Invertebrate retroviruses: *ZAM* a new candidate in *D.melanogaster*

P.LebLANC, S.DessET, B.Dastugue and C.Vaury¹

Unité INSERM U384, Faculté de Médecine, Place Henri Dunant, 63000 Clermont-Ferrand, France

¹Corresponding author

***ZAM*, a new retroelement of *Drosophila melanogaster*, was identified as a mutational insertion at the *white* locus. It displays all the structural features of a vertebrate retrovirus. Its three open reading frames encode predicted products resembling the products of the *gag*, *pol* and *env* genes of retroviruses. Its transcription gives rise to an 8.6 kb full-length RNA and a 1.7 kb spliced message for the *env* gene. The latter encodes an envelope protein that is typical of elements having an extracellular phase of the life cycle. The identification of a *ZAM* envelope retrogene provides evidence that *ZAM* is mobilized through a reverse transcriptional process in the germ line of flies. We report that *ZAM* is distributed differently among *D.melanogaster* strains. Two stocks out of >15 tested display a *ZAM* high copy number, with numerous copies distributed on chromosomal arms. This high copy number is associated with a high transcriptional rate of *ZAM*. The existence of these two categories of strains offers a new genetic system in which the properties of a potential invertebrate retrovirus can be tested.**

Keywords: *Drosophila*/envelope/retrogene/retrovirus

Introduction

For a long time, retroviruses were considered to be restricted to vertebrates. Retrovirus-like elements had been described in eukaryotic genomes, but most of them displayed only two open reading frames (ORFs), similar to the *gag* and *pol* genes of retroviruses between the long terminal repeats (LTRs) (Finnegan, 1994). Recently, five elements from insects have emerged with an additional ORF corresponding in sequence, size and location to retroviral *env* genes. These elements are *gypsy* (Marlor *et al.*, 1986), *17.6* (Saigo *et al.*, 1984), *297* (Inouye *et al.*, 1986), *tom* (Tanda *et al.*, 1988) and *TED* (Friesen and Nissen, 1990). These similarities have prompted the suggestion that they may represent endogenous insect retroviruses. This has been demonstrated recently for *gypsy*. Kim *et al.* (1994) have demonstrated that *gypsy* is infectious, since it can move from one cell to another and even infect flies. Such an extracellular existence requires an Env-like product. Indeed, PéliSSon *et al.* (1994) have detected a 2.1 kb RNA with a structure corresponding to the spliced *env* mRNA of retroviruses, while Song *et al.*

(1994) have shown that this *env* mRNA is able to encode a protein.

The demonstration of *gypsy* infectivity has been possible for two reasons (Bucheton, 1995). First, two categories of stocks exist in *Drosophila melanogaster*, i.e. flies devoid of functional *gypsy*, and flies in which the *gypsy* element is active and produces infectious particles. Second, an assay is available to detect the acquisition of *gypsy* elements, since the *ovo* gene of *Drosophila* contains a hot spot for its insertion. A *gypsy* insertion into a dominant female-sterile mutation of *ovo*, *ovo*^{DI} produces mutations with a recessive phenotype, restoring fertility to heterozygous females. The further characterization of new retrovirus-like elements will require similar genetic tools for the induction and visualization of new mobilization events.

Here we report the identification of such an element in *D.melanogaster*. *ZAM* displays all the structural characteristics of a retroviral provirus, with three ORFs corresponding to the *gag*, *pol* and *env* genes, flanked by two LTRs. The transcriptional pattern of *ZAM* is similar to that seen in active retroviruses. It gives rise to two kinds of transcript. One corresponds to a full-length RNA that will be reverse transcribed before its integration into the genome. The second corresponds to a subgenomic *env* RNA. The *ZAM* protein potentially encoded by this latter mRNA displays all the characteristics of retroviral Env proteins determined by their transmembrane localization and function (length, hydrophobicity, cleavage site, potential *N*-glycosylation and cysteine residues). These data strongly suggest that *ZAM* has the capacity for extracellular transmission.

Furthermore, we present evidence that *ZAM* has recently been mobilized and amplified in a strain of *D.melanogaster* where its recent *de novo* insertion upstream of the *white* gene has been visualized. The existence of a parental strain with a low copy number of *ZAM* and its descendant with a high copy number may permit investigation of the mechanism of retrovirus-like element mobilization.

Results

A phenotypic reversion of the mutated eye phenotype of flies bearing the w^{IR6} allele is due to a de novo 9 kb insertion

In a search for second site modifier genes affecting alleles due to *I factor* insertions, we performed a PM mutagenesis (Robertson *et al.*, 1988) on the line bearing the *w^{IR6}* allele (the mutagenesis will be described elsewhere). This allele is due to the insertion of the non-LTR retrotransposon *I factor* into the first intron of the *white* gene, and gives a brown–orange eye phenotype to flies (Lajoinie *et al.*, 1995).

From the mutagenesis, individual flies displaying a modified eye colour phenotype were established as lines.

Fig. 1. The structure of the w^{IR6Rev} allele is represented as follows: the dark boxed regions indicate exons while the thin lines joining them indicate introns. Exons 1 and 2 of the *white* gene are the only *white* exons represented. DNA upstream of the gene is shown as a dashed line. The insertion sites of the *I factor* and the 9 kb sequence are indicated by a triangle. The orientation of the *I factor* is indicated by an arrow.

One of these lines carried a mutation on the third chromosome that partially suppressed the brown–orange w^{IR6} phenotype to dark brown (*Su78* strain). After >20 generations, a male with wild-type eye colour was recovered from this line. The molecular event that caused the reversion in this male was analysed. The cause of the reversion was genetically localized to the X chromosome close to the *white* gene. This finding prompted the examination of the genomic DNA corresponding to the *white* locus in order to define the area of the locus affected in the revertant. Through Southern blot experiments using a battery of restriction enzymes, we found that the entire transcriptional unit of the w^{IR6} allele was unchanged: each exon and intron was still present as well as the complete *I factor* as judged from their fine restriction map. However, the *white* locus had suffered a 9 kb insertion ~3 kb upstream from the *white* start site of transcription between the *Pst*I and *Hpa*I sites at positions 6523 and 7305 respectively according to O'Hare *et al.* (1984) (Figure 1). This 9 kb insertion proved to be absent from this site in the w^{IR6} strain and in the lines used for or isolated from the PM mutagenesis (data not shown). This new allele of *white* has been called w^{IR6Rev} .

The 9 kb insertion is a new retrovirus-like element

In order to isolate and clone the entire 9 kb insertion, PCR experiments were performed (see Materials and methods). Oligonucleotides (o1 and o2) located at the *Pst*I and *Hpa*I restriction sites (Figure 2) were used as primers for the amplification. The restriction map of the insertion was established and verified by Southern blotting of genomic DNA (Figure 2A). The PCR products were cloned as sub-fragments encompassing the complete insertion (Figure 2B). The insertion displays the typical structure of a *gypsy*-like retrotransposon (Figure 2A), but clearly represents a family distinct from transposable elements described so far in *D.melanogaster*. We call this novel retrovirus-like element *ZAM*. *ZAM* is flanked at both ends by LTRs that are 473 bp long and are identical in sequence. Each of them displays 7 bp long inverted repeats (IR) at their 5' and 3' extremities.

