FEBS Letters 350 (1994) 147–150

FEBS 14371

Identification of spliced RNA species of *Drosophila melanogaster gypsy* retrotransposon

New evidence for retroviral nature of the *gypsy* element

Sergei N. Avedisov*, Yurii V. Ilyin

V.A. Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Vavilov str. 32, Moscow 117984, Russian Federation

Received 4 July 1994

Abstract

We have identified a novel RNA species of *Drosophila melanogaster gypsy* retrotransposon that is ca. 2 kb in length and corresponds to the third open reading frame (ORF3) of the *gypsy* element. This RNA is generated by splicing of the primary *gypsy* transcript, as is the case for retroviral *env* gene expression. Therefore, the striking resemblance between *gypsy* and retroviruses has now been extended by this study to the expression strategies of these retroelements. The primary structure of spliced RNA was determined, and its analysis shows that both *gypsy* subfamilies (6K and 7K) apparently are able to encode functionally active ORF3 translation products.

Key words: Splicing; *env*-gene; Retrotransposon *gypsy*; *Drosophila melanogaster*

1. Introduction

gypsy (mdg4) is one of the best characterized *Drosophila* retrotransposons [1–4]. A 7-kb *gypsy* transcript is expressed at relatively high levels in a variety of *D. melanogaster* cell types [2] and contains three long open reading frames (ORFs). The first and second ORFs are revealed in all known retrotransposons and retroviruses reflecting the universalism of the strategies by which they transpose [5]. The presence of the third long ORF is a feature of only few retrotransposons, and the functions of ORF3 are currently unknown. The third ORF in size and location corresponds to the *env* genes of retroviruses, which encode the viral glycoproteins and permit efficient cell-to-cell (horizontal) transmission, i.e., infection [6]. However, there is no direct experimental data for pronounced infectivity of *gypsy* element, as is the case for vertebrate retroviruses, although a recent report provides some genetic evidence to support this idea [7]. Therefore, it is clear that the role of the third *gypsy* ORF must provide a clue to understanding the principal differences between retrotransposons and retroviruses. As a first and necessary step toward solving this problem, we demonstrate that, as in retroviruses, the ORF3 expression is mediated by splicing of the full-length *gypsy* RNA.

*Corresponding author. Fax: (7) (095) 135 1405.

2. Materials and methods

2.1. RNA isolation and Northern blot analysis

Total RNA was extracted from adult flies, pupae, larvae, embryos and cultured cells by the method of Chomczynski et al. [8], and poly(A)$^+$ RNA was purified by poly(U)-sepharose chromatography, electrophoresed in 1% formaldehyde/agarose gel, transferred to Hybond N nylon membrane (Amersham) and hybridized to ^{32}P-labeled probes as described by Maniatis et al. [9].

2.2. cDNA-PCR amplifications

For reverse transcription, 50–200 ng of polyadenylated RNA treated with RNase-free DNase (Promega) was denatured at 70°C and cooled to 40°C in the presence of 20 pmol of oligomer p1. cDNA synthesis was carried out with 100 U of M-MLV reverse transcriptase (GIBCO BRL) following the instructions of the manufacturer. Reverse transcription mixture was included in a final volume (100 μl) of PCR reaction containing 20 mM Tris-HCl (pH 8.7), 15 mM ammonium sulfate, 0.1 mg/ml gelatine, 6 mM $MgCl_2$, a 0.1 mM concentration of each dNTP, a 100 nM concentration of p1 and p2 primers, 150 ng/ml BSA, and 4U of *Tth* polymerase (Fermentas). After 30–40 cycles at 95°C for 1 min, 55°C for 1 min and 70°C for 1–2 min, DNA fragments amplified in PCR experiments were purified on agarose gels and cloned into pUC19 by standard procedures [9].

2.3. DNA sequencing

DNA sequencing was performed on double-stranded plasmid DNA, by the dideoxy chain termination method [10].

