The EMBO Journal vol.9 no.2 pp.535 – 541, 1990

Virus-like particle formation of *Drosophila copia* through autocatalytic processing

Katsuji Yoshioka, Hiromi Honma, Makoto Zushi, Shunzo Kondo[1], Shin Togashi[2], Tadashi Miyake[2] and Tadayoshi Shiba

Laboratory of Molecular Biology, School of Hygienic Sciences, Kitasato University, 1-15-1 Kitasato, Sagamihara, Kanagawa 228, [1]Technical Section and [2]Laboratory of Cell Biology, Mitsubishi Kasei Institute of Life Sciences, 11 Minamioya, Machida, Tokyo 194, Japan

Communicated by D.J.Finnegan

Northern blot and nucleotide sequence analyses of *copia* RNA from a transfectant made by introducing a genomic *copia* into *copia*-free cells showed that the 2 kb RNA, one of the major transcripts of *copia*, is generated through splicing. Using the polymerase chain reaction (PCR), we have also found that the position of the splice sites used in *Drosophila* larvae and cultured cells originally containing *copia* is the same as that used in the transfectant. To investigate the function of the 2 kb RNA, we constructed mutant *copia*s which harboured a single point mutation at the splice site or ~3 kb deletion of the internal region corresponding to the spliced out sequence. Analyses of transfectants made by introducing these mutant *copia*s into *copia*-free cells demonstrated that the spliced 2 kb RNA contains sufficient information to make *copia* virus-like particles (VLPs). Furthermore, when *copia* RNA corresponding to the spliced RNA was translated *in vitro*, the major VLP protein was found to be released autocatalytically from its own precursor. A single amino acid substitution at the putative protease active site in the precursor prevented the processing, and resulted in accumulation of the mutant precursor *in vitro*. From these results, we conclude that *copia* VLPs are produced through autocatalytic processing of the precursor polyprotein encoded by the spliced *copia* RNA.

Key words: copia/Drosophila/retrotransposon/splicing/VLP

Introduction

The *Drosophila* transposable element *copia* is a retrotransposon that is structurally similar to retroviral proviruses (for reviews, see Rubin, 1983; Finnegan, 1985). It is ~5 kb long with long terminal repeats (LTRs) of 276 bp (Finnegan *et al.*, 1978; Levis *et al.*, 1980). Major transcripts of *copia* are 5 kb and 2 kb in length in cultured cells (Carlson and Brutlag, 1978). They start in the 5′ LTR (Flavell *et al.*, 1981) and the larger one probably extends to the 3′ LTR without splicing (Schwartz *et al.*, 1982; Emori *et al.*, 1985), analogous to retroviral genomic RNA. *Copia* contains sequences that could serve as a tRNA primer binding site adjacent to the 5′ LTR and an oligopurine tract, which could prime second strand synthesis, adjacent to the 3′ LTR (for a review, see Varmus, 1983). Indeed, we previously demonstrated that *copia* RNA directed DNA synthesis primed by a methionine tRNA (Kikuchi *et al.*, 1986). This result suggests that *copia* replicates by a mechanism similar to that of retroviral replication.

Copia RNA is the most abundant cytoplasmic poly(A)$^+$ RNA in *Drosophila* cultured cells (Finnegan *et al.*, 1978; Falkenthal and Lengyel, 1980) and can be translated *in vitro* (Falkenthal and Lengyel, 1980; Flavell *et al.*, 1980). Moreover, nucleotide sequence analysis of genomic *copia*s has revealed a single large open reading frame (ORF) capable of encoding a 1409 amino acid polyprotein (Emori *et al.*, 1985; Mount and Rubin, 1985). This polyprotein seems to encode protein products similar to those of retroviruses including a nucleic acid-binding protein, a protease, an integrase and a reverse transcriptase (Mount and Rubin, 1985). The similarity of *copia* and retroviruses has been further strengthened by the finding of virus-like particles (VLPs) containing *copia* RNA and reverse transcriptase activity in *Drosophila* cultured cells (Shiba and Saigo, 1983). The causal relationship between *copia* and the VLPs was confirmed by DNA transfection experiment (Miyake *et al.*, 1987). To date, however, the mechanism of *copia* VLP formation has been unknown.

The amounts of *copia* RNAs are modulated during the development of *Drosophila* and the highest level is in the larval stage (Flavell *et al.*, 1980; Schwartz *et al.*, 1982; Parkhurst and Corces, 1987). Schwartz *et al.* (1982) also demonstrated that the 2 kb *copia* RNA was absent or present at very low levels at each developmental stage of the organisms. From this result, together with the observation that *copia* VLPs have not been found in wild-type *Drosophila* organisms (our unpublished data), we assumed that the 2 kb RNA would relate to *copia* VLP production.

In this paper, we describe the structure and the function of 2 kb RNA. We show that the 2 kb RNA is generated through splicing and encodes sufficient information to make *copia* VLPs in *Drosophila* cultured cells. Moreover, we have found that the major VLP protein is released autocatalytically from its own precursor and the processing is prevented by a single amino acid substitution at the putative protease active site in the precursor. These results suggest that *copia* VLPs are produced through autocatalytic processing of the polyprotein precursor encoded by the spliced 2 kb RNA.

Results

The 2 kb copia RNA is generated through splicing

To investigate the structure of the 2 kb *copia* RNA, we first analysed RNA prepared from transfectant NC-1 (Miyake *et al.*, 1987) because of the following two reasons: (i) as NC-1 was made by introducing plasmid cDm2055, which carries a genomic *copia* of *D.melanogaster*, into a *copia*-free *D. hydei* cell line (KUN-DH-33), it should contain homogeneous *copia*s; (ii) the other more important feature is that NC-1 has been confirmed to produce *copia* VLPs quite similar in