Like all other retroviral LTRs, those of *ZAM* are

bordered by a primer-binding site (PBS), complementary to a cellular tRNA, and a polypurine tract (PPT). The *ZAM* PBS has 12 bp complementary to the 3' end of *D.melanogaster* tRNA^{Ser} (Cribbs *et al.*, 1987) at positions 473–484. Computer analysis of the *ZAM* LTRs with the BISANCE program (Dessen *et al.*, 1990) showed two strong consensus motifs corresponding to a TATA box. The first is located at position 208 (TATAAA) and the second, which corresponds to a retroviral TATA box, is at position 312. This latter is a GAATAAAG signal that is potentially recognized as a polyadenylation signal as well. The *ZAM* insertion site in the w^{IR6Rev} allele shows a target duplication of CGCGCG surrounding the element.

Predicted gene products of *ZAM* resemble those of retroviruses

ZAM sequence analysis with the DNA strider program (Marck, 1988) showed that *ZAM* has three ORFs able to encode polypeptides similar to the Gag, Pol and Env proteins found in vertebrate retroviruses such as human immunodeficiency virus type 1 (HIV-1) or Moloney murine leukaemia virus (MoMuLV) (Shinnick *et al.*, 1981; Morrow *et al.*, 1994) or in retrovirus-like elements such as 297, 17.6, *gypsy*, *tom* and *TED*. In each case, these similarities extend over domains containing amino acids that are highly conserved among the retroviruses.

ZAM ORF1 is 343 amino acids long. An asparagine-rich domain located in the N-terminal region and a more significant basic segment in the C-terminus of the predicted protein have been identified. There is no distinctive acidic domain in the C-terminus of the protein as reported for most of the elements described so far. *ZAM* ORF2 encodes a 1217 amino acid protein including sequences similar to other Pol polypeptides (Xiong and Eickbush, 1990). Analysis with the BLAST-X program (Altschul *et al.*, 1990) revealed four identifiable domains: the protease domain (prt) displays the short sequence Asp–Thr–Gly (DTG), which is thought to form the active site of aspartyl (or acidic) proteases (Rawlings and Barrett, 1995); the reverse transcriptase (rt) characterized by the conserved motif, called the YxDD box, that identifies its catalytic centre (Yuki *et al.*, 1986); the RNase H (rnH); and the

Fig. 2. ZAM organization and molecular structure. (A) Restriction map of the 9 kb insertion. Boxes represent long terminal repeats (LTRs), dotted lines the upstream non-transcribed region of *white*. Gene organization: *gag*-, *pol*- and *env*-like open reading frames are symbolized by rectangles. (B) Subcloning strategy of the insertion. Black lines represent fragments of the insertion cloned in the pBluescript vector or amplified by long template PCR (P3). Oligonucleotides used for the PCR amplification of the complete insertion (o1 and o2) or the P3 fragment (5U and p2) are symbolized by arrowheads.

integrase domain (int). This latter displays potential sites for binding of zinc ions (so-called 'zinc fingers') and a DD35E motif characteristic of the active site of integrases (Polard and Chandler, 1995). Figure 3A–D depicts partial alignments of these domains with those of several invertebrate retrovirus-like elements or vertebrate retroviruses.

Sequence analysis of ORF3 (546 amino acids) revealed similarities with Env-like proteins of 297, 17.6, *gypsy*, *tom* and *TED*. A domain conserved in these retrovirus-like elements is presented in Figure 3E. It displays an Arg-x-Lys-Arg conserved domain (RxKR) which is considered to be a consensus proteolytic cleavage site (Klenk and Garten, 1994).

ZAM is present at low copy number in most *D.melanogaster* strains but not in the *w^{IR6Rev}* strain

The genomic distribution of ZAM has been studied by Southern blotting experiments on various *D.melanogaster* strains. Typical results obtained by probing equal amounts of *Pst*I–*Bgl*II-digested DNAs with the BH probe (Figure 2B) are presented in Figure 4. Numerous hybridizing bands were visualized, indicating that ZAM is repeated in the *Drosophila* genome. However, the copy number of ZAM is variable among the strains. Most of them (we have tested >15 independent strains including each strain used for or isolated from the PM mutagenesis) display a low copy number of ZAM elements (LCN strains). Two exceptions have been noted, the line bearing the *w^{IR6Rev}* allele and *Charolles*, an old laboratory strain captured in the wild around 1955, which display a high copy number of ZAM in their genome (HCN strains). The *Su78* strain, from which a male with the *w^{IR6Rev}* allele was recovered, displays an intermediate status according to its ZAM copy number. It is higher than in *w^{IR6}* but lower than in *w^{IR6Rev}* (data not shown). A faint pattern of common bands is observed in all the strains examined, while additional bands that differ from one strain to the other are present.

The pattern of bands observed in *w^{IR6Rev}* and its parental strain *w^{IR6}* clearly indicates that ZAM has been mobilized and its copy number increased recently.

The chromosomal location of the ZAM elements present in various strains was analysed by *in situ* hybridization to polytene chromosomes of third instar larvae. Figure 5 shows some typical results obtained when chromosomes were hybridized with the *Hpa*I–*Pst*I probe containing the complete ZAM element surrounded by upstream *white* sequence (Figure 2). The high-copy *Charolles* strain (Figure 5A) shows ~15 hybridization signals dispersed on the chromosomal arms and surrounding the chromocentre. A weak hybridization signal may be seen in the heterochromatic part of the genome. Like *Charolles*, the *w^{IR6Rev}* strain (Figure 5B) displays ~15 hybridization signals, but they are located mainly on the X chromosome. In contrast, low-copy *Canton S* (Figure 5C) shows only faint hybridization signals located in the chromocentre; while no signal has been identified on the chromosome arms of this strain. The *w^{IR6}* strain (Figure 5D) displays the same pattern of signal hybridizations as *Canton S* except for two signals located on the X and third chromosome respectively.

ZAM RNAs are quantitatively and qualitatively variable among the strains

In order to determine the pattern of transcription of ZAM, we prepared total RNAs from adult *w^{IR6}* and *w^{IR6Rev}* flies, LCN and HCN strains, respectively. The nylon filter was first probed with a BH riboprobe (Figure 6B) corresponding to the ZAM *env* gene, and secondly probed with a clone of the actin gene as a control of loading. Typical results are presented in Figure 6A. Two main transcripts of 8.6 and 1.7 kb have been identified in the *w^{IR6Rev}* strain. The 8.6 kb transcripts are not detected in *w^{IR6}*. Although a larger quantity of RNA was loaded in the case of *w^{IR6}* (see the intensity of the band revealed by the actin control of charge), only the 1.7 kb transcript was observed in this strain.

