Nucleic Acids Research, 1998, Vol. 26, No. 17 **4019–4024**

A hotspot for the *Drosophila gypsy* retroelement in the *ovo* locus

Kimberley J. Dej, Tatiana Gerasimova, Victor G. Corces and Jef D. Boeke[1,*]

Department of Biology, Johns Hopkins University and [1]Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

Received April 24, 1998; Revised and Accepted July 22, 1998

ABSTRACT

The *Drosophila* retroelement *gypsy* has a number of unusual features including an unusual LTR terminal sequence and an apparent target sequence preference. The *ovo* locus is a known hotspot for *gypsy* insertion. We examined the target sequence preference of *gypsy* within *ovo* by isolating 26 new insertions and sequencing the *gypsy/ovo* junctions. Insertions were found at multiple sites within the *ovo* locus. The insertions clustered within an ~150 bp region in the non-translated region of the *ovo*β transcript, with most insertions falling within the first intron. There were seven sites of insertion within this region and these mostly conform to the consensus sequence YRYRYR (where Y = pyrimidine and R = purine). However, this target sequence is at best necessary but not sufficient to specify a hotspot, as there were several other sequences conforming to this consensus in the *ovo* locus that were not hit. The results indicate that *gypsy* may have a higher degree of target specificity than most infectious LTR retroelements.

INTRODUCTION

The *gypsy* LTR retroelement of *Drosophila melanogaster* is of great interest because it represents a genetically tractable retroelement that is closely related to the non-infectious LTR retrotransposons but is infectious like retroviruses (1,2). Like retroviruses, *gypsy* encodes genes equivalent to *gag*, *pol* and *env*, whereas most retrotransposons encode only *gag* and *pol*. Moreover, the envelope protein is encoded by a subgenomic mRNA, just as is the case in retroviruses (3). The latter feature is shared with another *Drosophila* retroelement, *Tom* (4). Although this aspect of *gypsy*'s biology has attracted considerable attention, relatively little is known about how *gypsy*, like all retrotransposons and retroviruses, inserts its DNA into the host genome during the essential integration phase of its life cycle. Like these other elements, *gypsy* encodes an integrase (IN) protein that presumably mediates the integration reaction. Unlike the other elements, *gypsy* has unusual terminal sequences, and may have sequence specificity for insertion.

LTR retroelement life cycles can be divided into a few discrete stages, reviewed in (5). These are: (i) the gene expression stage, when element RNA(s) and proteins are produced, (ii) the assembly stage, when virion or virus-like particles are put together and matured via a series of endoproteolytic cleavages, (iii) the reverse transcription stage, when element DNA is produced from the genomic RNA in the particle and (iv) the integration stage, during which the newly synthesized element DNA is integrated at a novel site in the host chromosome. An absolutely critical aspect of the integration stage is the identification of target sequences in host DNA. Craigie (6) pointed out that non-infectious elements are bound to the genome of the host organism they inhabit, and the progeny thereof. Hence non-infectious elements are likely to have evolved mechanisms for identifying 'safe havens' in the host DNA where they do little if any damage to the host. In contrast, the infectious retroviruses are relatively free of this constraint because of their ability to move horizontally to other members of the host population. Indeed, recent studies of both avian and mammalian retroviral integration suggest that only modest specificity of integration relative to host DNA targets is observed (7,8). In contrast, recent studies of the integration specificity of the non-infectious elements Ty1, Ty3 and Ty5 of yeast have borne this prediction out rather spectacularly. Ty1 and Ty3 both target tRNA upstream regions, which in their host, *Saccharomyces cerevisiae*, represent gene-free 'safe havens' (9–12). Interestingly Ty1 and Ty3, which are extremely distantly related elements (13), target these regions by what appear to be distinct mechanisms (14), and so tRNA targeting appears to be an example of convergent evolution for these two elements. Ty5 on the other hand, targets regions of silent chromatin, a different type of 'safe haven' (15,16). Because the *gypsy* element lies at the phylogenetic interface between the retrotransposons and retroviruses, analysis of its integration specificity is of special interest.

