DOI: 10.1534/genetics.108.094318

Insulator and Ovo Proteins Determine the Frequency and Specificity of Insertion of the *gypsy* Retrotransposon in *Drosophila melanogaster*

Mariano Labrador,[1] Ky Sha,[2] Alice Li[3] and Victor G. Corces[4]

Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218

Manuscript received July 23, 2008
Accepted for publication September 11, 2008

ABSTRACT

The *gypsy* retrovirus of Drosophila is quite unique among retroviruses in that it shows a strong preference for integration into specific sites in the genome. In particular, *gypsy* integrates with a frequency of >10% into the regulatory region of the *ovo* gene. We have used *in vivo* transgenic assays to dissect the role of Ovo proteins and the *gypsy* insulator during the process of *gypsy* site-specific integration. Here we show that DNA containing binding sites for the Ovo protein is required to promote site-specific *gypsy* integration into the regulatory region of the *ovo* gene. Using a synthetic sequence, we find that Ovo binding sites alone are also sufficient to promote *gypsy* site-specific integration into transgenes. These results indicate that Ovo proteins can determine the specificity of *gypsy* insertion. In addition, we find that interactions between a *gypsy* provirus and the *gypsy* preintegration complex may also participate in the process leading to the selection of *gypsy* integration sites. Finally, the results suggest that the relative orientation of two integrated *gypsy* sequences has an important role in the enhancer-blocking activity of the *gypsy* insulator.

RETROVIRAL DNA integration into the host genome is an essential step for production and replication of viral RNA. It has been traditionally difficult to study the factors controlling selection of integration sites, since most retroviruses integrate throughout the genome with no apparent DNA sequence specificity (Bushman *et al.* 2005). However, detailed analysis of multiple genomic integration sites *in vivo* has revealed that retroviruses have a strong preference for certain genomic regions. In particular, retroviruses integrate preferentially into actively transcribed DNA, which will thereafter facilitate transcription of the provirus (Scherdin *et al.* 1990). The distribution of retroviral integration sites along chromosomes suggests that open chromatin favors retroviral insertion, since integration events are favored in transcriptionally active chromatin and are rare in DNA sequences associated with heterochromatin (Lewinski *et al.* 2005; Yant *et al.* 2005). However, chromatin state or DNA accessibility could not be the only factor influencing integration, since different retroviruses manifest preferences for integration that are unlikely to be only the result of chromatin organization. For example, both HIV and murine leukemia virus (MLV) integrate in actively transcribing DNA, but HIV integrates with equal frequency throughout all transcribed DNA, whereas MLV integrates preferentially into transcription start sites (Schroder *et al.* 2002; Wu *et al.* 2003).

The mechanism of integration of retrotransposons is fundamentally identical to that of retroviruses (Bushman 2003). However, constraints imposed by small genome sizes have led some retrotransposons to the acquisition of mechanisms for site-specific integration. The best examples of site-specific integration are found in non-LTR retrotransposons such as the Drosophila telomeric elements *HeT-A*, *TART*, and *TAHRE* or rDNA elements such as the Drosophila *R1* and *R2* non-LTR retrotransposons (Yang *et al.* 1999; Christensen and Eickbush 2005; George *et al.* 2006). A number of examples involving integrases from yeast LTR retrotransposons have also shown that retroviral-like integrases have evolved to acquire strong site-specific integration properties (Sandmeyer 2003; Zhu *et al.* 2003; Brady *et al.* 2008; Gao *et al.* 2008). For example, the integration of *Saccharomyces cerevisiae* retrotransposons *Ty1* and *Ty3* is associated with RNA polymerase III transcription (Yieh *et al.* 2000; Bachman *et al.* 2005; Mou *et al.* 2006) and the *Tf1* retrotransposon from *Schizosaccharomyces pombe* integrates near RNA polymerase II promoters (Singleton and Levin 2002; Bowen *et al.* 2003). In other examples, transposable elements are targeted to heterochromatic sites by tethering mechanisms involving interactions between the integrase and DNA binding proteins. Targeting of the *Ty5* retrotransposon from *S. cerevisiae* to heterochromatin, for example, requires a six-amino-acid motif at the C terminus of the *Ty5* integrase that interacts with the heterochromatin protein Sir4 (Gai and Voytas

[1]*Present address:* Department of Biochemistry, Cellular and Molecular Biology, The University of Tennessee, Knoxville, TN 37996.

[2]*Present address:* Department of Biology, MIT, 77 Massachusetts Ave., Cambridge, MA 02139.

[3]*Present address:* Mount Sinai School of Medicine, New York, NY 10029.

[4]*Corresponding author:* Department of Biology, Emory University, 1510 Clifton Rd. NE, Atlanta, GA 30322. E-mail: vcorces@emory.edu

1998; Xie *et al.* 2001). It has also been proposed that the chromodomain (CHD domain) from certain transposable element chromointegrases targets the retrotransposon for insertion into sites bearing the specific epigenetic marks recognized by the CHD domain (Gao *et al.* 2008).