A

PROT_ZAM	48	T L N L O O K F P L S F L I D T G S N N S F I
PROT_TED	26	P Y T E F S D P P L K F L I D T G A N Q S F I
PROT_Gypsy ^m	74	I E R R L A G R T L K M L I D T D A A K N Y I
PROT_297	16	I K I V Y K G R S Y K C L I D T G S T I N M I
PROT_17.6	16	I T I K Y K E N N L K C L I D T G S T V N M T
PROT_tom	15	T H T I T L F D K T Y P L I D T D S N K G Q F
PROT_MoMuLV	13	I T L K V G G Q P V T F L V D T G A Q H S V L
PROT_HIV2	96	V T A Y I E G Q P V E V L E D T G A D D S I V

B

RT_ZAM	443	M P F G L K N A P A T F Q R V M D S V L G D L N G T - - - I C L E V Y L D D I I F S E S L Q K H L
RT_TED	417	M P M G L K N S P S T F Q R V M D N V L R G L Q N N - - - I C L V Y L D D I I V Y S T S L Q E H L
RT_Gypsy ^m	309	L P F G L R N A S S I F Q R A L D V L R E Q I G K - - - I C Y V Y V D D V I I F S E N E S D H V
RT_297	333	M P F G L R N A P A T F Q R C M N N I L R P L L N K - - - H C L V Y L D D I I F S T S L T E H L
RT_17.6	334	M P F G L K N A P A T F Q R C M N D I L R P L L N K - - - H C L V Y L D D I I V F S T S L D E H L
RT_tom	294	M P F G L R N A P A T F Q R C M N N I L R P L L N K - - - H C L V Y L D D M I I F S T S L D E H L
RT_MoMuLV	308	L P Q G F K N S P T L F D E A L H R D L A D F R I Q H P D L I L L Q Y V D D L L L A A T S E L D C Q
RT_HIV2	333	L P Q G W K G S P A I F Q Y T M R Q V L E P F R K A N Q D V I I I Q Y M D D I L I A S D R T D L E -

C

RNH_ZAM	619	L T T D - A S N F A L G A V L S Q G S L Q N D R P V C F A S R T L S D T E V N Y S T I E K E M L A I T
RNH_TED	593	L T T D - A S N F A I G A V L S Q G P I G S D K P V C Y A S R T L N E S E L N Y S T I E K E L L A I V
RNH_Gypsy ^m	497	L T T D - A S A S C I G A V L S Q - - - E G R P I T M I S R T L K Q P E Q N Y A T N E R E L L A I V
RNH_297	511	L T T D - A S N L A L G A V L S Q - - - N G H P I S F I S R T L N D H E L N Y S A I E K E L L A I V
RNH_17.6	511	L T T D - A S D V A L G A V L S Q - - - D G H P L S Y I S R T L N E H E I N Y S T I E K E L L A I V
RNH_tom	471	L T T D - A S N L A L G A V L S Q - - - D N H P I S F I S R T L N D H E L N Y S T I E K E L L A I V
RNH_MoMuLV	488	W Y T D G S S L L Q E G Q R K A G - - - A A V T T E T E V I W A K A L P A G T S A Q R A E L I A U T
RNH_HIV2	642	F Y T D G S C N R Q S K E G K A G - - - Y V T D K G R D K V K V E Q T T N Q A E L E A F A M A U T

D

IN_ZAM	870	H 4aa H 29aa CETC	938	D 52aa D 35aa E
IN_TED	890	H 5aa H 29aa CTIC	960	D 52aa D 35aa E
IN_gypsy ^m	746	H 3aa H 29aa CRVC	813	D 52aa D 36aa E
IN_297	758	H 4aa H 29aa CNIC	826	D 52aa D 33aa E
IN_17.6	758	H 4aa H 29aa CSIC	826	D 52aa D 33aa E
IN_tom	719	H 4aa H 29aa CRVC	787	D 52aa D 33aa E
IN_MoMuLV	848	H 3aa H 32aa CKAC	917	D 58aa D 35aa E
IN_HIV2	755	H 3aa H 23aa CAQC	826	D 51aa D 35aa E

Zinc Finger

DDE Domain

E

ENV_ZAM	101	R E K R G L I N G L G S L V K V V T G N M D A N D A K N I
ENV_TED	114	R V K R G L I D G L G S I V K S V T G N L D Y Q D A L K Y
ENV_Gypsy ^m	119	R I A R - S L D F L G T A L K V V A G T E D A T D L F K I
ENV_297	102	R N K R G F I N I V G S G F K Y L F G T L D E N D R V E I
ENV_17.6	87	R N K R G L I N I V G S V F K Y L F G T L D E N D R V D I
ENV_tom	93	R Q K R G L F N F V G S A F K F L F G T L D D N D R I Q F

Fig. 3. Multiple alignment of the deduced amino acid sequence of the *pol* regions. Comparison of the partial amino acid sequence of conserved regions of protease (A), reverse transcriptase (B), RNase H (C), integrase (D) of the *pol* ORF and a conserved region (E) of the *env*-like ORF. Amino acids that are conserved between ZAM and the others members analysed are shaded black or in bold, whereas amino acids which are similar are shaded in grey. Dashes indicate gaps introduced to preserve the alignment. MoMuLV, Moloney murine leukemia virus; HIV2, human immunodeficiency virus type 2. Accession numbers (for Pol and Env respectively) are as follows: *TED*, B36329 and C36329; *gypsy*, GNFFG1 and ENV1_DROME; 297, B24872 and ENV2_DROME; 17.6, GNFF17 and Y172_DROME; *tom*, S34639 and S34640; MoMuLV, POL_MLVMO and HIV2, 1072794.

The 5' ends of the transcripts from *w^{IR6Rev}* were determined by primer extension (see Materials and methods). Transcription starts at position 329 of the 5' LTR, suggesting that the TATA box at position 312 predicted by computer analyses is used for initiation (see Figure 6C). The 3' end of the transcripts has been determined by RT-PCR. ZAM RNAs are polyadenylated. Polyadenylation

occurs 26 bp downstream of the AATAAA site of the 3' LTR, at sequence CAAGCAGC-(A)_n (position 8306, Figure 6C).

Only one site corresponding to an initiation start site of transcription and one site of polyadenylation have been identified in our RNA preparations, suggesting that the 8.6 and 1.7 kb transcripts possess identical 5' and 3' ends.

Fig. 4. Genomic distribution of *ZAM*. Southern transfer experiment showing the degree of heterogeneity in the distribution of the *ZAM* elements in the genomes of *D.melanogaster* strains. Equal amounts of DNAs were digested with *Pst*I and *Bgl*II and probed with the BH fragment of *ZAM* (Figure 2). Strains used for these experiments are noted at the top of the autoradiograph.

In addition, these results indicate that the 8.6 kb *ZAM* RNA transcript encompasses the full length of *ZAM*, starting in the 5' LTR and ending in the 3' LTR, and potentially could correspond to an intermediate of transposition.

RNAs migrating at ~1.7 kb are present in all the strains studied. Since envelope mRNA of retroviruses such as HIV-1 or *gypsy* are produced by splicing of the full-length transcript that eliminates the *gag* and *pol* ORFs (Varmus and Brown, 1989; Pélissier *et al.*, 1994), we have analysed the structure of the *ZAM* 1.7 kb transcript.