The *gypsy* element is unusual among the LTR retroelements in another way. All retroviruses and nearly all LTR retrotransposons share the terminal sequences TG...CA (usually embedded within a somewhat longer imperfect, inverted repeat sequence; Fig. 1). However *gypsy* is one of a very small number of elements, thus far known only from insects, that have non-TG...CA terminal sequences. Interestingly, and perhaps not coincidentally, all of these elements share the property of having three reading frames corresponding to retroviral *gag*, *pol* and *env*. All of these elements

*To whom correspondence should be addressed at: Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, 617 Hunterian, 725 N. Wolfe Street, Baltimore, MD 21205, USA. Tel: +1 410 955 2481; Fax: +1 410 614 2987; Email: jef.boeke@qmail.bs.jhu.edu

Element	U3 end	U5 end	# "core" ORFs
gypsy/Ty3 family			
gypsy	AGTTAACAACTAACAATGTA....CTATTGGAACTTATATAATT		3
gypsyDv	AGTTAACAACTAAGCATAAA....ACATACATTAATGAATAACT		3
gypsyDs	AGTTAAGAACTAAGTACATA....GGGACCGGATTGGATTAATT		3
17.6	AGTGACATATTCACATACAA....GCAACCATTTATTTGCAATT		3
297	AGTGACGTATTTGGGTGGTC....AATAAAACAACAATTTTACT		3
Tom	AGTGACATATTCACTCTCAC....AAAATATAAACTATTTTACT		3
TED	TGTTAGGTATGGAGCCTTAA....CGTCTGCAGTATACGTAATT		3
Ty3	TGTTGTATCTCAAAATGAGA....CTCGAGCCCGTAATACAACA		2
Tf1	TGTCAGCAATACTACACTAC....ACCAAACTGCGTAGCTAACA		1
412	TGTAGTATGTGCCTATGCAA....AATTGGAGTTCATCATTACA		2
1731	TGTTGAATATAGGCAATGCC....CACCTTTTCAACAAACCACA		2
del	TGTGAAACCCTGAATTTTCG....TAAAATGTCGGGTCGTGAAA		
copia/Ty1 family			
Ty1	TGTTGGAATAGAAATCAACT....CAACATTCACCCAATTCTCA		2
copia	TGTTGGAATATACTATTCAA....ATAAATTATAAATTACAACA		1
Tnt1	TGATGATGTCCATCTCATTG....TAAGGGGTTTATTCCCAACA		1
Ty5	TGTTGAATGTGATAACCCAA....CGTAATTCACTACGTCAACA		1
Retrovirus family			
HIV-1	TGGAAGGGCTAATTTGGTCC....AGTGTGGAAAATCTCTAGCA		3
MoMLV	TGAAAGACCCCACCTGTAGG....CCGTCAGCGGGGGTCTTTCA		3
RSV	TGTAGTCTTATGCAATACTC....GAATGAAGCAGAAGGCTTCA		3
Spuma	TGTGGCAGACAGCCACTAAA....ATACACAATATTCCACGACA		3

Figure 1. Retroelement (LTR) termini. *gypsy* and related insect elements with three open reading frames corresponding to *gag*, *pol* and *env* have unusual termini relative to other retrotransposons and retroviruses, which universally begin and end with the dinucleotide inverted repeat TG ... CA. Interestingly, *gypsy* sequences from different *Drosophila* species do not completely conserve the terminal dinucleotide, suggesting that *gypsy*'s IN may have less stringent terminal sequence requirements than those of the other LTR retroelements. Alternatively, the INs of different *gypsy* species could recognize the individual termini with high specificity.

are presumed to be infectious, although this has only been demonstrated, thus far, for *gypsy*.

The *gypsy* element can insert into many different loci in genomic DNA. Based on the sequencing of a relatively small number of genomic *gypsy* elements it was concluded that *gypsy* was highly sequence-specific in its insertion into genomic DNA; in nearly every example, the target sequence TATATA or TACATA (17–19) was duplicated upon insertion. A recent study in which three *gypsy* insertions into the *ovo* gene were sequenced (2) suggested that other target sequences could also be recognized, and suggested a relaxation of the proposed recognition sequence to YRYRYR. The *gypsy* element moves at high frequency during crosses of *flamenco* strains to other strains (20). A particularly useful genetic assay for *gypsy* movement was developed by Mével-Ninio *et al*. (21), who found that dominant alleles of the X-linked *ovo* locus, such as ovo^{D1}, could be reverted by retrotransposon insertion. Our previous studies indicate that the *ovo* region represents a major hotspot for *gypsy* insertion, with as many as 1/4 of all insertions in the *Drosophila* genome occurring within this region (2). The *ovo* locus, required for differentiation of the female germ line, codes for two OVO protein isoforms that differ in their N-termini (22,23). Dominant female-sterile alleles, including ovo^{D1}, are the result of point mutations that create new initiation codons in the 5′ region of ovo (24,25). We have used the ovo^{D1} reversion assay system to generate a larger collection of *gypsy* insertions, and report here the insertion sites of a total of 26 new insertion mutations within the *ovo* locus. Remarkably, almost all of the insertions recovered cluster within a region of ~150 bp and most of these fall within the first intron of the major *ovo* transcripts.

