in Fig. 1. The approximate locations of the major 5-kb and 2-kb *copia*-specific RNAs on the DNA genome are shown.

copia-specific mRNA, the DNA sequences in fragment A that were responsible for hybrid arrest must lie in the approximately 250 nucleotides between the end of the direct repeat and the clustered *HinfI* sites separating fragments A and B. This suggests that no part of the protein coding sequences for P-51 lies in the *copia* direct repeat.

The synthesis of P-38, P-35, and P-31 was specifically arrested by DNA fragments B and C but not by fragment A. This suggests that the sequences encoding the amino termini of these polypeptides lie at least 250 base pairs downstream from that of P-51. The results for the remaining polypeptides were less clear and will require further study. However, fragments D and E did not arrest the synthesis of any *copia*-specific proteins. This is consistent with the observation that the 5-kb RNA was not efficiently translated in our system (see below) because fragment D sequences were represented only in 5-kb RNA (ref. 10; unpublished results).

P-51 Is Encoded by mRNA Approximately 2 kb long. To determine which *copia*-specific RNA encodes P-51, partially purified *copia* mRNA was size-fractionated by sedimentation on a sucrose gradient containing 50% (vol/vol) formamide. A portion of each gradient fraction was assayed for the presence of the major 5-kb and 2-kb *copia*-specific mRNAs by electrophoresis on an agarose gel containing methyl mercury hydroxide denaturant followed by transfer of the RNA to activated paper (31) and hybridization to ³²P-labeled *copia* DNA. The 5-kb and 2-kb *copia*-specific RNA species were well resolved from each other, although after prolonged exposure of the autoradiogram, a minor degree of cross-contamination

(estimated at less than 5%) of each RNA species with the other was observed (data not shown). A second portion from each fraction was translated and the [³⁵S]methionine-labeled products were separated on NaDodSO₄/polyacrylamide gels as before (Fig. 3). Despite contamination of *copia*-specific polypeptides with proteins encoded by total cell RNA, several *copia*-specific polypeptides, including P-51, were easily visible. Several conclusions can be drawn from this experiment. First, polypeptides P-51, P-38/P-35 (which comigrated on this particular gel), P-28, P-19.5, and P-19 are encoded by RNA of approximately 2 kb in length. Second, P-18 is encoded by a somewhat smaller RNA. Last, only low levels of ³⁵S-labeled polypeptides (including P-51) were detected in the translation products of fractions containing predominantly 5-kb RNA.

Abundance of *copia*-Specific RNAs Varies During Development. We have examined the expression of *copia* at several stages of the *Drosophila* life cycle. Total cellular RNA was isolated from embryos, larvae, and adults (22). Equal quantities of these RNAs and RNA from Kc₀ tissue culture cells were electrophoresed through agarose gels containing 6 mM methyl mercury hydroxide, then transferred to diazobenzylmethyl-oxy-cellulose filter paper, and hybridized to ³²P-labeled *copia* DNA. Prominent hybridization was observed only in those lanes containing tissue culture and larval RNA. To demonstrate the existence of hybridizable RNA in the samples containing no detectable *copia*-specific sequences, the *copia* probe was melted off the filter, which was then rehybridized to a restriction fragment containing both a ribosomal protein gene and a second adjacent undefined gene encoding mRNAs of 500 and 3000 nucleotides, respectively (P. O'Connell, C. Vaslet, and M. Rosbash, personal communication). Hybridization was observed to both species of mRNA in all samples. Although some variability in the hybridization intensity is discernible, it was much less than that observed for *copia*. Thus, we conclude that the expression of *copia*-specific RNAs is modulated at least severalfold during the life cycle of *Drosophila melanogaster*, with the highest level occurring during the larval stages.

DISCUSSION

The major conclusions from this work are twofold: first, the transposable element *copia* encodes the information for several polypeptides; second, *copia*-specific RNAs are modulated during development. Therefore it is likely that protein gene products of this transposable element are present in varying amounts during the life cycle of the organism. However, such proteins have not been identified to date and their function remains obscure.