The *ZAM* element encodes a spliced ORF3 subgenomic transcript with the characteristics of a functional retroviral *env* mRNA

Northern blots showed that the 1.7 kb transcript does not hybridize with the *pol*-specific BB riboprobe (Figure 2C), suggesting that this 1.7 kb band could correspond to a specific envelope transcript (data not shown). Through RT-PCR experiments, we isolated cDNA fragments amplified with a primer targeted to the 5' untranslated region of *ZAM* (5U) in combination with primers corresponding to various sites in the *env* gene (E1, E2, E3, E4 and E5) (Figure 6C). Fragments of the sizes expected for an amplification from a spliced RNA lacking the *gag* and *pol* sequences were observed after electrophoresis of the

PCR products (data not shown). One generated from the primers 5U and E5 was sequenced and shown to correspond to the predicted leader region-ORF3 fusion product (Figure 6D).

The identified spliced junction displays typical consensus 5' and 3' splice sites characterized by the presence of the dinucleotides GT at the donor site and AG at the acceptor site. The donor splice site is located at position 531, 59 bp downstream from the beginning of the tRNA PBS; the acceptor splice site is located at position 6387 (Figure 6D). The leader region and ORF3 are in-frame in the spliced message, suggesting that the first methionine codon in the leader region (position 494) is used as a translation initiation site (Figure 6D). Translation of this cDNA revealed a potential 551 amino acid polypeptide. Figure 7 gives the amino acid sequence of the predicted translation product of the subgenomic RNA and highlights the structural motifs typical of functional envelope polypeptides found in the *ZAM* envelope (Hunter and Swanstrom, 1990). The N-terminal stretch of 22 hydrophobic amino acids (position 3–24) fits statistical predictions for signal peptides. One potential cleavage site has been identified with the signalp program, which predicts the presence and location of peptide cleavage sites (Von Heijne, 1983). A second hydrophobic region of 19 amino acids from residues 485 to 503 is located at a position expected for a transmembrane (TM) domain (Coffin *et al.*, 1990). The cytoplasmic tail would correspond to residues 506–551. Eleven putative *N*-glycosylation sites conforming to the consensus sequence Asp-x-Ser or Thr (N-x-S/T), and six cysteines residues similar to those known in retroviruses to mediate attachment between the surface and the transmembrane proteins via disulfide bonds, are found upstream of the TM domain.

Evidence of a processed *ZAM env* retrogene

In the course of characterizing the spliced *env* mRNA, we performed an analysis of the *ZAM* transcripts present in *Charolles*. The transcripts observed on Northern blots were similar to those observed in the *wJR6Rev* strain. However, in RT-PCR experiments using the 5U and E1 oligonucleotides, a 1.5 kb fragment was amplified that hybridized with the BH probe. Its sequence revealed that this fragment has been amplified from an RNA identical in its splicing pattern to the *env* mRNA of the *wJR6Rev* strain. However, one difference was discovered: a 148 bp deletion located between positions 7308 and 7456 in the *ZAM* envelope sequence.

In order to test if this deleted ORF3 fragment could have been generated from the amplification of an endogenous deleted *ZAM* element present in *Charolles*, we performed a PCR experiment using the oligonucleotides 5U and E1 on genomic DNA from *Charolles*. We also performed this amplification starting from DNA preparations of *wJR6Rev* and two other strains (Figure 8A). The blot was hybridized with a complete *ZAM* probe. With all the strains, genomic PCR amplified a 7.5 kb fragment which corresponds to full-length *ZAM*. A few other bands were observed, including a 1.5 kb fragment detected in *Charolles* and not in the other strains tested.

Sequence analysis of this product showed that it displays the same structural characteristics as the 1.5 kb RT-PCR

Fig. 5. *In situ* hybridization on polytene chromosomes of four *D.melanogaster* strains using the complete *ZAM* element as a probe. (A) *Charolles*, (B) *w^{IR6Rev}*, (C) *Canton S* and (D) *w^{IR6}*. Arrowheads show the hybridization signals, CC the chromocentre and X symbolizes the X chromosome.

product described above, i.e. the *env* subgenomic structure plus the 148 bp deletion.

The existence of such a genomic *ZAM* envelope sequence strongly suggests that it arose by the reverse transcription of a *ZAM* ORF3 transcript and subsequent integration into the host genome. The internal deletion presumably represents an independent mutation. Experiments were designed to elucidate its exact pathway of formation. Two possibilities were considered: it could have been formed via a true retrovirus-like process (retrotransposon process) occurring during *ZAM* mobilization (Figure 8B and B1; Yoshioka *et al.*, 1991; Goodchild *et al.*, 1995) or via a less specific mechanism involving its reverse transcription by a potential nuclear reverse transcriptase leading to a so-called processed pseudogene (Figure 8C, Tchénio *et al.*, 1993). Processed pseudogenes generated by this latter mechanism are characterized by three canonical features: they lack upstream promoter sequence, have no intron and display a poly(A) tail at their 3' end. A sequence integrated by a process related to transposition of an LTR-containing retrotransposon would also lack introns, but would be surrounded by complete LTRs and lack a poly(A) tail.

To distinguish between these two pathways, we analysed the 3' end of the element found in *Charolles* by an inverse PCR approach using specific primers. One primer (ASO) was targeted to the splice junction, and the second (SDO) to the sequence junction created by the 148 bp deletion

(Figure 8B1). A 1.2 kb fragment was amplified, cloned and sequenced. We found that this deleted *ZAM* element displays a complete reconstituted 3' LTR and no poly(A) tail. This demonstrates that this element was created by retrotransposition. In addition, this deleted element is flanked by a CGCG sequence which is identical to 4 bp of the target duplication generated by *ZAM* insertion upstream of *white*.

Discussion

ZAM structure is similar to that of mammalian retroviruses

Here we have described the identification of *ZAM*, a new element in *D.melanogaster* that displays all the structural and functional characteristics of an invertebrate retrovirus. The structure of *ZAM* is comprised of three ORFs corresponding to the *gag*, *pol* and *env* genes, flanked by two LTRs. Each of the genes encodes putative products homologous to retroviral proteins.