MATERIALS AND METHODS

Isolation of ovo^{D1} revertants

Ovo^{D1} revertants were isolated by two methods, described in Figure 2. The two methods are genetic crosses (Fig. 2A) and virus particle 'feeding' (Fig. 2B). In both the feeding experiments and the genetic crosses, the assay of ovo^{D1} reversion, originally described by Mével-Ninio *et al*. (21) and subsequently improved by identification of the *flamenco* gene as an essential element responsible for high level *gypsy* mobility (3,20), was used. The genetic crosses involved three steps. First, females of one of three strains, listed in Table 1, were crossed by *y v f mal* $flam^1$ males (F_0). Second, progeny females were crossed to ovo^{D1} *v* males (F_1); the F_2 female progeny were then scored for fertility (ovo^{D1} reversion). The high frequency of reversion observed (Table 1) indicates that the three strains from which the females were derived must bear *flamenco* mutations. In the feeding experiments, *gypsy* particles were first isolated from females of the *y v f mal flam* strain; these were fed to SS strain larvae as described previously (2).

PCR amplification

Genomic DNA was prepared from 50 flies for each insertion stock by using a potassium acetate quickprep. Flies were homogenized in 500 µl of extraction buffer (0.1 M Tris–HCl, pH 9.0; 0.1 M EDTA; 1% SDS) using a glass mortar and pestle and the homogenate was incubated in a 1.5 ml microcentrifuge tube for 30 min at 70°C. Potassium acetate was added to a final concentration of ~0.1 M and the mixture was placed on ice for 30 min. This was spun down at 14 000 r.p.m. in an Eppendorf 5415C microfuge for 15 min at 4°C. The supernatant was transferred to a fresh 1.5 ml tube and the DNA was precipitated by adding 0.5× vol of isopropanol.

Approximately 50 ng of genomic DNA was used per PCR reaction. The PCR protocol was as follows: 91°C, 1 min; 58–62°C, 1 min; 72°C, 2–10 min; number of cycles, 32. Generally each sample was run for a short extension time (2–3 min) at 72°C with *Taq* polymerase and a long extension time (5–10 min) with *Taq* polymerase with *Taq* extender in order to look for short and long extension products. *Taq* polymerase and *Taq* polymerase extender buffers were used accordingly with 200 µM dNTPs and 10 ng/µl of each primer.

Primers used:
P1 5′-CAACATGACCGAGGACGGTCATAAAC-3′ (*gypsy* 3′)
P2 5′-CTCCCGCTCTGCGGGCTTCTCTTT-3′ (*ovo* 5′)
P3 5′-CTTTGCCGAAAATATGCAATG-3′ (*gypsy* 5′)
P4 5′-CGGCTTTTTCAGCGGCTAACC-3′ (*ovo* 5′)
POA 5′-TCGCCATCTCGCTCTGTTG-3′ (*ovo* 5′)
POB 5′-CTCATTGCTCTACGCGTTCTG-3′ (*ovo* 3′)
POC 5′-GCTACCAACCGAACAAGCTGC-3′ (*ovo* 3′)
SPGL5′: 5′-GCGTGGAGCGTTGAACCC-3′
SPGL3′: 5′-TCAAGCCCTCCAACCTAA-5′