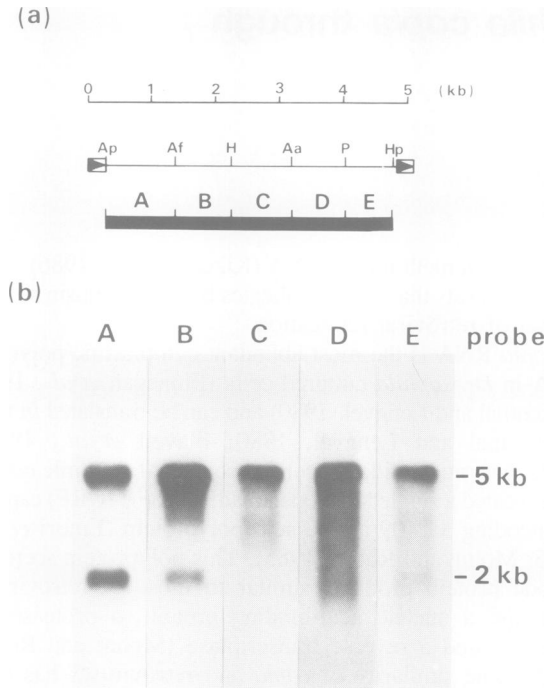


Fig. 1. Analysis of *copia* RNA using different probes. (**a**) Restriction map of *copia* in cDm2055. The arrowhead indicates the *copia* LTR. The solid bars indicate fragments used to probe Northern experiments (b). Restriction site abbreviations are Ap, *Apa*I; Af, *Afl*II; H, *Hin*dIII; Aa, *Aat*I; P, *Pst*I; Hp, *Hpa*I. (**b**) Northern blot analysis of *copia* RNA using different probes. Poly(A)$^+$ RNA from NC-1 was electrophoresed on a 1% formaldehyde−agarose gel and transferred to nitrocellulose for hybridization with ^{32}P-labelled fragments (a). Probe A, the 1.1 kb *Apa*I−*Afl*II fragment; probe B, the 0.82 kb *Afl*II−*Hin*dIII fragment; probe C, the 0.96 kb *Hin*dIII−*Aat*I fragment; probe D, the 0.82 kb *Aat*I−*Pst*I fragment; probe E, the 0.74 kb *Pst*I−*Hpa*I fragment.

various aspects (except for lower reverse transcriptase activity) to *copia* VLPs appearing spontaneously in cultured cells of *D.melanogaster*. Northern blot analysis of poly(A)$^+$ RNA from NC-1 using five different fragments (A−E in Figure 1a) as probes is shown in Figure 1b. The 2 kb RNA was hybridized with probes A, B and E, but not with probes C and D. This result reveals that the 2 kb RNA is generated through splicing and the splice sites are in B (the 5′ splice site) and in E (the 3′ splice site) respectively.

For further investigation, we constructed an NC-1 cDNA library and isolated 10 cDNA clones of the 2 kb RNA as being positive with probes A and E, and negative with probe D. Nucleotide sequence analysis of these cDNA clones and the parental genomic *copia* indicated that the 2 kb RNA is generated through splicing and contains a single ORF capable of encoding a 426 amino acid protein of 48 000 daltons (48 kd) (Figure 2). Taking this result together with our previous study (Emori *et al.*, 1985), this ORF should include the entire coding region of the major VLP protein (Figure 2). The 5′ and 3′ splice sites are at nucleotides 1604 and 4501, respectively, in NC-1 *copia* (where consensus sequences for 5′ and 3′ splice sites are found; Mount, 1982).

As a next step, we examined whether the position of the splice sites used in *Drosophila* larvae and other cultured cells originally containing *copia* is the same as that used in NC-1. To study this, the relevant RNA sequences are enzymatically amplified using the polymerase chain reaction (PCR, Saiki

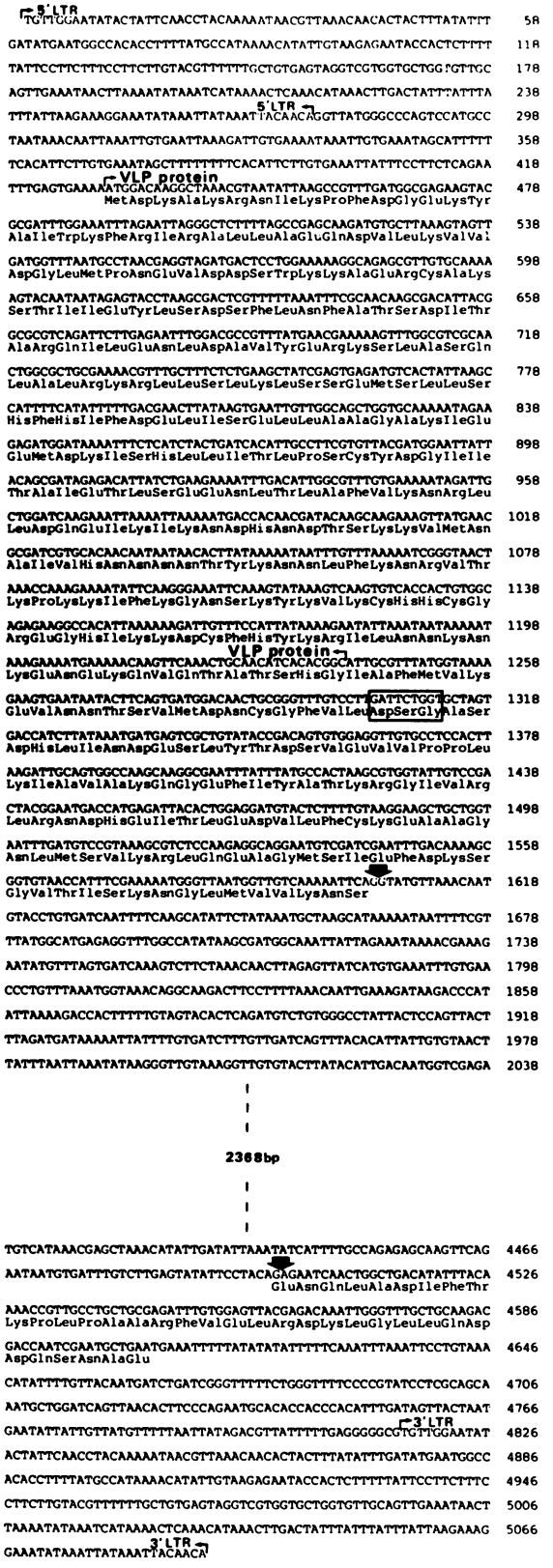


Fig. 2. Nucleotide sequence of *copia* in cDm2055. The DNA strand corresponding to the *copia* RNAs is shown. The vertical arrow indicates the splice site of the 2 kb RNA. The translation of ORF2 from the methionine codon at nucleotide 431 to the termination codon at nucleotide 4605 is shown on the bottom line. The ranges of 5′ LTR, 3′ LTR and the expected coding region of the major VLP protein (Emori *et al.*, 1985) are shown by L-shaped arrows. The box indicates the putative protease active site.