The predicted translation product of *ZAM* ORF1 displays some similarities to the *gag* product of the retrovirus-like elements 297 and *gypsy*. In the *ZAM* Gag protein, we failed to identify Cys-His motifs (Covey *et al.*, 1986). Nevertheless it has numerous basic amino acids that have been hypothesized to play a role in the interaction with genomic RNA (Mesnard and Carrière, 1995). *gypsy*, 17.6, 297, *tom* and *TED* display a strong acid domain in the

Fig. 6. Expression and structure of ZAM transcripts. (A) Northern blot of total RNA isolated from *wIR6* and *wIR6Rev* adult flies probed with the BH riboprobe (B). The sizes of the two major transcripts observed are shown (in kb) on the right hand margin. The lower part of the panel shows the level of actin transcript when the same blot was hybridized with the *Drosophila Actin 5C* gene as a control for the amount of RNA loaded in each lane. (B) Organization of the ZAM element with locations of the different ORFs. (C) Structure of genomic and subgenomic transcripts of the ZAM element. The schematic drawing shows the unspliced genomic RNA as well as the spliced ORF3 transcript. Numbers refer to the nucleotides at the splicing donor (531) and acceptor sites (6387), the position where polyadenylation occurred (8306) and the site where the initiation of transcription begins (329). 5U, E1, E2, E3, E4 and E5 are oligonucleotides used for PCR amplifications and sequence analysis. (D) Nucleotide sequence of the ZAM splice sites. The two genomic sequences are those of the leader and the beginning of ORF3. The sequence of the subgenomic RNA is that of the PCR-amplified cDNA.

C-terminus of the Gag protein, while ZAM does not, but instead has a region very rich in asparagine. Identical results have been observed for the *woot* retrotransposon of the floor beetle *Tribolium castaneum* (Beeman *et al.*, 1996).

The *pol*-like ORF of ZAM is the most similar to other retrovirus-like elements, and includes regions identified with the retroviral enzymes protease (prt), reverse transcriptase (rt), RNase H (rnH) and integrase (int) (for a review, see Varmus and Brown, 1989). These domains contain the same or conserved amino acids at sites that have been shown to be invariant among many retroviruses and retroviral elements (Xiong and Eickbush, 1990). In addition, the organization of these different domains is prt–rt–rnH–int, which is the order found in retroviruses and the *gypsy-Ty3* group of elements and which differs from the prt–int–rt–rnH order found in the *copia-Ty1* group of retrotransposons (Finnegan, 1994).

ZAM possesses an additional ORF similar in position and structure to the retroviral *env* gene responsible for the infectious properties of a virus. The envelope glycoprotein of active retroviruses is composed of two polypeptides:

an external, glycosylated hydrophilic polypeptide (SU) and a membrane-spanning protein (TM). They are synthesized in the form of a polyprotein precursor that is proteolytically cleaved during its transport to the cell surface (Hunter and Swanstrom, 1990). The predicted protein encoded by ZAM ORF3 displays the SU and TM domains, strongly suggesting that ZAM is capable of extracellular transmission. In addition, the Env polypeptide of ZAM has potential *N*-glycosylation sites and cysteine residues that may be involved in protein folding. It carries a signal peptide for targeting to the endoplasmic reticulum in its NH₂ end and a hydrophobic domain of the size expected for a TM domain near its COOH end.

Nucleic acid and protein comparisons have shown that ZAM is very close to the lepidopteran *TED* retroelement. Despite their residence in invertebrates that are considered relatively distant phylogenetically (orders Diptera and Lepidoptera, respectively), the marked similarities between ZAM and *TED* support the contention that these elements could derive from a common ancestral transposon. On the other hand, available data cannot rule out the intriguing possibility that the observed similarity between these

MENTLLNLLL VLLSCHGAYQ SIFIHNF[▼]ST NLLAKVPVGK
 TLVIGNYKKI ^{SP}SHIIDLSEYT NQIEKLYHTI DTLRQDETTLT
 DSISILNAKL AQTQSKIDAL TPF^{SR}HKRG^L INGLGSLVKV
 VTGNMDANDA KNIETEINHL KSQSTTISDN FEIQNSFNDE
 VQLRFK^NLTR HINNEQNLIK NFFENTQNTI YTKIYNNEEE
 IKKLQYINRL NYNIDLLVSH LSDIESTLL AKINVIPKLI
 LDKTEITKIK QIFKTON^YTI KSEQHIYNLL KMNALNYQNK
 IIFSIKIPIF LSCNYEMARL IPLP^NISTQF VIAPKYLIYN
^NKNSMFMSTM YKCPVIEEQF VCEIDSINNL K^NNTCLGHLI
 QNKTSYCDIK ETGLTTDVFE PEKGFILVFN GNNLP^{IISS}^N
 QTITSINGSA IIKYN^NCTLK INEINYDNRA VSTEEHPDFF
 LPPMRKLIK^N ATINILT^LER LHLDTLTTSN KLLVVAAGNS
 RHST^{TL}YLIF TVSLVAVILT WTLRRDTHIF HTGPDHILPI
 VAPPIPPSMA FAPNWGGRSY RPIGTIHHPS L
 TM

Fig. 7. Predicted amino acid sequence of the spliced ORF3 product. Thick underlines indicate the location of the signal peptide (SP) of the precursor protein and the transmembrane domain (TM). The asparagine residues (N) in putative *N*-glycosylation sites conforming to the consensus sequence N-X-S/T are boxed. A putative protease cleavage site and cysteine residues are thickly underlined. The black arrowhead indicates the possible cleavage site that eliminates the signal peptide.

elements is due to a more recent horizontal transfer between different organisms of an ancestral element that has diverged into the present day elements.

ZAM RNA expression is reminiscent of the pattern seen in active retroviruses

Reverse transcription is a central process in the life cycle of infectious retroviruses and results in the synthesis of proviral DNA copies from the viral genomic RNA. *ZAM* transcription gives rise to two kinds of RNA: a full-length 8.6 kb transcript and a 1.7 kb subgenomic *env* transcript. The 8.6 kb transcript has all the features of a full-length RNA that will be reverse transcribed before its insertion into the genome. One unusual feature is that the same sequence seems to be used as a TATA box and as a polyadenylation signal in the *ZAM* LTRs. To our knowledge, this is the first case of a LTR retroelement in which a single LTR signal assures two functions: the initiation and the termination of transcription of *ZAM*.

The predicted R region that is duplicated at the ends of the RNA is 16 nucleotides long. An even smaller R region of 13 nucleotides has been identified in the Mason–Pfizer monkey virus (Sonigo *et al.*, 1986) and appears sufficient for template switching during reverse transcription.

The main functional difference between a retrovirus and a retrotransposon is that the former has an extracellular existence as infectious virus particles whereas the latter does not. Components of the viral envelope are translated from subgenomic RNAs from which *gag* and *pol* coding sequences have been removed specifically by splicing events. The 1.7 kb transcript identified on Northern blots

corresponds to this subgenomic RNA precursor of the *Env* polypeptide; it contains 203 nucleotides from the 5' end of the *ZAM* transcript spliced in-frame with the *env* gene. The complete translation product of this RNA, starting with the methionine codon found at position 494, has an *M_r* of 63 kDa. While this *Env* protein could enable *ZAM* to leave the host cell as a retrovirus, its infectious potential has yet to be demonstrated.

ZAM is mobilized in the germline of *Drosophila* strains via an RNA intermediate

In the course of our *ZAM* expression study, we have detected a *ZAM env* RNA-derived sequence present in the genome of the *Charolles* strain. The structure of this intronless element and the presence of a complete LTR at its 3' end suggest that it is of proviral origin. The following scheme for its formation can be put forward: a spliced *ZAM env* RNA was reverse transcribed to form a cDNA copy of the transcript. The initiation of its reverse transcription occurred at the *ZAM* tRNA PBS, as observed for transcription of all retroviruses. A 'jump' of the reverse transcriptase led to the regeneration of complete LTRs just as for the full-length viral RNA (Varmus and Brown, 1989).