The information required to encode all the *copia*-specific proteins described here demands a DNA complexity of about 6 kb, somewhat greater than that of the entire *copia* element. Furthermore, less than 50% of the element appears to be translated into polypeptides. It is possible that each of these products derives from a different individual copy of the element in the cell. There are about 150 copies of the *copia* element in tissue culture cells (6), and minor variations in the DNA sequence of the direct repeats (8) and in the restriction maps of cloned individual elements (6) are known. We think it likely that some of these proteins may derive from such variants. It is also possible that some of the proteins are related by premature translation termination, by initiation at different AUG codons on a single mRNA, or by proteolysis of the P-51 protein. The last explanation is unlikely, as protein degradation occurs rarely in reticulocyte lysates (unpublished results). We cannot discount premature translation termination as the origin of some of the smaller *copia*-specific polypeptides. However, P-38, P-35, and P-31 cannot be derived from P-51 in this manner because their synthesis is insensitive to hybrid arrest by DNA fragment

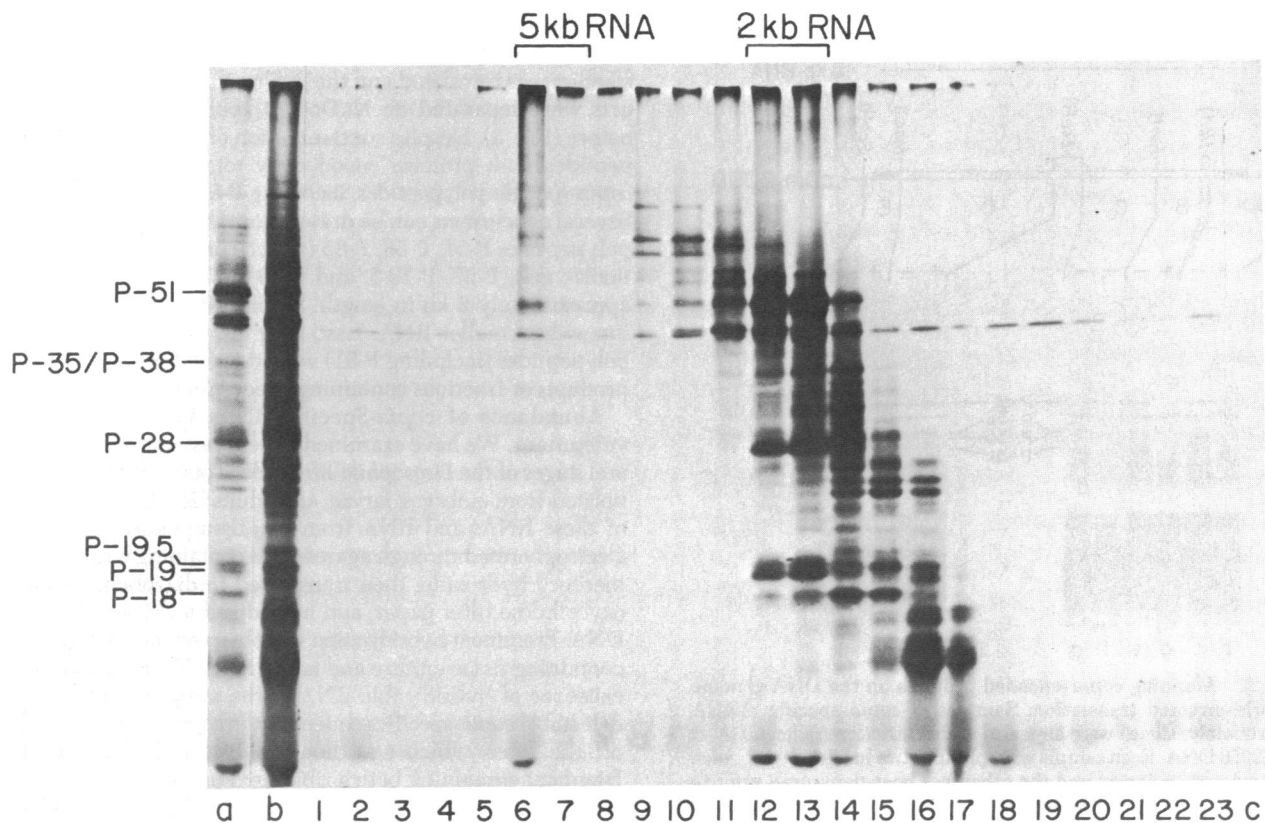


FIG. 3. *In vitro* translation of size-fractionated *copia*-specific mRNAs. mRNA enriched in *copia*-specific sequences was isolated from 400 μ g of *Kc0* tissue culture cell total mRNA. This RNA was fractionated by sedimentation on a 5–20% (wt/vol) sucrose gradient containing 50% formamide. Ten percent of each fraction was translated by using [35 S]methionine, and the products were sized on a polyacrylamide gel. Fluorography was for 15 hr. Lanes: a, 1% of the input *copia*-enriched RNA; b, 0.16 μ g of total *Kc0* mRNA; 1–23, gradient fractions 1–23 (fraction 1 is at the bottom of the gradient); c, no addition of mRNA. *copia*-specific polypeptides are indicated as in Fig. 1, and the approximate locations on the gradient of the 5-kb and 2-kb *copia*-specific mRNAs are shown.