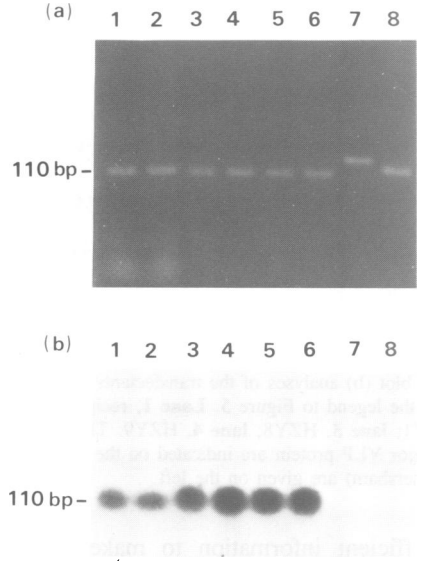


Fig. 3. Analysis of PCR amplified cDNAs. (**a**) Ethidium bromide-stained agarose gel of PCR amplified cDNAs and cDm2055 DNAs. *D.melanogaster* larvae (**lane 1**), *D.simulans* larvae (**lane 2**), NC-1 (**lane 3**), GM_2 (**lane 4**), Kc (**lane 5**), 83-8b (**lane 6**) poly(A)$^+$ RNAs oligo(dT) primed. These cDNAs were amplified with COP30 and COP34. cDm2055 DNA was amplified either with COP30 or COP57 (**lane 7**) or with COP34 and COP58 (**lane 8**). Amplified samples were electrophoresed on 4% NuSieve GTG Agarose (FMC) gel and blotted onto Hybond-N. (**b**) Southern blot of gel (a) hybridized sequentially with COP46. Final wash was at 65°C in 6 × SSC, 0.1% SDS. GM_2 (Mosna and Dolfini, 1972) and Kc (Echalier and Ohanessian, 1970) are cell lines established from embryos of *D.melanogaster*. 83-8b (T.Miyake, unpublished) is a cell line established from embryos of *D.simulans*.

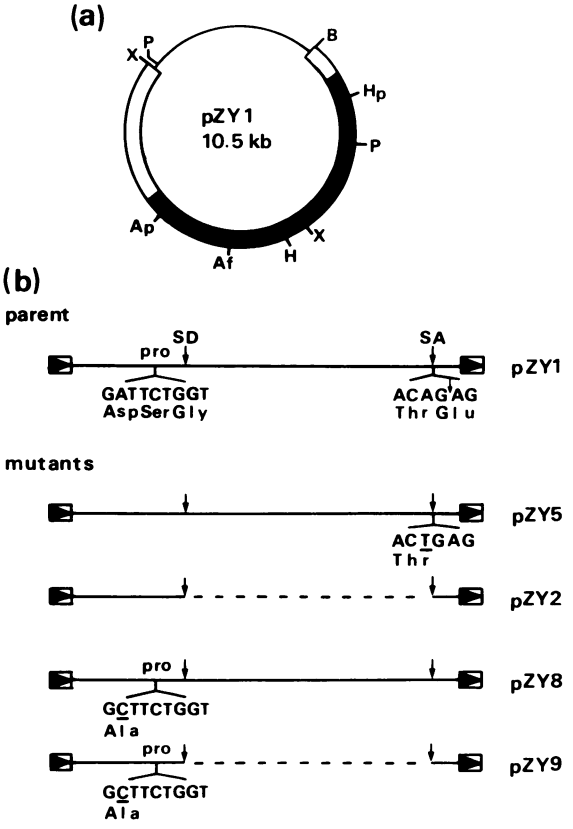
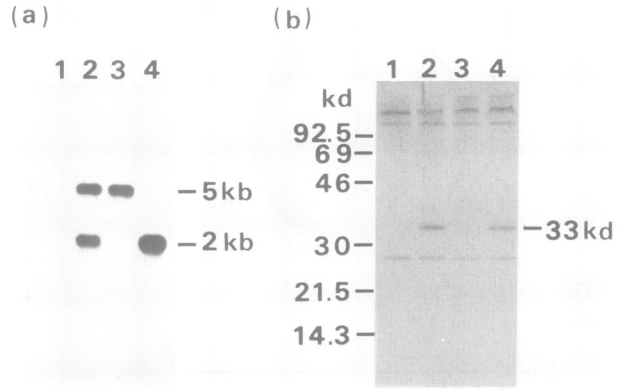


Fig. 4. (**a**) Restriction map of pZY1. Closed box indicates *copia* sequence. Open boxes indicate *Drosophila* sequence. Thin line indicates pUC18 vector. B, *Bam*HI; H, *Hin*dIII; other restriction sites are designated as in Figure 1. (**b**) Structure of *copia*s in plasmids used for DNA transfection experiment. The arrowhead indicates the *copia* LTR. The arrow indicates the splice site. SD, splice donor; SA, splice acceptor; pro, putative protease active site. The dashed line indicates the deletion of the internal region corresponding to the spliced put sequence.

et al., 1985) and screened with a synthetic oligonucleotide designed to detect specifically the spliced sequence. Total RNA was isolated from larvae and cultured cells (see legend to Figure 3) and poly(A)$^+$ RNA was selected on oligo(dT)−cellulose (Pharmacia). Double-stranded cDNA primed with oligo(dT) was amplified using specific primers COP30 and COP34 as described in Materials and methods. COP30 and COP34 bind upstream of the 5′ splice site and downstream of the 3′ splice site, respectively, used in NC-1. If the 2 kb *copia* RNA from different tissues originates through splicing and the position of the splice sites is the same as that used in NC-1, then PCR with the cDNA using COP30 and COP34 will amplify a 110 bp DNA fragment. The predicted 110 bp fragments were visualized with ethidium bromide staining after electrophoresis on an agarose gel (Figure 3a, lanes 1−6). As controls, we also amplified cDm2055 DNA using specific primers COP30 and COP57, or COP58 and COP34. COP57 and COP58 bind downstream of the 5′ splice site and upstream of the 3′ splice site, respectively, used in NC-1. The PCR with cDm2055 DNA using these two sets of primers amplifies 138 bp and 121 bp DNA fragments, respectively (Figure 3a, lanes 7,8). Furthermore, the agarose gel was blotted onto Hybond-N (Amersham), and hybridized sequentially with oligonucleotide COP46, which is specific for the spliced sequence. All 110 bp bands hybridized with COP46 but, under the same conditions, the control PCR fragments did not (Figure 3b). These results suggest that the position of the splice sites used in *Drosophila* larvae and cultured cells originally containing heterogeneous *copia* is the same as that used in transfectant NC-1.