In retroviruses or LTR-containing retrotransposons, reverse transcription of the RNA into double-stranded DNA takes place in ribonucleoprotein particles (Eichinger and Boeke, 1988). Signals exist on RNAs of a given retroviral family specifically to package viral RNAs in the presence of many kinds of cellular RNAs (Mesnard and Carrière, 1995). Thus, it is very likely that such an encapsidation signal, referred to as a psi (ψ) domain, is present in the *ZAM* retrogene and is sufficient for its encapsidation by *ZAM* gene products. In avian sarcoma leukosis virus (ASLV), this ψ region has been localized to 270 nucleotides of the 5' leader region (Aronoff *et al.*, 1993). If the ψ sequence of *ZAM* is in a similar location, between the PBS and the splice junction, that region is only 58 nucleotides. Recent analyses of the ψ region of HIV-1 have shown that it is a multipartite element composed of sequences localized to the 5' UTR region plus the *gag* gene (McBride and Panganiban, 1996). If the complete *ZAM* ψ region is composed of different subdomains located in and downstream of the leader region, it may be that the 58 nucleotide untranslated region portion would be less efficient than the full-length ψ domain and result in a less efficient packaging of the *ZAM* ORF3 RNA. This could explain the very rare occurrence of spliced *ZAM* genomic sequences in *D. melanogaster*.

Finally, the fact that the retrogene has been found integrated in the genome of *Charolles* indicates that its formation occurred in the germ line of the strain. This last point, added to the fact that *ZAM* has been identified in the genome of the *w^{IR6Rev}* strain after its insertion upstream of *white*, provides evidence that *ZAM* mobilization is at least taking place in the germ line of HCN strains of *D. melanogaster*.

ZAM expression is regulated differently depending on the *Drosophila* strains

The distribution analysis of *ZAM* in its host genome has shown that its copy number is highly variable among *D. melanogaster* strains. Two strains (*Charolles* and *w^{IR6Rev}* strains) have relatively high copy number and display

Fig. 8. Envelope retrogene formation. (A) The autoradiograph of the blot is presented on the left and a simplified structure of ZAM and its *env* mRNA are presented on the right. Long template PCR was performed on genomic DNA of low (*Canton S* and *wIR6*) and high (*wIR6Rev* and *Charolles*) copy number strains using the 5U and E1 oligonucleotides. The two pathways of formation of the *Charolles* spliced deleted envelope genomic DNA element are presented in (B) (retrotransposon formation) and (C) (pseudogene formation). (B1) The ZAM retrogene structure identified in the *Charolles* strain. Open and striped boxes represent complete and incomplete LTRs respectively. ASO (antisense splice oligonucleotide) and SDO (sense deletion oligonucleotide) are the specific primers used for inverse PCR experiment. ASO and SDO are located respectively on the splice junction (SJ) and on the deletion site (triangle). SD and SA are respectively the donor splice site and the acceptor splice site. H symbolizes the *HhaI* restriction endonuclease sites used for inverse PCR.

some 15 euchromatic insertions of ZAM. All other strains had low copy number and showed very few or no signals on the chromosomal arms. All the strains have some copies of ZAM located in the heterochromatic part of the genome. Interestingly, *wIR6* and *wIR6Rev* strains, which derived from one another, belong to two different groups, i.e. LCN and HCN strains respectively.

The question we can address is then: what is the molecular basis of the mechanism that induced or allowed

ZAM mobilization leading to a massive invasion of the *wIR6* genome in producing the *wIR6Rev* derivative? In the cases of the *P* element and *I* factor families of *D.melanogaster*, elements are mobilized in crosses between strains that have and that lack the elements (Engels, 1989; Bucheton, 1990). This type of mobilization has not been observed for ZAM (unpublished results).

ZAM distribution is similar to the gypsy retrovirus distribution in *D.melanogaster*. All the *D.melanogaster*

strains contain *gypsy* elements located in pericentromeric heterochromatin, and a few strains contain additional copies of *gypsy* located on chromosomal arms (Bucheton, 1995). A high copy number of *gypsy* on chromosomal arms is correlated with high rates of mobilization. Prud'homme *et al.* (1995) have shown that the stability of *gypsy* is associated with a gene located on the X chromosome called *flamenco* (*flam*) which controls its mobility. The *flam*¹ allele is permissive for *gypsy* transposition and the *flam*⁺ allele is not.

The mobilization of *ZAM* could be controlled by a similar process. At present, we know that *flam* is not involved in *ZAM* mobilization for two reasons. First, *w^{IR6Rev}* carries a wild-type allele of *flam* (A. Péliçon, personal communication). Second, *flam* has been described to act on *gypsy* mobilization by controlling the synthesis of its ORF3 *env* transcripts. *ZAM* control is certainly performed by a different process since its ORF3 *env* transcripts are synthesized in all the strains, albeit in lower amounts in the LCN strains. In contrast, it is the amount of full-length *ZAM* RNA that differs most between HCN and LCN strains. Thus, the question remains of what was the molecular basis of *ZAM* invasion in the *w^{IR6Rev}* strain? The primary P-mediated element mutagenesis applied on the *w^{IR6}* strain could have been responsible for *ZAM* amplification. We can postulate that in the fly that gave rise to the *Su78* line, a P element has inserted a key locus involved in the control of *ZAM* expression or life cycle. This event would have allowed *ZAM* mobilization, the increase of its copy number and thus the probability of its insertion upstream of *white*. It will be interesting to study the P insertions and their affected loci visualized in this strain in order to know if one of them may be responsible for *ZAM* deregulation. At present, it is tempting to think that it will allow the identification of a novel system offered by an invertebrate genome to control the relationships between retroviruses and their hosts.

Materials and methods

Fly stocks

Fly stocks were maintained on cornmeal–glucose–yeast media at 20°C. The *Charolles*, *SV-XX* (*wv¹⁴⁴*, *C(1)DX*, *yf*), *w^{IR6}* and *Canton S* strains are from the collection of the INSERM U384. The *23500* strain is from the Umea Stock Center.

Drosophila DNA preparation and Southern blotting

Genomic DNA was prepared from strains of *Drosophila* according to the protocol of Udomkit *et al.* (1995). The DNA was then transferred to Hybond N⁺ membrane by capillary action in 3.6 M NaCl, 0.2 M Na phosphate, 0.02 M EDTA pH 7.7. After hybridization at 42°C, filters were washed in 2× SSC, 0.1% SDS at 42°C, and in 0.1× SSC, 0.1% SDS at 42°C. Fragments used as probe were gel purified and labelled with [α -³²P]dCTP by random priming (Stratagene).