3. Results

In order to establish the expression strategy of *gypsy* ORF3, we carried out a transcription analysis of endogenous *gypsy* copies for several *D. melanogaster* strains, including three so-called 'Mutator strains' (MS, MSn^1 and MSn^2) [11]. These strains are known to contain extraordinarily high copy number of *gypsy* element

SSDI 0014-5793(94)00758-6

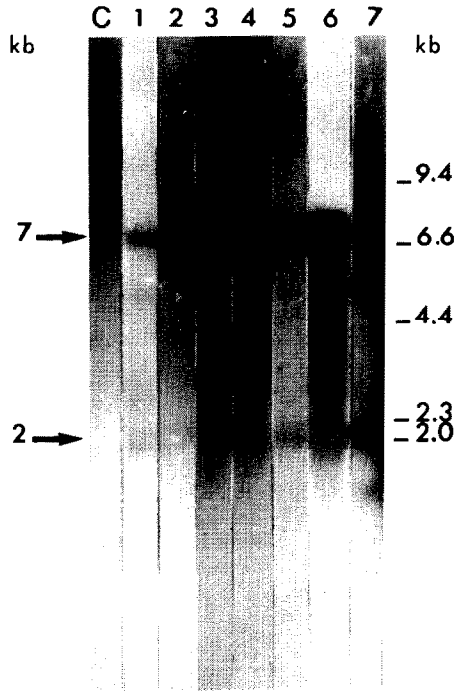

Fig. 1. Northern blot analysis of *gypsy* poly(A)$^+$ RNA. The following strains were used: Oregon R flies (1); MS flies (2); MSn^2 flies (3); MSn^1 flies (4); MSn^1 embryos (5); MSn^1 larvae (6); Schneider 2 cell culture (7); control, *D. hydei* cell culture (C). The full-length and spliced transcripts of *gypsy* are marked by the arrows.

and are characterized by high frequency of *gypsy* transposition. The results of Northern blot analysis with probe from the *gypsy* ORF3 (0.90-kb *NdeI–EcoRI* fragment) are presented in Fig. 1. Identical results were obtained, when the 0.23-kb *XhoI–BglII* fragment of the *gypsy* LTR was used as a probe. In all cases, the analysis revealed the presence of 2-kb RNAs besides the well characterized 7-kb (full-length) *gypsy* transcripts. At the same time this new RNA species did not hybridize to the probes derived from the ORF1 or ORF2 (data not shown). These results strongly suggest that the ORF3 expression is mediated by RNA splicing similar to that seen for retroviral *env* gene expression, as shown in Fig. 2.

Sequence analysis of the *gypsy* copy reported by Marlor et al. [3] (this copy belongs to the 6K [12] *gypsy* subfamily) reveals the presence of a motif similar to 5′ splice site consensus [13] within the leader untranslated region of *gypsy*. Two sequences similar to 3′ splice site consensus of *Drosophila* and separated by 65 bp were found within the 5′ portion of *gypsy* ORF3 (Fig. 3). We have recently sequenced the full-length copy from another (7K) *gypsy* subfamily (Avedisov and Ilyin, manuscript in preparation). All three splice site consensus-like sequences mentioned above are also conserved in this 7K copy, but a point substitution (nucleotide position 5,504; we use the coordinates that have been reported by Marlor et al. [3]) introduces the translation stop (UAA) at the beginning of the ORF3. Thus, in the case of the upstream located translation initiation site, this mutation must prevent ORF3 translation. Therefore, the hypothetical implication of the cryptic splice site in RNA processing might result in substantial alterations of the third ORF expression pattern for 7K *gypsy* copies.

The cDNA-PCR technique was used to analyze the structure of 2-kb transcripts. In all cases, the 1-st strand synthesis was carried out using primer p1, which hybridizes to the central part of *gypsy* ORF3 (position 5918 to 5937). cDNA preparations were amplified using p1 and p2 (oligomer specific to the leader untranslated region of *gypsy*; position 527 to 548) as primers. The results (Fig. 4a) show that, in all cases, p1 and p2 amplify only one DNA fragment of 428 bp. Sequencing of these fragments have allowed us to determine the nucleotide sequence of 2-kb RNA at the splice junction (Fig. 4b). The 5′ splice site is located within the leader region of *gypsy* element and far upstream of the ORF1 translation initiation site. The 3′ splice site, at least for all *Drosophila* strains examined, is located downstream of the nonsense mutation found in 7K copies. Besides, the presumed initiator ATG codon is generated directly by splicing. These data suggest that the nonsense mutation positioned downstream of the beginning of ORF3 should not influence ORF3 expression, and the sequence located between the 3′ end of the ORF2 and the 3′ splice site is untranslated. Our data indicate, therefore, that both *gypsy* subfamilies (6K and 7K) not only are transpositionally competent in cultured cells [4], but apparently are able to encode functionally active translation products of ORF3.