The spliced RNA encodes sufficient information to make copia VLPs in Drosophila cultured cells

To elucidate the function of the spliced RNA, we constructed plasmids harbouring mutations in *copia* DNAs (Figure 4). Plasmid pZY1, which was made by recloning the *copia* sequence of plasmid cDm2055 (Miyake *et al.*, 1987) into plasmid pUC18 (for details, see Materials and methods), is the source from which all plasmids with mutated *copia*s originate. Plasmid pZY5 harbours a single point mutation at the 3′ splice site (AG → TG), which causes no change in amino acid sequences putatively encoded by the 5 kb *copia* RNA. Plasmid pZY2 was constructed by partially replacing pZY1 with a cDNA of the spliced RNA. Therefore, pZY2 harbours an ~3 kb deletion of the internal region corresponding to the spliced out sequence. These plasmids were transfected into *copia*-free cells and the resultant transfectants were designated HZY1, HZY2 and HZY5, respectively. Northern blot and Western blot analyses of these transfectants were carried out and the results are shown in Figure 5a and b, respectively. In the control RNA from HYZ1, the 5 kb and 2 kb RNAs were detected using a 1.1 kb *Apa*I−*Afl*II fragment of pZY1 (equal to fragment



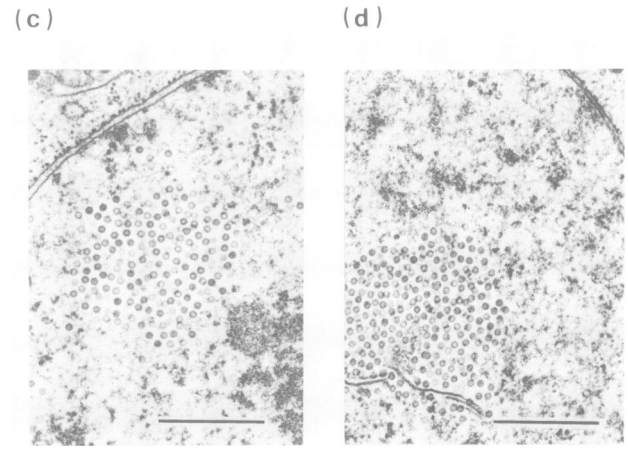


Fig. 5. Analysis of transfectants HZY1, HZY5 and HZY2. (**a**) Northern blot analysis of the transfectants. Poly(A)$^+$ RNAs from recipient KUN-DH-33 (**lane 1**), HZY1 (**lane 2**), HZY5 (**lane 3**) and HZY2 (**lane 4**) were electrophoresed on a 1% formaldehyde−agarose gel and transferred to nitrocellulose for hybridization with ^{32}P-labelled 1.1 kb *Apa*I−*Afl*II fragment of pZY1. (**b**) Western blot analysis of the transfectants. Cell lysates of recipient KUN-DH-33 (**lane 1**), HZY1 (**lane 2**), HZY5 (**lane 3**), HZY2 (**lane 4**) were electrophoresed on a 12.5% SDS−polyacrylamide gel, transferred to nitrocellulose, and subjected to immunoblot analysis using anti-VLP serum. The position and mol. wt of the major VLP protein are indicated on the right. Mol. wt markers (Amersham) are given on the left. (**c**), (**d**) Electron micrographs of the transfectants HZY1 (c) and HZY2 (d). Specimens were prepared for electron microscopy as described previously (Miyake *et al.*, 1987). Scale bar is 0.5 μm.

A in Figure 1) as a probe (Figure 5a, lane 2). As expected, only 5 kb RNA and only 2 kb RNA, structurally equivalent to the spliced RNA, were detected by the probe in RNA from HZY5 and HZY2, respectively (Figure 5a, lanes 3 and 4). The major VLP protein of relative molecular mass 33 kd was visualized using anti-VLP serum in HZY1 (Figure 5b, lane 2) in a similar manner to previous studies (Shiba and Saigo, 1983; Miyake *et al.*, 1987). Interestingly, the major VLP protein was also detected in HZY2, but not in HZY5 (Figure 5b, lanes 3 and 4). Electron microscopy indicated VLPs to be present not only in HZY1, but also in HZY2 (Figure 5c and d). In both transfectants these VLPs exist in the nuclei and appear to have some features in common. However, in HZY5 very few, if any, VLPs were detected (data not shown). Thus, the spliced 2 kb *copia* RNA should

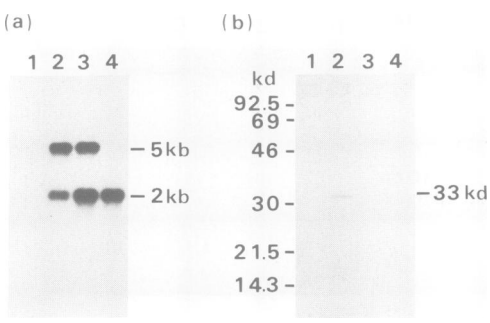


Fig. 6. Analysis of transfectants HZY8 and HZY9. Northern blot (**a**) and Western blot (**b**) analyses of the transfectants. Procedures were as described in the legend to Figure 5. **Lane 1**, recipient KUN-DH-33; **lane 2**, HZY1; **lane 3**, HZY8; **lane 4**, HZY9. The position and mol. wt of the major VLP protein are indicated on the right. Mol. wt markers (Amersham) are given on the left.

encode sufficient information to make *copia* VLPs in *Drosophila* cultured cells.