PCR amplifications

PCR amplification was performed with the Expand™ Long Template PCR System of Boehringer. Conditions and procedures are those described in the manufacturer's protocol. Primers o1 [(*Pst*I site) 5' GTT GTC CCC TGC AGT AAA TGC 3' (6510; 6530)] and o2 [(*Hpa*I site) 5' GAG CCA GTT AAC TGG CAT TCC 3' (7311; 7291)] were used to amplify the complete *ZAM* element at the *w^{IR6Rev}* locus; 5U [5' CAG CCG GAA AAC TGG AAT GGA 3' (478; 498)] and p2 [5' GGG TTG TAG AAT ATG TCG CGA 3' (3658; 3678)] amplified the P3 fragment corresponding to the *ZAM* 5' region (Figure 2).

DNA sequencing

The sequence of the *ZAM* element was determined by subcloning the *Pst*I–*Hpa*I PCR product in the SK pBluescript vector (Stratagene). Double-stranded templates were prepared using QUIAGEN™ columns and these were sequenced on both strands by dideoxynucleotide chain termination using the DNA Dye Terminator Cycle Sequencing Kit (Perkin Elmer); the samples were loaded on an ABI377 sequence analyser. Two or more sequences of the same fragment cloned from independent amplifications were determined in order to eliminate potential errors induced during PCR.

RNA extraction and poly(A)⁺ isolation

RNAs of 0–24 h adult flies were extracted using a guanidine-HCl method (Cox, 1968). After treatment for 10 min at 65°C, total RNA was loaded on an oligo(dT)–cellulose column, and poly(A)⁺ RNA was isolated (Aviv and Leder, 1972).

Northern blot experiments

The antisense RNA probe BH was synthesized from the *ZAM* ORF3 *Hind*III–*Bgl*II fragment (Figure 2) cloned in the pBluescriptII vector using procedures presented by Maniatis *et al.* (1989). A DNA probe from the *Actin-5C* gene was used as a control for RNA loading. Northern blotting was performed according to the protocol used by Lajoinie *et al.* (1995).

Reverse transcription–PCR analysis

Poly(A)⁺ RNA (500 ng) from adult flies was reverse transcribed into single-stranded cDNA using the first strand cDNA synthesis kit of BRL following the protocol provided by the manufacturer. Two µl of the cDNA pool was used for PCR amplification with the oligonucleotide 5U and the following primers: E1, 5' TGG TGT ATG GTA CCG ATG GGT 3' (7971; 7991); E2, 5' TGT GAG TGT ATC CAG GTG 3' (7732; 7749); E3, 5' ATG TCG CAG TAG CTG GTC 3' (7437; 7454); E4, 5' CAG GTT GCC GGT GAC GAC 3' (6706; 6723); and E5, 5' CAG GTT GCC GGT GAC GAC 3' (6706; 6723).

Determination of *ZAM* transcript 5' and 3' ends

Primer extension. For mapping of the 5' end of RNAs initiated from the 5' *ZAM* LTR, the reverse complement 5U^{rc} primer [5' TCC ATT CCA GTT TTC CGG CTG 3' (478; 498)] was radiolabelled at its 5' end with T4 polynucleotide kinase and [γ -³²P]ATP. The end-labelled primer was annealed to *w^{IR6Rev}* total RNA and extended with avian myeloblastosis virus (AMV) reverse transcriptase. Product analysis was performed according to Triezenberg (1992). Only one band corresponding to the extension product has been observed after autoradiography, indicating that transcription starts at the same position for both transcripts.

RT–PCR. First strand cDNA synthesis was performed using the first strand cDNA synthesis kit of BRL on total *w^{IR6Rev}* RNA. A modified oligo(dT) primer containing a 5' primer adaptor [5' GAC TCG AGT CGA CAT CGA (dT)₁₇ 3'] was used for the reverse transcription step. PCR amplification was performed using the primer adaptor (5' GAC TCG AGT CGA CAT CG 3') and a *ZAM* envelope-specific oligonucleotide [5' TCA ACA GAA GAG CAC CC 3' (7642; 7658)].

Inverse PCR experiment

Charolles DNA extraction was performed from one fly according to Gloor and Engels (1992). DNA was treated with *Hha*I restriction endonuclease and the 3' end of the *ZAM* retrogene was identified by the inverse PCR method (Gloor *et al.*, 1983). ASO (5' CCC CAT GGC AAG ATA ATA GAA C 3') and SDO [5' GTA CTT AAT ATA TAA AGG AAA CGG G 3'] were the primers used in the PCR experiment.

In situ hybridization on polytene chromosomes:

In situ hybridization on polytene chromosomes of *D. melanogaster* strains was carried out following the method described in Biémont (1994).

Accession numbers

The accession number for *ZAM* is AJ000387.

Acknowledgements

We are particularly grateful to D. Carroll for critical review of the manuscript. We thank J.M. Deragon for his helpful comments, C. Biémont for assistance in polytene chromosome hybridization and M. and X. Wilhelm for their helpful information. This work was supported by

INSERM grants and by a grant from MESER (ACC-CV). P.L. received a graduate grant from MESER.