As controls, all the poly(A)$^+$ RNA preparations (and, in some cases, the p1-primed cDNA preparations) were amplified in parallel using the following sets of primers: p2 and p3, p1 and p4, p1 and p5. In all cases, no fragment

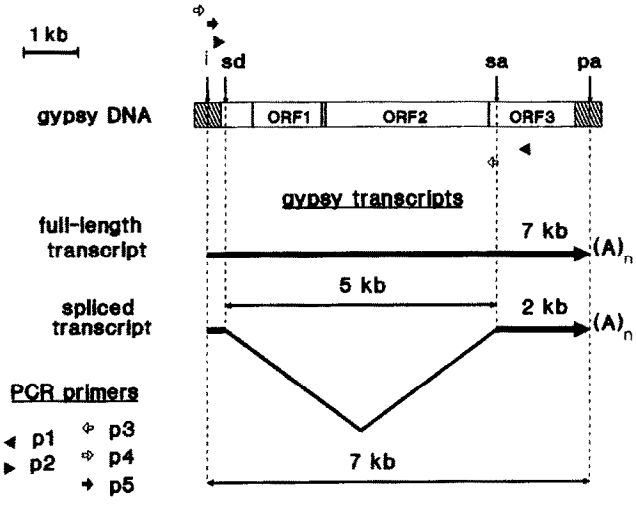

Fig. 2. Genetic organization and transcription products of *gypsy* retrotransposon. The LTRs of the 7.5-kb element are marked as hatched boxes. ORF1, ORF2 and ORF3 are shown as open boxes. The *gypsy* 7-kb genomic and 2-kb subgenomic transcripts are indicated directly below the element. The transcription initiation site (i) polyadenylation signal (pa), and splice sites (sd and sa) are indicated by vertical arrows.

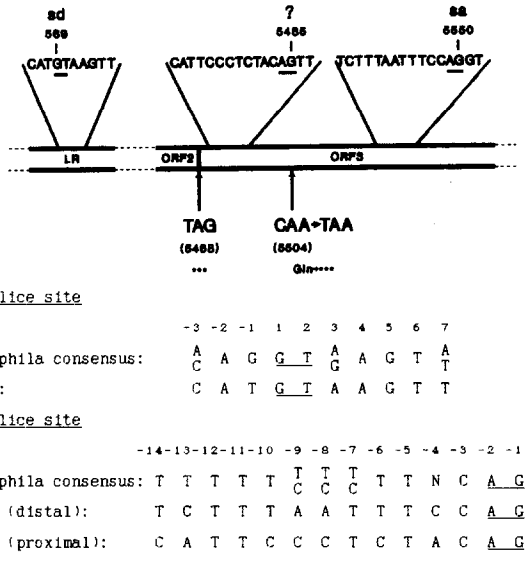

Fig. 3. Hypothetical alternative splicing during *gypsy* ORF3 expression. The nucleotide sequences of *gypsy* leader region (LR) and ORF3 5′ part are shown. The putative 3′ (sa) and 5′ (sd) splice sites are compared to *Drosophila* splice consensus sequences (12), where N can be any nucleotide. A question mark above the proximal 3′ splice site means that it is not functional. The termination codons are marked by the upward arrows.

was amplified with p1 and p4 (p4 is located just upstream of the start point of *gypsy* transcription [14], position 187 to 206), while PCR analysis using p1 and p5 (p5 is located just downstream of the transcription start point, position 254 to 273) yielded a single fragment, the size of which correctly corresponds to the predicted size of the spliced RNA fragment. These observations, combined with the known length of the spliced transcript, therefore suggest that it is the full-length *gypsy* RNA that is subject to splicing. To control for detection of any alternatively spliced RNA, primers p2 and p3 were used (p3 is located between the acting 3′ splice site of *gypsy* and cryptic splice acceptor site; position 5,514 to 5,533). In this case, no fragment was amplified by PCR, further suggesting in favor of the absence of alternative splicing during ORF3 expression for all *D. melanogaster* strains and developmental stages tested.