A, which contains information necessary for translation of P-51 (Fig. 2). Therefore, it appears that P-38, P-35, and P-31 may derive from translational initiation at an alternative site(s).

Last, it is possible that some of the polypeptides are encoded by discrete mRNA species. These species could derive from different individual RNAs transcribed from a common DNA template. Such heterogeneity could be generated by the presence of several transcriptional promoters on the DNA sequence, giving variable 5' ends to the *copia*-specific RNA species. Alternatively these discrete species could derive from a common precursor RNA by variable splicing patterns, giving access to either different reading frames or new translational initiation sites. We have observed size heterogeneity in the 2-kb RNA upon gel electrophoresis and both the 5-kb and 2-kb RNAs possess heterogeneous 5' ends (unpublished data). A more accurate study of the individual *copia* DNA elements and the topography of their transcripts will be required to determine what contribution each of the above effects makes to the generation of the observed family of *copia*-specific proteins. Whereas *copia*-specific transcripts constitute about 3% of the tissue culture cell mRNA, the translational efficiency of this fraction is low, and it may be that a large proportion of the *copia* transcripts are not functional messages.

The results presented here suggest that *copia* RNA is most abundantly expressed in the larval stages of development. *copia* RNA has been detected at low levels in embryos harvested at 22 hr, shortly before hatching (9). However, we did not detect any *copia*-specific transcripts isolated from 8- to 9-hr-old embryos (Fig. 4). This discrepancy might be due either to our assay being less sensitive or to an acceleration of *copia* transcription in the later stages of embryogenesis. It will be of interest to

determine in more detail the developmental time course and possible tissue specificity of *copia* RNA levels.

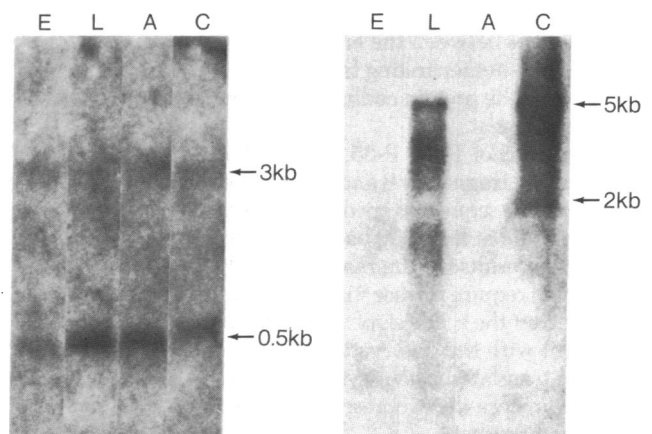


FIG. 4. Hybridization of *copia* to total-cell RNA from different developmental stages. Equal amounts of total cell RNA from 8- to 9-hr embryos (lanes E), a mixture of first-, second-, and third-instar larvae (lanes L), adults (lanes A), and *Kc0* tissue culture cells (lanes C) were electrophoresed on a 1.5% (wt/vol) agarose gel containing methyl mercury hydroxide and were transferred to diazobenzoyloxymethyl cellulose paper. (Right) *copia*: 32 P-labeled *copia* DNA (10^8 cpm/ μ g) was hybridized to the bound RNA. (Left) Control: 32 P-labeled *copia* DNA was melted off the paper, which was then rehybridized to a 32 P-labeled DNA fragment containing a ribosomal protein gene (0.5 kb) and an adjacent uncharacterized gene (3 kb). The position of the RNA species is indicated. The 2-kb *copia*-specific RNA comigrates with *Drosophila* rRNAs, which inhibit RNA transfer to the paper because of their abundance.