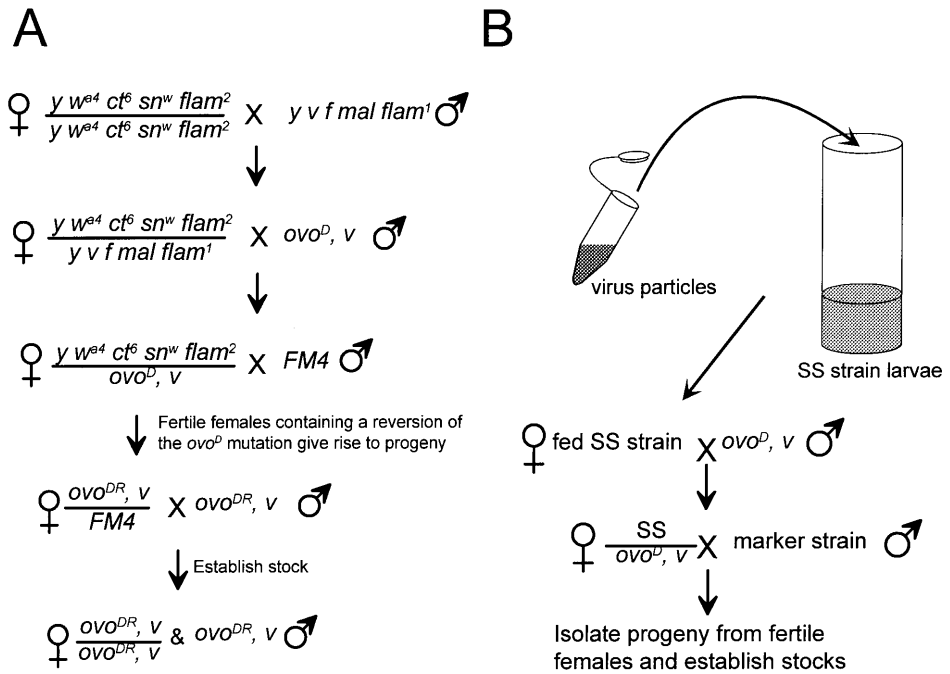

Figure 2. Isolation of *ovo*^{D1} revertants caused by *gypsy* insertions (diagram of two methods, crossing and feeding). (**A**) Method 1: genetic cross. A genetic cross between females of a donor strain (*y w*^{a4} *ct*^{6} *sn*^{w} *flam*^{2}) and *ovo*^{D1} males was carried out as outlined in the figure. Stocks of *ovo*^{DR} (revertant) lines were then generated. (**B**) Method 2: 'feeding' experiment. *gypsy* particles, isolated as described, were fed to the SS strain of flies. Fed SS females were mated with male *ovo*^{D1} flies and fertile female progeny were identified and studied further. PCR was performed on the DNA from the progeny of these *ovo*^{D1} (revertant) fertile females.

Table 1. Frequency of *ovo*^{D1} reversion

Crosses F_0 female strain	Total females tested	Fertile F_2 females	*ovo*^{D1} reversion (%)
ct^{MR2}	420	27	6.4
y w^{a4} *ct*^{6} *sn*^{w}	450	30	6.7
Df(1) Pgd-kz	397	12	3.1
Feeding			
NA	928	20	2.2

PCR products were run on 0.8% agarose gels and these gels were dried and hybridized directly with *ovo* genomic DNA. The 18 kb *ovo* clone D1B2NR was used as a probe (26). This clone covers the 7.2 kb *Sal*I–*Hin*dIII rescue fragment and 8 kb of upstream and 3 kb of downstream flanking DNA. PCR products that hybridized with the *ovo* DNA were then isolated from the gel by electroelution and sequenced directly.

The P4 and P2 primers and P4 and POB primers generated products of the expected length using genomic miniprep DNA as template (~1 and 6 kb respectively) showing that this protocol was capable of amplifying large products and thus of detecting any *gypsy* insertion within this region. Reactions were done using the *Taq* Extender PCR additive (Stratagene) which was used to generate up to an 8 kb product in control experiments using cloned DNA and vector primers.

DNA sequencing

PCR products were electroeluted from agarose gels and sequenced on an ABI Model 373 DNA sequencer using fluorescent dye terminator chemistry using the primers SPGL 5′ (reading outward from the 5′ LTR of *gypsy*) or SPGL 3′ (reading outward from the 3′ LTR of *gypsy*) as appropriate. When necessary, PCR products were directly cloned in the TA vector cloning system (InVitrogen, San Diego, CA) and the double-stranded plasmids were sequenced using the same method.