4. Discussion

The presence of the *env* gene, which functions to permit the entry of the virion into a specific cell, is considered as the only essential difference between retroviruses and retrotransposons; therefore, the elucidation of ORF3 function(s) and expression features must help to clear up whether are there any principal differences between these two classes of retroelements, or whether the *gypsy* element has an extracellular phase and is infectious. If the latter possibility is indeed the case, it would show that the *gypsy* is a retrovirus. The identification of spliced *gypsy* RNA provides further evidence of striking *gypsy* resemblance to retroviruses in expression strategies.

Data presented here also demonstrate that the regulation at the level of RNA splicing may be a factor contributing substantially to the retrotransposon expression. It is the full-length *gypsy* RNA that serves as the template for ORF1 and ORF2 protein synthesis and represents an intermediate for reverse transcriptase-mediated transposition [4]. At the same time, the ratio of the full-length to spliced transcript band intensities (as may be seen in Fig. 1) varies from one *Drosophila* strain to another. Thus, the quantitative ratio of ORF3 translation products to ORF1 and ORF2 proteins is possibly controlled by splicing efficiency. Therefore, the host splicing factors may be involved in the regulation of *gypsy* transposition frequency.

The further important observation to note concerns the obvious absence of alternative splicing during ORF3 expression. The N-terminal portions of native surface proteins encoded by the *env* genes are known to be extremely important in productive infectivity of retroviral

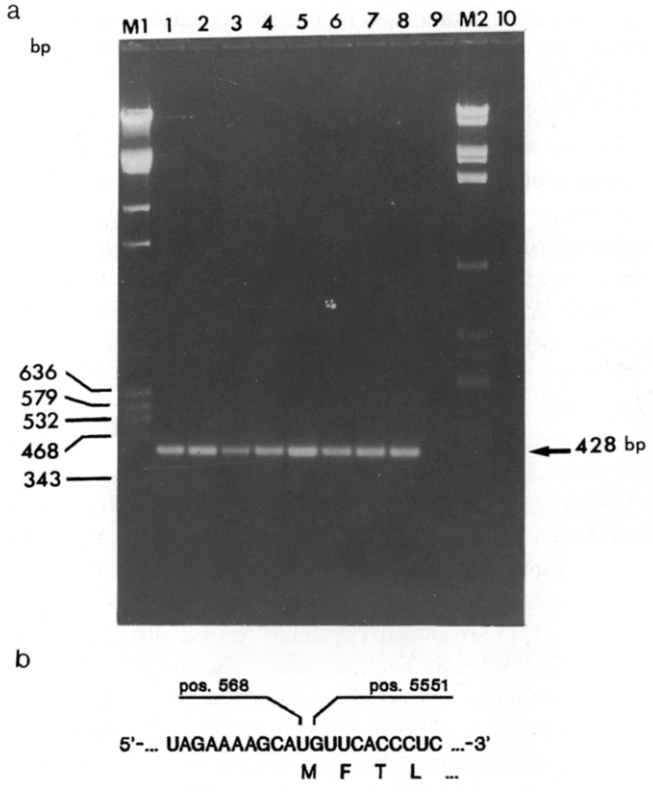

Fig. 4. cDNA-PCR analysis of *gypsy* spliced RNA. (a) EtBr-stained agarose gel of cDNA-PCR amplified subgenomic *gypsy* RNA. The RNAs from the following strains were used: Oregon R flies (1); MS flies (2); MSn^2 flies (3); MSn^1 flies (4,9,10); MSn^1 embryos (5); MSn^1 larvae (6); MSn^1 pupae (7); Schneider 2 cell culture (8). Poly(A)$^+$ RNA was amplified using the following primers: p1 and p2 (1–8); p1 and p4 (9), or p2 and p3 (10). The arrow represents the 428-bp amplified fragment. Lanes M1 and M2: lambda phage DNA digested with *Pvu*II and *Ssp*I. (b) Nucleotide sequence of spliced ORF3-transcript in the splice junction.