The VLP protein is produced through autocatalytic processing of the precursor polyprotein encoded by the spliced RNA

Comparison of the putative *copia* polyprotein sequence with retroviral proteases reveals that *copia* encodes a protease (Mount and Rubin, 1985). As the putative protease active site (Asp-Ser-Gly) is situated upstream of the 5′ splice site (see Figure 2), it is likely that *copia* protease participates in the formation of VLPs by the processing of its own precursor, that is, autocatalytic processing. To examine this possibility, we constructed plasmid pZY8 harbouring GAT → GCT substitution which corresponds to an Asp → Ala mutation at the putative protease active site (Figure 4). pZY8 carries the genomic *copia* identical to the parental *copia* in PZY1 except for the mutation. To clarify the effect of the mutation on the spliced RNA, we also constructed plasmid pZY9 which harbours the same GAT → GCT mutation and an ~3 kb deletion of the internal region corresponding to the spliced out sequence (Figure 4). Transfectants of these plasmids (termed HZY8 and HZY9, respectively) were obtained and subjected to Northern and Western blot analyses. The results are shown in Figure 6. No VLP protein could be detected in either HZY8 or HZY9 (Figure 6b, lanes 3 and 4), although the 2 kb *copia* RNA was present in each transfectant (Figure 6a, lanes 3 and 4). Electron microscopy showed that there were no (or very few if any) VLPs in these transfectants (data not shown).

Moreover, we carried out *in vitro* translation of the spliced 2 kb RNA and its relevant mutant RNA harbouring the mutation at the putative protease active site. For this experiment, the 1.5 kb *Apa*I−*Hpa*I fragment of pZY2 (wild-type) or pZY9 (Asp → Ala mutation), which contains the whole ORF of the spliced RNA, was recloned into plasmid Bluescript SK M13$^+$ (Stratagene), respectively. Resulting plasmids were linearized with *Not*I and used as templates for *in vitro* transcription. Transcripts were made using T7 RNA polymerase, and were translated *in vitro* in a rabbit reticulocyte lysate. In the case of the wild-type RNA, the main translation products after 60 min of incubation were a polypeptide of relative molecular mass 33 kd, which was immunoprecipitated with anti-VLP serum, and a polypeptide of relative molecular mass 23 kd (Figure 7a, lanes 2 and

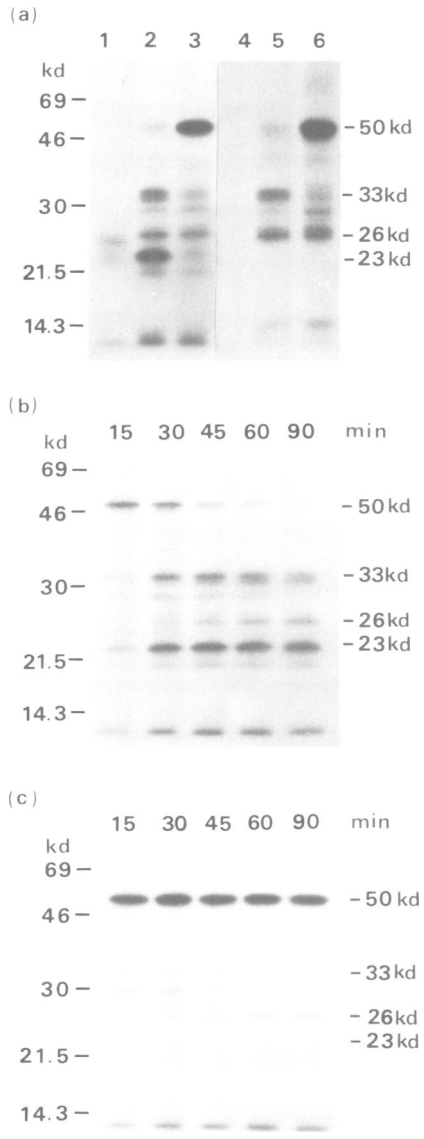


Fig. 7. Analysis of *in vitro* translation products. (**a**) Analysis of *in vitro* translation products with the spliced wild-type RNA and its relevant mutant RNA harbouring the mutation at the protease active site. T7-generated RNAs were translated for 60 min at 30°C in reticulocyte lysate in the presence of [^{35}S]methionine, and analysed on a 12.5% SDS−polyacrylamide gel. The translation products with no RNA (**lane 1**), the wild-type RNA (**lane 2**) and the mutant RNA (**lane 3**) are shown. The anti-VLP serum immunoprecipitated translation products of **lanes 1−3** are shown in **lanes 4−6**, respectively. (**b**), (**c**) Kinetic analysis of *in vitro* translation products with the wild-type RNA (b) and the mutant RNA (c). *In vitro* translation reactions were performed for the indicated periods of time and analysed as in (a) without immunoprecipitation. The positions and mol. wts of main translation products are indicated on the right. Mol. wt markers (Amersham) are given on the left.

5). On the other hand, a polypeptide of relative molecular mass 50 kd was mainly produced by the translation after 60 min of incubation with mutated RNA (Figure 7a, lane 3). This polypeptide reacted with anti-VLP serum (Figure 7a, lane 6), and its size is similar to the predicted size of full length precursor encoded by the spliced RNA. A polypeptide of relative molecular mass 26 kd which also reacted with anti-VLP serum was also detected (Figure 7a, lanes 2,3,5,6). To clarify the relationship between these products, kinetic analysis of *in vitro* translation products was performed. In the case of the wild-type RNA, the full length precursor was first detected at 15 min of incubation. With time, the precursor disappeared, and the predominant 33 kd and 23 kd polypeptides appeared after 30 min of incubation (Figure 7b). In contrast, the precursor accumulated with translation of the mutated RNA (Figure 7c).

These results suggest that the spliced 2 kb *copia* RNA is translated initially as the polyprotein precursor which corresponds to retroviral *gag* polyprotein precursor, followed by the production of the 33 kd major VLP protein and the 23 kd protein by autocatalytic processing. The Asp → Ala mutation at the putative protease active site in the precursor prevented the processing, and resulted in accumulation of the mutant precursor *in vitro*. Thus, the 2 kb *copia* RNA also encodes protease and the 23 kd polyprotein which did not react with anti-VLP serum seems to be the protease, because our previous study (Emori *et al.*, 1985) has shown that the 33 kd polyprotein comes from the amino-terminal end of the 50 kd precursor and does not contain the putative protease active site. In addition, the 26 kd product which also reacted with anti-VLP serum would be either a processed product of *copia* protease or a degradation product of the precursor. So far we have not been able to observe *trans* processing of the mutant (Asp → Ala) full length polypeptides by the translation products with the wild-type RNA as in the case for cauliflower mosaic virus (Torruella *et al.*, 1989).