RESULTS AND DISCUSSION

Introduction of *gypsy* to *ovo*^{D1} strains and identification of insertions

Insertions of *gypsy* into *ovo*^{D1} were obtained in either of two ways, by crossing as described (21) or by feeding of *gypsy* virus particles as described by Song *et al.* (2). The two methods used in this study are specifically outlined in Figure 2; in all cases a high level of *ovo*^{D1} reversion activity was observed (Table 1), and *gypsy* insertions constituted 61% of the total *ovo*^{D1} reversion events observed using both methods. The remaining (non-*gypsy*) revertants were not characterized further. The progeny of revertant female flies were studied for whether or not they bore *gypsy* elements within the *ovo* locus by a series of PCR amplification experiments using one primer reading outward from one of the *gypsy* ends and one primer chosen from among a set of primers spanning the *ovo* locus (Fig. 3). Finally, selected

```
   1  GTCGACTGCAACAGTTGGCCATGACCAACAGAGCGAGATGGCGAATGCGTGATCGGCATTTGTAATTGCACACACTCCTGTAATTGCACAAACAACACAT   100

 101  GCATGCACATACAAATATACAAATGCAATTACACACTCGCAAAGGGCCAAGAAAAACACGCACACCCACGCCACAAAGCAAAAACATACATATATATACA   200
                                                                                                  site A

(600 bp noncoding sequence deleted for brevity)
                                                                 TSSβ
                                                                  ↓
 801  TTGAAGAATGTAAGGAAGAAAGGGAGTGTGATCGAAAGTCCGTTCCTTTTTACAGTTACATAGCAATCGTCCGAGCGAACGGACAGACAAATTTCTGAGA   900
                                                         site B

                      Major SD                Minor SD
 901  ATCGCACTTCTTTGCTTCTCTCATTTTCGGTGATTTTCCGTTGCTTTTTTATTGTGTGTGCACTCGAAAGTTCTATTAGGTTCCACAGGGTTTTTATACA   1000
                                                                      site C                site E     site F
                                                                        site D                           site G
                                                                                                           site H

                                                                                   SA
1001  TATGATTAATTCGTATTTAAAGTAACTTTGTGCACCTGATGCTAATTTAATTTTCTTCTCCTTTCTTGCAGAAAACTTTGCCGCCGAGTTGCTGCAGCGT   1100
(site H, ctd.)

(1000 bp of noncoding sequence deleted for brevity)

2101  GCAATCAGAACGCCGGCAGCGGCGCTGGTTCACCTGGATCTGGTGCCGGTGGCGGTGGTGGTATGGGCGGTGGTCGCGACGGCCGAGGGAACTACGGACC   2200
                                                                           M  G  G  G  R  D  G  R  G  N  Y  G  P
```