particles because just these portions recognize the specific receptors located at the surface of host cell [6]. It is the variability of the N-terminal *env*-peptide structure that provides for a wide variety of retroviral target-cells. This variability can arise not only because of high mutability level (the outside portions of the envelope proteins are changing the most rapidly of all retroviral functional domains) [15] but, at least in some cases [16], appears to be the result of alternative splicing mechanisms involved in *env*-gene expression. However, during analysis of ORF3-specific transcripts we found no evidence for alternative splicing: two sequences similar to 3′ splice site consensus were found in ORF3 of *gypsy*, but only one of them, distal, functions as 3′ splice site. Moreover, the splice junction structure of *gypsy* subgenomic RNA independently determined by A. Pelisson and V. Corces for another *D. melanogaster* strain (personal communication) is the same as we have determined. If ORF3 translation products are indeed analogous in function to the retroviral *env* proteins, this may give evidence for high target-cell specificity of ORF3-proteins. One can thus hypothesize that the distribution of *gypsy*-like retrotransposons among the different *Drosophila* species should reflect a such specificity. In this case, the detailed structural characterization of these transposable elements should be useful in the functional analysis of ORF3 proteins and in defining evolutionary relationships between retroviruses and ORF3-containing retrotransposons.

Acknowledgements: We thank Drs. A.I. Kim and N.G. Shostack for providing the fly stocks. This work was supported by Grant 93-04-06180 from the Russian Foundation of Fundamental Research and by Grant N M9C000 from the International Science Foundation.

References

[1] Bayev, A.A., Lyubomirskaya, N.V., Dzhumagaliev, E.B., Ananiev, E.V., Amiantova, I.G. and Ilyin, Y.V. (1984) Nucleic Acids Res. 12, 3707–3723.
[2] Parkhurst, S.M. and Corces, V.G. (1987) EMBO J. 6, 419–424.
[3] Marlor, R.L., Parkhurst, S.M. and Corces, V.G. (1986) Mol. Cell. Biol. 6, 1129–1134.
[4] Lyubomirskaya, N.V., Avedisov, S.N., Surkov, S.A. and Ilyin, Y.V. (1993) Nucleic Acids Res. 21, 3265–3268.
[5] Boeke, J.D. and Corces, V.G. (1989) Annu. Rev. Microbiol. 43, 403–434.
[6] Varmus, H. and Brown, P. (1989) in: Mobile DNA (Berg, D.E. and Howe, M.H. Eds.) pp. 53–108, American Society for Microbiology, Washington, DC.
[7] Kim, A.I., Terzian, C., Santamaria, P., Pelisson, A., Prud'homme, N. and Bucheton, A. (1994) Proc. Natl. Acad. Sci. USA 91, 1285–1289.
[8] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[9] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
[10] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[11] Kim, A.I., Lyubomirskaya, N.V., Belyaeva, E.S., Shostack, N.G. and Ilyin, Y.V. (1994) Mol. Gen. Genet. 242, 472–477.
[12] Lyubomirskaya, N.V., Arkhipova, I.R., Ilyin, Y.V. and Kim, A.I. (1990) Mol. Gen. Genet. 223, 305–309.
[13] Mount, S.M., Burks, C., Hertz, G., Stormo, G.D., White, O. and Fields, C. (1992) Nucleic Acids Res. 20, 4255–4262.
[14] Arkhipova, I.R., Mazo, A.M., Cherkasova, V.A., Gorelova, T.V., Schuppe, N.G. and Ilyin, Y.V. (1986) Cell 44, 555–563.
[15] McClure, M.A., Johnson, M.S., Feng, D.-F. and Doolittle, R.F. (1988) Proc. Natl. Acad. Sci. USA 85, 2469–2473.
[16] Ciminale, V., Pavlakis, G.N., Derse, D., Cunningham, C.P. and Felber, B.K. (1992) J. Virol. 66, 1737–1745.