Discussion

We have shown in this paper that the 2 kb *copia* RNA from transfectant NC-1 made by introducing a genomic *copia* into *copia*-free cells is generated through splicing and the splice sites used in NC-1 are also used in *Drosophila* larvae and cultured cells originally containing *copia*. Recently, Miller *et al.* (1989) reported the nucleotide sequence of some 2 kb *copia* RNA. The position of the splice sites is found to be the same from our experiments. Furthermore, we have found that the spliced RNA contains sufficient information to make *copia* VLPs in *Drosophila* cultured cells and the major VLP protein is released autocatalytically from its own precursor. Since integrase and reverse transcriptase domains of *copia* are situated in the internal region corresponding to the spliced out sequence, *copia* VLP formation does not depend on the expression of these enzymes.

The proteolytic factor of the autolysis should be *copia* protease, because the Asp → Ala mutation at the putative protease active site in the polyprotein precursor similarly prevented the processing with HIV protease expressed as *pol* polyprotein precursor in *Escherichia coli* (Mous *et al.*, 1988). Based on sequence homologies with aspartic proteases, including retroviral proteases (Toh *et al.*, 1985; Mount and Rubin, 1985) and the mutagenesis study of the putative protease active site (this paper), *copia* protease would belong to the aspartic protease family. Retroviral proteases have been mapped either at the 5′ end of the *pol* polyprotein or at the 3′ end of the *gag* polyprotein, or between the *gag* and *pol* polyproteins. In *copia*, the protease is mapped within the *gag* polyprotein as is the case for avian retroviruses (Dittmar and Moelling, 1978; Vogt *et al.*, 1979). However, *copia gag* polyprotein shows features different from those of avian retroviruses. Rous sarcoma virus *gag* polyprotein synthesized *in vitro* is stable and does not exhibit

autocatalytic processing (von der Helm, 1977). In addition, although most retroviral *gag* polyproteins are encoded by the first ORF of unspliced genome length RNAs, *copia gag* polyprotein is encoded by the spliced RNA.

Nucleotide sequence analysis of genomic *copia*s has revealed a single large ORF capable of encoding a 1409 amino acid protein (Emori *et al.*, 1985; Mount and Rubin, 1985). However, the *copia* in cDm2055 is 5091 bp in length and contains a 53 bp deletion (between nucleotides 1993 and 2000 in Figure 2) within the ORF. Because of the deletion, the ORF is split in two. Analysis of *D.melanogaster* genomic DNAs by the PCR using specific primers (corresponding to nucleotides 1892–1911 and 2098–2117 in Figure 2) has indicated that a large number of *D.melanogaster* genomic *copia*s do not contain the deletion (data not shown). Here, we redesignate the ORF capable of encoding 1409 amino acid protein as ORF1, and the ORF of the spliced RNA, which could encode a 426 amino acid protein corresponding to *gag* polyprotein, as ORF2. The amino acid sequence of ORF2 is in large part colinear with that of ORF1. Translation of ORF1 would produce a large polyprotein similar to retroviral *gag–pol* polyprotein as previously pointed out by Mount and Rubin (1985).

In retroviruses it is well known that proteolytic processing of *gag* and *gag–pol* polyproteins is mediated by a virus-encoded protease and that the processing is essential for their replication (for a review, see Kräusslich and Wimmer, 1988). HIV protease expressed in *E.coli*, for example, shows autolysis of its own precursor and cleavage of *gag* and *pol* polyproteins in *trans* (Debouck *et al.*, 1987; Graves *et al.*, 1988; Mous *et al.*, 1988; Hansen *et al.*, 1988; Le Grice *et al.*, 1988; Kräusslich *et al.*, 1989). In *copia*, the protease is encoded by ORF2 (*gag*) and shows autolysis as well. Thus, it should be quite reasonable to consider that *copia* protease also mediates the processing of the polyprotein encoded by ORF1 (*gag–pol*); that is, it may release an active integrase, reverse transcriptase and so on. Since the major VLP protein has been mapped at the amino-terminal region of both ORF1 and ORF2 (Emori *et al.*, 1985), we had expected anti-VLP serum to detect polyprotein precursors in transfectants HZY8 and HZY9, which harbour protease mutant *copia*s, respectively. The inability to do so (Figure 6b, lanes 3 and 4), reveals that the mutant precursors are labile *in vivo*.

An open question is why the translation products of the wild-type RNA (corresponding to the spliced 2 kb *copia* RNA) could not complement the processing of the mutant full length polypeptides. This result may reflect the idea that the autocatalytic processing occurs only during or after assembly of the wild-type full length precursors themselves. The assembly of the *gag* precursors to produce immature particles has been known in yeast Ty element and mammalian retroviruses, such as HIV-1 (Adams *et al.*, 1987; Gheysen *et al.*, 1989). Furthermore, there was little effect on the autocatalytic processing by pepstatin A at the concentration of 1 mM in the assay system used in this study (data not shown). Pepstatin A is a specific inhibitor of aspartic proteases including not only non-viral proteases, such as pepsin, but also retroviral proteases. The 50% inhibitory concentration (IC_{50}) of pepstatin A for pepsin, cathepsin D and renin is 1.4×10^{-8} M, 8.8×10^{-9} M and 6.6×10^{-6} M, respectively (Aoyagi *et al.*, 1973). In contrast, the IC_{50} values for retroviral proteases are in the range 1.0×10^{-4} M–5.0×10^{-4} M (Katoh *et al.*, 1987). As mentioned by Katoh *et al.* (1987), less effectiveness of pepstatin A against retroviral proteases could reflect the strict substrate specificity of retroviral proteases. Thus, the result of the inhibition study by pepstatin A in this paper may reflect higher specificity of *copia* protease than retroviral proteases. Characterization of *copia* protease is in progress.

Although it is not yet known how the ratio of the 5 kb RNA (which is probably unspliced genome length RNA; Emori *et al.*, 1985) to the spliced 2 kb RNA is controlled, both *copia* RNAs should be required for *copia* replication. Therefore, an increase in the amount of the spliced RNA, which is absent, or present in a much smaller amount, than that of the 5 kb RNA at each developmental stage of *Drosophila* (Schwartz *et al.*, 1982 and our unpublished data), may lead to higher frequencies of *copia* transposition in the organism. Experiments of this nature are presently under way.