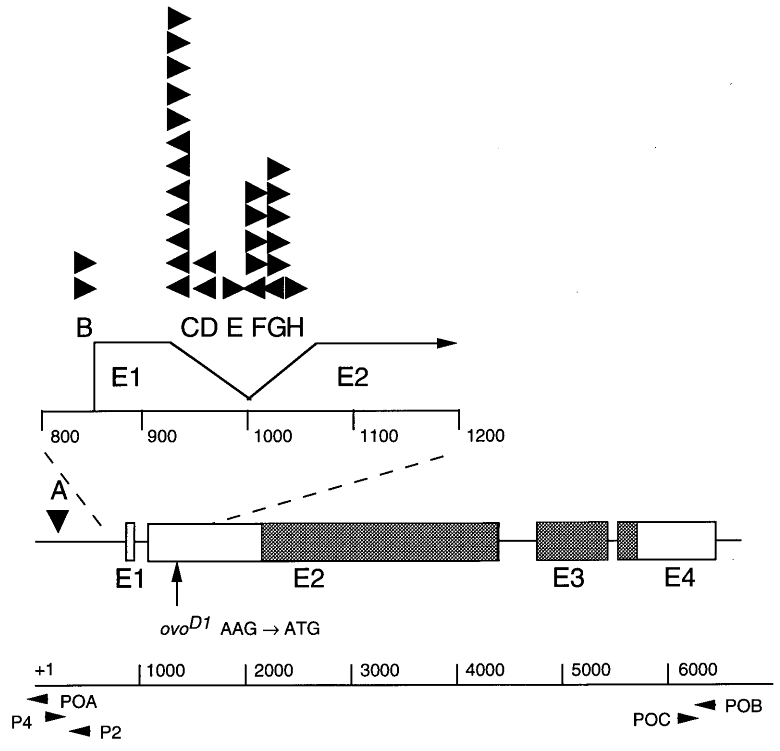

Figure 3. *gypsy* insertions in the *ovo* locus. The DNA sequence indicates the sites of insertion of *gypsy* in the *ovo* locus studied, as well as the features of the major ovo transcript found in the female germline, called variously the *ovo*β and *ovoB* transcripts by different authors. This is the transcript relevant for the *ovo*^{D1} phenotype because *ovo*^{D1} only affects female fertility; a second transcript, *ovo*α or *ovoA*, is expressed in both sexes and has a different 5′ exon. The features included are the major transcription start site (TSS; vertical arrow), major and minor splice donor sites (SD; overline) for the first intron (two other extremely minor splice donors are not shown), the splice acceptor site in intron 1 (SA; overline), the position of the new ATG formed by the *ovo*^{D1} mutation (24,25) (arrow pointing upward) and a single *gypsy* insertion (site A, underline). Also indicated are seven sites of *gypsy* insertion uncovered during this study (sites B–H); the inferred target site duplications are underlined. The diagram indicates the position of sites A–H within the *ovo* locus; insertions are symbolized by triangles, exons by boxes and the ORF by shading. Rightward triangles are *gypsy* elements in the (+) orientation, leftward in the (–) orientation; the orientation of the insertion in site A has not been reported and is represented by a vertical triangle. The arrows below the number line indicate the approximate positions of the primers used.

PCR products were chosen for DNA sequence analysis, and the DNA sequencing provided final proof that the observed products indeed corresponded to *gypsy* insertions into the *ovo* locus.

Sequencing the insertion points of 26 *gypsy* elements in the *ovo* locus

A total of 26 PCR products corresponding to *gypsy–ovo* junction fragments were directly sequenced (or cloned and then sequenced) and the sequences of the junction between the end of the *gypsy* element and *ovo* were tabulated (Fig. 3 and Table 2). Remarkably, all of the new insertion sites clustered within a 150 bp region, and seven different target sites, defined by their inferred 6 bp target site duplication, were observed; four of these target sites consisted of either a single insertion or insertions in a single orientation and the other three included insertions in both orientations. The (+) orientation (defined as the one in which *gypsy* and *ovo* are transcribed in the same direction) was represented by 62% of the insertions. All but one of the newly defined target sites fall within the first intron of the female germ line-specific *ovo*β transcripts of the *ovo* locus recently defined by Garfinkel *et al*. (23) and Mével-Ninio *et al*. (24). The remaining target site lies at the major *ovo*β transcription start site mapped in the same studies. All of the insertions lie within the boundaries of the other class of transcripts, the *ovo*α transcript. Interestingly, the *yct* allele contains two separate *gypsy* insertions ~40 bp apart, in opposite orientations. The insertion sites obtained by crossing and feeding methods were very similar to each other, and also agree well with the positions of previously restriction mapped insertions (1,21).

Table 2. Consensus sequence of target site duplication

Consensus	% of insertions conforming
TATATA	0
TACATA	6.7
YRYRYR	67
YRYRYR 5/6 match	23

It is expected that insertion of *gypsy* at these positions in *ovo* will lead to premature termination of transcription within the *gypsy* sequences. The *ovo* locus contains two transcription start sites at positions 361 and 852 bp (22). The transcript initiating at 361 bp codes for the OVO-A isoform, a 1222 amino acid protein believed to be expressed late in oogenesis (25). The second transcript codes for the OVO-B isoform, a smaller protein truncated at its N-terminus. The OVO-B isoform is believed to be expressed early in oogenesis (25). It has been suggested that the expression of OVO-A may downregulate OVO-B late in oogenesis. Several dominant female sterile alleles of *ovo*, such as ovo^{D1}, result from point mutations that create new in-frame translation initiator codons upstream of the OVO-B initiator site and result in translation of a new protein with an N-terminal extension (24,25). This new protein is thought to act like OVO-A and downregulate OVO-B; however, the mutant OVO-D isoform would be expressed inappropriately early in oogenesis and downregulation of OVO-B would occur prematurely, resulting in female sterility. Insertion of *gypsy* sequences that terminates transcription of the ovo^{D1} mRNA would prevent production of the OVO-D isoform and restore fertility.