References

- Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403.
- Aronoff,R., Hajjar,A.M. and Linial,M.L. (1993) Avian retroviral RNA encapsidation: reexamination of functional 5' RNA sequences and the role of nucleocapsid Cys-His motifs. *J. Virol.*, **67**, 178–188.
- Aviv,H. and Leder,P. (1972) Purification of biologically active globin messenger RNA by chromatography on oligothymidine acid-cellulose. *Proc. Natl Acad. Sci. USA*, **69**, 1408–1412.
- Beeman,R.W., Thomson,M.S., Clark,J.M., DeCamillis,M.A., Brown,S.J. and Denell,R.E. (1996) *Woot*, an active Gypsy-class retrotransposon in flour beetle *Tribolium castaneum*, is associated with a recent mutation. *Genetics*, **143**, 417–426.
- Biémont,C. (1994) Dynamic equilibrium between insertion and excision of P elements in highly inbred lines from an M' strain of *Drosophila melanogaster*. *J. Mol. Evol.*, **39**, 466–472.
- Bucheton,A. (1990) I transposable elements and I-R hybrid dysgenesis in *Drosophila*. *Trends Genet.*, **6**, 16–21.
- Bucheton,A. (1995) The relationship between the *flamenco* gene and gypsy in *Drosophila*: how to tame a retrovirus. *Trends Genet.*, **11**, 349–353.
- Coffin,J.M. (1990) Retroviridae and their replication. In Fields,B.N., Knipe,D.M., Chanock,R.M., Hirsh,M.S., Melnick,J.L., Monath,T.P. and Roizman,B. (eds), *Fields Virology*. 2nd edn. Raven Press, New York, pp. 1437–1500.
- Covey,S.N. (1986) Amino acid sequence homology in Gag region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. *Nucleic Acids Res.*, **14**, 623–633.
- Cox,R.A. (1968) The use of guanidium chloride in the isolation of nucleic acids. *Methods Enzymol.*, **12**, 120–129.
- Cribbs,D.L., Gillam,I.C. and Tener,G.M. (1987) Nucleotide sequences of three tRNA^{ser} from *Drosophila melanogaster* reading the six serine codons. *J. Mol. Biol.*, **197**, 389–395.
- Dessen,P., Fondrat,C., Valencien,C. and Mugnier,C. (1990) BISANCE: a French service for access to biomolecular sequences data bases. *Comput. Applic. Biosci.*, **6**, 355–356.
- Eichinger,D.J. and Boeke,J.D. (1988) The DNA intermediate in yeast *TyI* element transposition copurifies with virus-like particles: cell-free *TyI* transposition. *Cell*, **54**, 955–966.
- Engels,W.R. (1989) P elements in *Drosophila melanogaster*. In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. ASM Publications, Washington, DC, pp. 437–484.
- Finnegan,D.J. (1994) Retroviruses and transposons. Wandering retroviruses? *Curr. Biol.*, **4**, 641–643.
- Friesen,P.D. and Nissen,V. (1990) Gene organization and transcription of *TED*, a Lepidopteran retrotransposon integrated within the baculovirus genome. *Mol. Cell. Biol.*, **10**, 3067–3077.
- Gloor,G.B. and Engels,W.R. (1992) Single fly DNA preps for PCR. *Dros. Inform. Serv.*, **71**, 148–149.
- Gloor,G.B., Preston,C.R., Johnson-Schlitz,D.M., Nassif,N.A., Phillis,R.W., Benz,W.K., Robertson,H.M. and Engels,W.R. (1983) Type I repressors of P element mobility. *Genetics*, **135**, 81–95.
- Goodchild,N.L., Freeman,J.D. and Mager,D.L. (1995) Spliced HERV-H endogenous retroviral sequences in human genomic DNA: evidence for amplification via retrotransposition. *Virology*, **206**, 164–173.
- Hunter,E. and Swanstrom,R. (1990) Retrovirus envelope glycoproteins. *Curr. Top. Microbiol. Immunol.*, **157**, 187–253.
- Inouye,S., Yuki,S. and Saigo,K. (1986) Complete nucleotide sequence and genome organization of a *Drosophila melanogaster* transposable genetic element, 297. *Eur. J. Biochem.*, **154**, 417–425.
- Kim,A., Terzian,C., Santamaria,P., Péllisson,A., Prud'homme,N. and Bucheton,A. (1994) Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA*, **91**, 1285–1289.
- Klenk,H.D. and Garten,W. (1994) Host cell proteases controlling virus pathogenicity. *Trends Microbiol.*, **2**, 39–43.
- Lajoinie,O., Drake,M.E., Dastugue,B. and Vaury,C. (1995) Aberrant pre-mRNA maturation is caused by LINE insertions into introns of the *white* gene of *Drosophila melanogaster*. *Nucleic Acids Res.*, **23**, 4015–4022.
- Maniatis,T., Fritsch,E.F. and Sambrook,J. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Marck,C. (1988) DNA Strider: a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family computers. *Nucleic Acids Res.*, **16**, 1829–1836.
- 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.
- McBride,M.S. and Panganiban,A.T. (1996) The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J. Virol.*, **70**, 2963–2973.
- Mesnard,J.M. and Carrière,C. (1995) Comparison of packaging strategy in retroviruses and pararetroviruses. *Virology*, **213**, 1–6.
- Morrow,C.D., Park,J. and Wakefield,J.K. (1994) Viral gene products and replication of the human immunodeficiency type 1 virus. *Am. J. Physiol.*, **266**, 1135–1156.
- O'Hare,K., Murphy,C., Levis,R. and Rubin,G.M. (1984) DNA sequence of the *white* locus of *Drosophila melanogaster*. *J. Mol. Biol.*, **180**, 437–455.
- Péllisson,A., Song,S.U., Prud'homme,N., Smith,P.A., Bucheton,A. and Corces,V.G. (1994) Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila flamenco* gene. *EMBO J.*, **13**, 4401–4411.
- Polard,P. and Chandler,M. (1995) Bacterial transposases and retroviral integrases. *Mol. Microbiol.*, **15**, 13–23.
- Prud'homme,N., Gans,M., Masson,M., Terzian,C. and Bucheton,A. (1995) *Flamenco*, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. *Genetics*, **139**, 697–711.
- Rawlings,N.D. and Barrett,A.J. (1995) Retroviral proteases: families of aspartic peptidases, and those of unknown catalytic mechanism. *Methods Enzymol.*, **248**, 105–119.
- Robertson,H.M., Preston,C.R., Phillis,R.W., Johnson-Schlitz,D.M., Benz,W.K. and Engels,W.R. (1988) A stable genomic source of P-element transposase in *Drosophila melanogaster*. *Genetics*, **118**, 461–470.
- Saigo,K., Kugimiya,W., Matsuo,Y., Inouye,S., Yoshioka,K. and Yuki,S. (1984) Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable element in *Drosophila melanogaster*. *Nature*, **312**, 659–661.
- Shinnick,T.M., Lerner,R.A. and Sutcliffe,J.G. (1981) Nucleotide sequence of Moloney murine leukaemia virus. *Nature*, **293**, 543–548.
- Song,S.U., Gerasimova,T., Kurkulos,M., Boeke,J.D. and Corces,V.G. (1994) An Env-like protein encoded by a *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus. *Genes Dev.*, **8**, 2046–2057.
- Sonigo,P., Barker,C., Hunter,E. and Wain-Hobson,S. (1986) Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. *Cell*, **45**, 375–385.
- Tanda,S., Shrimpton,A.E., Chuch,L.L., Itayama,H., Matsubayashi,M., Saigo,K., Tobari,Y.N. and Langley,C.H. (1988) Retrovirus like features and site specific insertions of a transposable element, *tom*, in *Drosophila ananassae*. *Mol. Gen. Genet.*, **214**, 405–411.
- Tchenio,T., Segal-Bendirdjian,E. and Heidmann,T. (1993) Generation of processed pseudogenes in murine cells. *EMBO J.*, **12**, 1487–1497.
- Triebenberg,S.J. (1992) Primer extension protocol. In *Current Protocols in Molecular Biology*. J.Wiley and Sons, Inc., Vol. 1, pp. 4.8.1–4.8.5.
- Udomkit,A., Forbes,S., Dalgleish,G. and Finnegan,D.J. (1995) *BS'* a novel LINE-like element in *Drosophila melanogaster*. *Nucleic Acids Res.*, **23**, 1354–1358.
- Varmus,H. and Brown,P. (1989) Retroviruses. In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. ASM Publications, Washington, DC, pp. 53–108.
- Von Heijne,G. (1983) Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.*, **133**, 17–21.
- Xiong,Y. and Eickbush,T.H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.*, **9**, 3353–3362.
- Yoshioka,K., Kanda,H., Akiba,H., Enoki,M. and Shiba,T. (1991) Identification of an unusual structure in the *Drosophila melanogaster* transposable element copia: evidence for copia transposition through an RNA intermediate. *Genes*, **103**, 179–184.
- Yuki,S., Ishimaru,S., Inouye,S. and Kaoru,S. (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**, 3017–3030.

Received on July 29, 1997; revised on September 19, 1997