Materials and methods

RNA preparation and analysis

Total RNA from *Drosophila* larvae and cultured cells was prepared by the guanidium/caesium chloride method as described by Maniatis *et al.* (1982) and poly(A)$^+$ RNA was selected on oligo(dT)–cellulose. For Northern blot analysis, poly(A)$^+$ RNA was electrophoresed through 1% agarose–2.2 M formaldehyde gel and transferred to nitrocellulose. Hybridization was carried out at 42°C in 50% formamide, 5 × SSC, 50 mM sodium phosphate buffer (pH 6.8), 1 × Denhardt's solution and 100 μg/ml denatured calf thymus DNA. Blots were washed at 65°C in 0.5 × SSC, 0.1% SDS.

Construction and screening of cDNA library

An oligo(dT)-primed cDNA library of NC-1 was constructed using kits (Amersham, catalogue no. RPN1256 and RPN1257). This was screened with ^{32}P-labelled A, D and E fragments of *copia* (see Figure 1). Plaque hybridization was carried out at 65°C in 6 × SSC, 5 × Denhardt's solution, 0.5% SDS and 100 μg/ml denatured calf thymus DNA. Filters were washed at 65°C in 0.5 × SSC, 0.1% SDS.

DNA sequencing

DNA sequencing was performed by the dideoxy chain termination method (Sanger *et al.*, 1977) using Sequenase (USB).

Oligonucleotides and site-directed mutagenesis

The following oligonucleotides were synthesized on an Applied Biosystems 381A DNA synthesizer. SPLICE3, ATTCCTACTGAGAATCA 17 mer 5′ end at 4491; COP30, GCGGTGTAACCATTTCGAAA 20 mer, 5′ end at 1557; COP34, TCGTAACTCCACAAATCTCG 20 mer 5′ end at 4562; COP46, GTCAAAAATTCAGAGAATCAACTGGC 26 mer 5′ end at 1592, 3′ end at 4513; COP56, TGTCCTTGCTTCTGGTG 17 mer 5′ end at 1297; COP57, ACCTCTCATGCCATAAACGA 20 mer 5′ end at 1694; COP58, TCATTTTGCCAGAGAGCAAG 20 mer 5′ end at 4442. Numbers indicate nucleotide positions in Figure 2. Oligonucleotides were used for mutagenesis (SPLICE3 and COP56), for PCR amplification (COP30, COP34, COP57 and COP58), or as a probe (COP46). COP34 and COP57 are complementary to the 5 kb *copia* RNA sequences, and COP46 corresponds to the 2 kb *copia* RNA sequences. Others correspond to the 5 kb *copia* RNA sequences.

Mutagenesis was carried out essentially by the method of Kunkel (1985) using a kit (TAKARA, catalogue no. 6060). Mutated sequences were confirmed by DNA sequencing.

PCR amplification

One microgram of poly(A)$^+$ RNA was converted into double-stranded cDNA using a kit (Amersham, catalogue no. RPN1256). Oligo(dT) was used to make the first strand cDNA. The PCR was carried out using *Thermus thermophilus* DNA polymerase (TOYOBO). One-twentieth of double-stranded cDNA or 20 pg cDm2055 DNA was added to 0.2 mM of each dNTP, 100 pmol of each primer, enzyme buffer recommended by the manufacturers, one unit of polymerase in a 100 μl volume. The samples were initially denatured at 95°C for 5 min followed by successive cycles

of annealing at 40°C for 30 s, extension at 70°C for 1.5 min, and denaturation at 95°C for 30 s. *T.thermophilus* DNA polymerase was added after the first denaturation step. The sample was overlaid with several drops (~100 μl) of mineral oil and subjected to 30 cycles. After the last cycle, the sample was incubated for an additional 5–10 min at 70°C.

Plasmid construction

cDm2055 was digested with *Bam*HI; the released 10 kb *Bam*HI fragment containing a genomic *copia* was subcloned into the *Bam*HI site of pUC18 to give rise to plasmid cDm2055-2. cDm2055-2 was digested completely with *Xba*I and self-ligated, then it was digested with *Hin*dIII, filled in with DNA polymerase I large fragment and self-ligated to give rise to plasmid cDm2055-3. pZY1 was constructed by inserting the 4.6 kb *Xba*I fragment of cDm2055 into the unique *Xba*I site of cDm2055-3. pZY2 was constructed by replacing the 3.3 kb *Afl*II–*Hpa*I fragment of pZY1 with the 0.45 kb *Afl*II–*Hpa*I fragment of a spliced *copia* cDNA clone prepared from NC-1 cDNA library. For site-directed mutagenesis, two plasmids pSK1 and pSK2 were constructed. The 1.6 kb *Pst*I–*Bam*HI fragment and the 2.0 kb *Apa*I–*Hin*dIII fragment of pZY1 were subcloned between the *Pst*I and *Bam*HI sites and the *Apa*I and *Pst*I sites to give rise to pSK1 and pSK2, respectively. The 1.8 kb *Hin*dIII–*Pst*I fragment of pZY1 was inserted between the *Hin*dIII and *Pst*I sites of pSK1 containing the mutation at the 3′ splice site (which was made using SPLICE3) to give rise to pSK3. pZY5 was constructed by replacing the 2.5 kb *Hin*dIII–*Hpa*I fragment of pZY1 with the 2.5 kb *Hin*dIII–*Hpa*I fragment of pSK3. pZY8 and pZY9 were constructed by replacing the 1.1 kb *Apa*I–*Afl*II fragments of pZY1 and pZY2 with the 1.1 kb *Apa*I–*Afl*II fragments of pSK2 containing the mutation at the putative protease active site (which was made using COP56), respectively. For *in vitro* transcription, 1.5 kb *Apa*I–*Hpa*I fragments of pZY2 and pZY3 were inserted between the *Apa*I and *Sma*I sites of Bluescript SK M13$^+$ to give rise to plasmids pHK2 and pHK9, respectively.

Western blot analysis

5×10^5 cells of each transfectant were resuspended in 10 μl of the lysis buffer, consisting of 50 mM Tris–HCl pH 6.8, 2% SDS, 5% glycerol, 5% 2-mercaptoethanol, 0.01% bromophenol blue. The protein was heated for 5 min at 100°C and subjected to 12.5% SDS–polyacrylamide gel electrophoresis (Laemmli, 1970). For the Western blot, the gel was electroblotted to nitrocellulose in 2.5 mM Tris base, 192 mM glycine, 0.1% SDS, 20% methanol. Specific bands were visualized using anti-VLP serum and the Vectastain ABC kit (Vector Systems).