Table 3. *gypsy* insertion alleles recovered at *ovo*

Allele name	Site[a]	Target site duplication[b]	Orientation in *ovo*[c]	Isolation method[d]	PCR primers[e]
lzl	A	TACATA	NR	NR	(23)
75-1	C	TGTGCA	+	2	p3, p4
ct5	B	GTAAAA	+	1	p3, p4
ct17	B	GTAAAA	+	1	p3, p4
ct16	C	TGCACA	–	1	p1, p4
ct28	C	TGTGCA	+	1	p3, p4
ct31	C	TGCACA	–	1	p1, p4
ct61	C	TGCACA	–	1	p1,p4
ct67	C	TGCACA	–	1	p1, p4
Df4	C	TGCACA	–	1	p1, p4
f2	C	TGTGCA	+	2	p3, p4
f5-1	C	TGCACA	–	2	p1, p4
f28-3	C	TGCACA	–	2	p1, p4
X	C	TGTGCA	+	2	p3, p4
Y	C	TGTGCA	+	2	p3, p4
R11	D	TGCACT[f]	–	2	(2) p1, p2
yct	D	TGCACT[f]	–	1	p3, pC
ct37	E	TGGAAC	+	1	p3, p4
5-1	F	TATAAA[f]	+	2	p3, p4
ct8	F	TATAAA[f]	+	1	p3, p4
f30-3	F	TATAAA[f]	–	2	p1, p4
yct	F	TATAAA[f]	+	1	p1, p4
Z	F	TATAAA[f]	+	2	p3, p4
ct7	G	TATACA	+	1	p3, p4
ct19	G	TATACA	+	1	p3, p4
ct30	G	TATACA	+	1	p3, p4
ct45	G	TATACA	+	1	p3, p4
ct90	G	TGTATA	–	1	p1, p4
R15	G	TATACA	+	2	(2) p3, p2
R9	H	TACATA	+	2	(2) p1, p2

[a]See Figure 3; the yct allele apparently contains two separate *gypsy* insertions in opposite orientations.
[b]Inferred from sequence of a single junction.
[c]+, *gypsy* and *ovo* transcribed in same direction; –, *gypsy* and *ovo* transcribed in opposite directions; NR, not reported.
[d]1, genetic cross; 2, 'feeding' experiment (Fig. 2).
[e]See figures 3 or 6 in Song *et al*. (2) as indicated.
[f]Single mismatch with YRYRYR consensus (Table 3).

We hypothesize that insertion of *gypsy* at these positions will lead to the premature termination of *ovo* transcription within the *gypsy* sequences. This is expected to be the case in the (+) orientation, in which the LTR termination sequences will be in their active orientation. It is not known whether *gypsy* inserted in the opposite orientation will result in truncation of the *ovo* mRNA, but we note that there are four copies of the sequence AATAAA, a sequence known to specify polyadenylation, on the

negative strand of the *gypsy* sequence. We believe these *gypsy* insertions may represent null alleles (or at least alleles that significantly reduce ovo^{D1} expression) because they phenotypically completely reverse the dominant female-sterile phenotype of the starting ovo^{D1} mutation.

gypsy integration specificity

Examination of the target sites within the *ovo* gene reveals that the previously reported 'consensus sequences' for *gypsy*, namely TATATA and TACATA, are poorly represented among the collection of *ovo* insertions sequenced (Table 3). However, if the consensus is relaxed to a slightly more degenerate one consisting of alternating pyrimidine (Y) and purine (R) residues, that is, a YRYRYR consensus target sequence, it is apparent that most of the *ovo* insertions conform. If this requirement is further relaxed by allowing a single mismatch to the consensus, ~90% of the insertions conform (as do all previously reported *gypsy* target site sequences). Interestingly, all of the single mismatches in our studies were in the 5th or 6th position of the target site duplication. The alternating purines and pyrimidines are a typical feature of DNA sequences able to adopt the Z conformation (27), raising the possibility that *gypsy* recognizes this DNA structural feature. Alternatively, other sequence variations, such as kinks have been associated with YR steps in DNA sequences (28), and these could conceivably be recognized by the *gypsy* integration machinery. Site C, our biggest hotspot, lies within a stretch of 11 alternating pyrimidines and purines and sites F, G and H lie within stretches of 10.

However, the target sequence and/or Z-DNA are unlikely to fully explain *gypsy* integration specificity in the *ovo* locus because there are many other sequences within *ovo* that conform to the consensus and yet do not appear in our collection of insertion mutations. In particular, there is a remarkable sequence of 39 consecutive alternating pyrimidines and purines (nt 184–222) that was not hit in our study, although a single insertion in this region has been reported previously as a mutation that reverts ovo^{D1} (23,29). Furthermore, our earlier study suggests that as many as 1/4 of all *gypsy* insertions occur within the *ovo* region cytologically (2). Thus, the *gypsy* integration machinery must recognize features of this region other than simply the DNA sequence or Z-DNA structure. In any case, it is very clear that *gypsy* sequences integrate very non-randomly in response to the selection for ovo^{D1} revertants.