Some prokaryotic transposons encode proteins that modulate their own transposition, and *cop*ia-specific proteins might have similar properties. If this were the case, one might expect a higher rate of *cop*ia transposition in tissues or developmental stages in which *cop*ia RNAs were abundant. It is also possible that *cop*ia can control the expression of adjacent genes in a developmentally regulated manner by read-through from promoters located within the element.

We thank H. Schwartz, M. W. Young, and M. Rosbash for communicating their results prior to publication, M. Rosbash for the restriction fragment mentioned in the text, and Brian Backner for his help in preparation of this manuscript. This research was supported by grants from the American Cancer Society and National Institutes of Health to G.M.R. and from the National Science Foundation to B.E.R.; J.J.T. was an American Cancer Society Postdoctoral Fellow, and S.W.R. was a National Institutes of Health Postdoctoral Fellow.

1. Calos, M. P. & Miller, J. H. (1980) *Cell* **20**, 579–595.
2. Fincham, J. R. S. & Sastry, G. R. K. (1974) *Annu. Rev. Genet.* **8**, 15–50.
3. Green, M. M. (1977) in *DNA Insertion Elements, Plasmids, and Episomes*, eds. Bukhari, A. I., Shapiro, J. A. & Adhya, S. A. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 467–482.
4. McClintock, B. (1956) *Cold Spring Harbor Symp. Quant. Biol.* **21**, 197–216.
5. Potter, S. S., Brorein, W. J., Dunsmuir, P. & Rubin, G. M. (1979) *Cell* **17**, 415–427.
6. Finnegan, D. J., Rubin, G. M., Young, M. W. & Hogness, D. S. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1053–1063.
7. Strobel, E., Dunsmuir, P. & Rubin, G. M. (1979) *Cell* **17**, 429–439.
8. Levis, R., Dunsmuir, P. & Rubin, G. M. (1980) *Cell*, **21**, 581–588.
9. Carlson, M. & Brutlag, D. (1978) *Cell* **15**, 733–742.
10. Young, M. W. & Schwartz, H. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45**, in press.
11. Dunsmuir, P., Brorein, W. J., Simon, M. A. & Rubin, G. M. (1980) *Cell*, **21**, 575–579.
12. Farabough, P. J. & Fink, G. (1980) *Nature (London)* **286**, 352–356.
13. Gafner, J. & Philippsen, P. (1980) *Nature (London)* **286**, 414–418.
14. Dhar, R., McClements, W. L., Enquist, L. W. & Vande Woude, G. F. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3937–3941.
15. Shimotohno, K., Mizutani, S. & Temin, H. M. (1980) *Nature (London)* **285**, 550–554.
16. Kleckner, N. (1977) *Cell* **11**, 11–23.
17. Heffron, F., McCarthy, B. J., Ohtsubo, H. & Ohtsubo, E. (1979) *Cell* **18**, 1153–1163.
18. Rothstein, S., Jorgensen, R., Postle, K. & Reznikoff, W. (1980) *Cell* **19**, 795–805.
19. Rubin, G. M. & Hogness, D. S. (1975) *Cell* **6**, 207–213.
20. Flavell, A. J., Cowie, A., Legon, S. & Kamen, R. (1979) *Cell* **16**, 357–371.
21. Wensink, P. C., Finnegan, D. J., Donelson, J. E. & Hogness, D. S. (1974) *Cell* **3**, 315–325.
22. Fyrberg, E. A., Kindle, K. L., Davidson, N. & Sodja, A. (1980) *Cell* **19**, 365–378.
23. Noyes, B. E. & Stark, G. R. (1975) *Cell* **5**, 301–310.
24. Paucha, E. & Smith, A. E. (1978) *Cell* **15**, 1011–1020.
25. Pelham, H. R. B. & Jackson, R. J. (1976) *Eur. J. Biochem.* **67**, 247–256.
26. Laemmli, U. K. (1970) *Nature (London) New Biol.* **227**, 680–685.
27. Bonner, W. M. & Laskey, R. A. (1974) *Eur. J. Biochem.* **46**, 83–88.
28. Miller, J. S., Ricciardi, R. R., Roberts, B. E., Paterson, B. M. & Matthews, M. B. (1980) *J. Mol. Biol.* **142**, 455–488.
29. Paterson, B. M., Roberts, B. E. & Kuff, E. L. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4370–4374.
30. Hastie, N. D. & Held, W. A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1217–1221.
31. Alwine, J. C., Kemp, D. J. & Stark, G. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5350–5354.