Transcription with T7 RNA polymerase and in vitro translation

DNA of pHK2 and pHK9 (1 μg of each) was digested with *Not*I, extracted with phenol, extracted with chloroform:isoamylalcohol (24:1), precipitated with ethanol, rinsed with 70% ethanol, dried and resuspended in water. Capped RNAs for *in vitro* translation were generated by T7 RNA polymerase using a kit (Pharmacia, catalogue no. 27-9276–01). *In vitro* translation and immunoprecipitation were carried out as described previously (Shiba and Saigo, 1983). *In vitro* complementation analysis was carried out by mixing an equal volume of two different kinds of translation mixtures made in rabbit reticulocyte lysates with wild-type and mutant RNAs for 60 min of incubation, respectively. Inhibition analysis of the autocatalytic processing was carried out as follows: pepstatin A was dissolved in dimethyl sulphoxide (DMSO) and added to the translation mixture so that the final concentration of DMSO was 2% (v/v). There was a slight precipitation of pepstatin A under these conditions.

DNA transfection and electron microscopy

DNA transfection and electron microscopy were carried out as described previously (Miyake *et al.*, 1987).

Acknowledgements

We thank N.Takamatsu for helpful discussion. This work was supported in part by a grant from the Kitasato Research Foundation to K.Y., and was also supported in part by a grant from the Ministry of Education, Sciences and Culture, Japan, to T.S.

References

Adams,S.E., Mellor,J., Gull,K., Sim,R.B., Tuite,M.F., Kingsman,S.M. and Kingsman,A.J. (1987) *Cell*, **49**, 111–119.
Aoyagi,T., Yagisawa,Y., Kumagai,M., Hamada,M., Morishima,H., Takeuchi,T. and Umesawa,H. (1973) *J. Antibiot.*, **26**, 539–541.
Carlson,M. and Brutlag,D. (1978) *Cell*, **15**, 733–742.
Debouck,C., Gorniak,J.G., Strickler,J.E., Meek,T.D., Metcalf,B.W. and Rosenberg,M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8903–8906.
Dittmar,K.J. and Moelling,K. (1978) *J. Virol.*, **28**, 106–118.
Echalier,G. and Ohanessian,A. (1970) *In Vitro*, **6**, 162–172.
Emori,Y., Shiba,T., Kanaya,S., Inoue,S., Yuki,S. and Saigo,K. (1985) *Nature*, **316**, 773–776.
Falkenthal,S. and Lengyel,J.A. (1980) *Biochemistry*, **19**, 5842–5850.
Finnegan,D.J. (1985) *Int. Rev. Cytol.*, **93**, 281–326.
Finnegan,D.J., Rubin,G.M., Young,M.W. and Hogness,D.S. (1978) *Cold Spring Harbor Symp. Quant. Biol.*, **42**, 1053–1063.
Flavell,A.J., Ruby,S.W., Toole,J.J., Roberts,B.E. and Rubin,G.M. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 7107–7111.
Flavell,A.J., Levis,R., Simon,M.A. and Rubin,G.M. (1981) *Nucleic Acids Res.*, **9**, 6279–6291.
Gheysen,D., Jacobs,E., de Foresta,F., Thiriart,C., Francotte,M., Thines,D. and De Wilde,M. (1989) *Cell*, **59**, 103–112.
Graves,M.C., Lim,J.J., Heimer,E.P. and Kramer,R.A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2449–2453.
Hansen,J., Billich,S., Schulze,T., Sukrow,S. and Moelling,K. (1988) *EMBO J.*, **7**, 1785–1791.
Katoh,I., Yasunaga,T., Ikawa,Y. and Yoshinaka,Y. (1987) *Nature*, **329**, 654–656.
Kikuchi,Y., Ando,Y. and Shiba,T. (1986) *Nature*, **323**, 824–826.
Kräusslich,H.-G. and Wimmer,E. (1988) *Annu. Rev. Biochem.*, **57**, 701–754.
Kräusslich,H.-G., Ingraham,R.H., Skoog,M.T., Wimmer,E., Pallai,P. and Carter,C.A. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 807–811.
Kunkel,T.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 488–492.
Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
Le Grice,S.F.J., Mills,J. and Mous,J. (1988) *EMBO J.*, **7**, 2547–2553.
Levis,R., Dunsmuir,P. and Rubin,G.M. (1980) *Cell*, **21**, 581–588.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Miller,K., Rosenbaum,J., Zbrzezna,V. and Pogo,A.O. (1989) *Nucleic Acids Res.*, **17**, 2134.
Miyake,T., Mae,N., Shiba,T. and Kondo,S. (1987) *Mol. Gen. Genet.*, **207**, 29–37.
Mosna,G. and Dolfini,S. (1972) *Chromosoma*, **38**, 1–9.
Mount,S.M. (1982) *Nucleic Acid Res.*, **10**, 459–472.
Mount,S.M. and Rubin,G.M. (1985) *Mol. Cell. Biol.*, **5**, 1630–1638.
Mous,J., Heimer,E.P. and Le Grice,S.F.J. (1988) *J. Virol.*, **62**, 1433–1436.
Parkhurst,S.M. and Corces,V.G. (1987) *EMBO J.*, **6**, 419–424.
Rubin,G.M. (1983) In Shapiro,J.A. (ed.), *Mobile Genetic Elements*. Academic Press, NY, pp. 329–361.
Saiki,R.K., Scharf,S., Faloona,F., Mullis,K.B., Horn,G.T., Erlich,H.A. and Arnheim,N. (1985) *Science*, **230**, 1350–1354.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schwartz,H.E., Lockett,T.J. and Young,M.W. (1982) *J. Mol. Biol.*, **157**, 49–68.
Shiba,T. and Saigo,K. (1983) *Nature*, **320**, 119–124.
Toh,H., Ono,M., Saigo,K. and Miyata,T. (1985) *Nature*, **315**, 691.
Torruella,M., Gordon,K. and Hohn,T. (1989) *EMBO J.*, **8**, 2819–2825.
Varmus,H.E. (1983) In Shapiro,J.A. (ed.), *Mobile Genetic Elements*. Academic Press, NY, pp. 411–503.
Vogt,V.M., Wight,A. and Eisenman,R. (1979) *Virology*, **98**, 154–167.
von der Helm,K. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 911–915.

Received on August 25, 1989; revised on November 17, 1989