ACKNOWLEDGEMENTS

We thank Maryellen Kurkulos for helpful discussions and Sun Song for doing some of the feeding experiments. Supported in part by NIH grant CA16519 (J.D.B.) and GM56022 (V.G.C.).

REFERENCES

1 Kim,A.I., Terzian,C., Santamaria,P., Pélisson,A., Prud'homme,N. and Bucheton,A. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 1285–1289.
2 Song,S.U., Gerasimova,T., Kurkulos,M., Boeke,J.D. and Corces,V.G. (1994) *Genes Dev.*, **8**, 2046–2057.
3 Pélisson,A., Song,S.U., Prud'homme,N., Smith,P., Bucheton,A. and Corces,V.G. (1994) *EMBO J.*, **8**, 4401–4411.
4 Tanda,S., Mullor,J.L. and Corces,V.G. (1994) *Mol. Cell. Biol.*, **14**, 5392–5401.
5 Boeke,J.D. and Stoye,J.P. (1997) In Varmus,H., Hughes,S. and Coffin,J. (eds), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 343–435.
6 Craigie,R. (1992) *Trends Genet.*, **8**, 187–189.
7 Carteau,S.C.H. and Bushman,F. (1998) *J. Virol.*, **72**, 4005–4014.
8 Withers-Ward,E.S., Kitamura,Y., Barnes,J.P. and Coffin,J.M. (1994) *Genes Dev.*, **8**, 1473–1487.
9 Chalker,D.L. and Sandmeyer,S.B. (1990) *Genetics*, **126**, 837–850.
10 Chalker,D.L. and Sandmeyer,S.B. (1992) *Genes Dev.*, **6**, 117–128.
11 Devine,S.E. and Boeke,J.D. (1994) *Nucleic Acids Res.*, **18**, 3765–3772.
12 Ji,H., Moore,D.P., Blomberg,M.A., Braiterman,L.B., Voytas,D.F., Natsoulis,G. and Boeke,J.D. (1993) *Cell*, **73**, 1007–1018.
13 Xiong,Y. and Eickbush,T.H. (1990) *EMBO J.*, **9**, 3353–3362.
14 Boeke,J.D. and Devine,S.E. (1998) *Cell*, **93**, 1087–1089.
15 Gai,X. and Voytas,D.F. (1998) *Mol. Cell*, **1**, 1051–1055.
16 Zou,S., Ke,N., Kim,J.M. and Voytas,D.F. (1996) *Genes Dev.*, **10**, 634–645.
17 Freund,R. and Meselson,M. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 4462–4464.
18 Bayev,A.A.,Jr, Lyubomirskaya,N.V., Dzhumagaliev,E.B., Ananiev,E.V., Amiantova,I.G. and Ilyin,Y.V. (1984) *Nucleic Acids Res.*, **12**, 3707–3723.
19 Marlor,R.L., Parkhurst,S.M. and Corces,V.G. (1986) *Mol. Cell. Biol.*, **6**, 1129–1134.
20 Prud'homme,N., Gans,M., Masson,M., Terzian,C. and Bucheton,A. (1995) *Genetics*, **139**, 697–711.
21 Mével-Ninio,M., Mariol,C.M. and Gans,M. (1989) *EMBO J.*, **8**, 1549–1558.
22 Mével-Ninio,M., Terracol,R., Salles,C., Vincent,A. and Payre,F. (1995) *Mech. Dev.*, **49**, 83–95.
23 Garfinkel,M.D., Wang,J., Liang,Y. and Mahowald,A.P. (1994) *Mol. Cell. Biol.*, **14**, 6809–6818.
24 Mével-Ninio,M., Fouilloux,E., Guénal,I. and Vincent,A. (1996) *Development*, **122**, 4131–4138.
25 Andrews,J., Levenson,J. and Oliver,B. (1997) *Dev. Genes. Evol.*, **207**, 482–487.
26 Chou,T.B., Noll,E. and Perrimon,N. (1993) *Development*, **119**, 1359–1369.
27 Haniford,D.B. and Pulleyblank,D.E. (1983) *J. Biomol. Struct. Dynam.*, **1**, 593–609.
28 Dickerson,R.E. (1998) *Nucleic Acids Res.*, **26**, 1906–1926.
29 Garfinkel,M.D., Lohe,A.R. and Mahowald,A.P. (1992) *Genetics*, **130**, 791–803.