In contrast to mammalian retroviruses, Drosophila retroviruses such as *gypsy*, *ZAM*, or *Idefix*, display a high rate of site-specific integration into certain regions of the genome (Mevel-Ninio *et al.* 1989; Desset *et al.* 1999; Leblanc *et al.* 1999; Conte *et al.* 2000). The mechanisms governing this specificity however are poorly understood, but the genetic tools available in Drosophila provide a unique opportunity to analyze retroviral site integration specificity in higher eukaryotes. In particular, *gypsy* insertions into the *ovo* locus occur in the germ line of ~10% of the female offspring from mothers carrying permissive mutations in the *flamenco* (*flam*) locus (Prud'homme *et al.* 1995; Dej *et al.* 1998). *Gypsy* integrations take place specifically into a sequence of ~1.3 kb spanning the 5′ regulatory region of the *ovo* gene (Dej *et al.* 1998). The *flam* locus is located in the heterochromatin of the X chromosome (Prud'homme *et al.* 1995) and produces a long noncoding RNA that controls transcription of the *gypsy* retrovirus through the piwiRNA pathway (Brennecke *et al.* 2007).

The process of *gypsy* transposition is maternally regulated, involving maternally inherited *gypsy* particles that originate in the developing oocyte of *flam* mutant females. These females fail to produce the *flam* RNA, allowing the transcription of euchromatic *gypsy* elements in the follicle cells surrounding the oocyte during oogenesis. Transcription of *gypsy* in follicle cells leads to the formation of virus particles that infect the oocyte and subsequently participate in the integration of *gypsy* in the germ line of the resulting embryo after fertilization (Song *et al.* 1994, 1997). These integration events take place preferentially in the *ovo* gene, whose product is necessary for the development of the female germ line and the normal progression of oogenesis (Lu *et al.* 1998). The *ovo* gene encodes two isoform proteins, Ovo-A and Ovo-B, which have a common DNA-binding domain but different N-terminal domains. Ovo-B positively regulates the *ovo* promoter, whereas Ovo-A functions as a negative regulator of the *ovo* promoter (Andrews *et al.* 2000). Adult females homozygous for a null mutation of the *ovo* gene do not develop germ line cells. The ovo^{D1} allele is caused by a point mutation that creates a new in-frame methionine codon in the 5′ region of *ovo*, adding an extra amino terminus domain to Ovo-B that is normally only present in the wild-type Ovo-A protein (Mevel-Ninio *et al.* 1996). The ovo^{D1} allele is dominant negative and causes female sterility even when heterozygous. The sterility is due to the expression of Ovo^{D1}B protein, which is made at the same time of development as Ovo-B but has the repressor activity of Ovo-A; the presence of Ovo^{D1}B is sufficient to arrest oogenesis at stage 4 (Lu *et al.* 1998).

Insertion of *gypsy* into the ovo^{D1} allele in a heterozygous female reverts to fertility by preventing the expression of the Ovo^{D1}B protein, although the reversion occurs only in those germ cells in which *gypsy* is inserted into the ovo^{D1} sequence. The ability of *gypsy* to integrate specifically into *ovo* sequences was analyzed by Dej *et al.* (1998). These studies concluded that *gypsy* integrates in at least seven different target sites localized within a 200-bp sequence present in the promoter region of the *ovo* gene. Close analysis of these sites reveals a very relaxed consensus sequence consisting of six alternating pyrimidines and purines. The weak conservation of the observed target sequence suggests that *gypsy* site-specific integration is not due to a direct interaction of the *gypsy* integrase with these sequences. Instead, it has been proposed that Ovo proteins may mediate *gypsy* insertion specificity by promoting protein–protein interactions between Ovo or an associated protein and the *gypsy* preintegration complex (Labrador and Corces 2001).

The *gypsy* retrovirus of Drosophila also exhibits the interesting property of blocking enhancers from activating promoters when *gypsy* is inserted between them. This property is referred to as insulator activity and resides in the Suppressor of Hairy wing [Su(Hw)] binding sites present in a 350-bp sequence located in the 5′-UTR of the *gypsy* retroviral genome (Geyer *et al.* 1986; Spana *et al.* 1988; Gerasimova and Corces 1998, 2001; Gerasimova *et al.* 2000). In addition, *gypsy* insulators are also able to buffer transgenes from position effects by preventing heterochromatin from spreading through the chromatin fiber (Roseman *et al.* 1993). There is mounting evidence suggesting that *gypsy* insulators function by creating chromatin domains most probably defined by the interaction between adjacent insulator sites in chromosomes. Molecular evidence for such interactions has been obtained by measuring the distance between adjacent *gypsy* insertions in wild-type and in *su(Hw)* mutant cells (Gerasimova *et al.* 2000). These experiments revealed that the two *gypsy* sequences were significantly closer during interphase when the Su(Hw) protein was present. It has been proposed that such interactions might create chromatin domains by looping out the DNA contained between two interacting insulators. Additional evidence in support of this model has been provided by showing the presence of DNA loops attached at their base to the nuclear matrix by the *gypsy* insulator in the nucleus of Drosophila imaginal disc cells (Byrd and Corces 2003). Interaction between *gypsy* insulators is also supported by data showing that two adjacent insulators were able to cancel each other, no longer exerting their enhancer blocking effect when located between the enhancer and the promoter of a reporter gene (Cai and Shen 2001; Muravyova *et al.* 2001; Kuhn *et al.* 2003).

The molecular basis for interactions between individual insulators is not well understood but it has been

suggested that Modifier of mdg4 [Mod(mdg4)] and CP190, both components of the insulator complex, might facilitate such interactions by mediating protein–protein contacts between the BTB domains present in the two proteins (Gause *et al.* 2001; Ghosh *et al.* 2001; Golovnin *et al.* 2007). These two properties of *gypsy*, site-specific integration and insulator activity, have been the subject of intense but unrelated studies during the past two decades (Mevel-Ninio *et al.* 1989; Bucheton 1995; Prud'homme *et al.* 1995; Dej *et al.* 1998; Gerasimova and Corces 2001; Labrador and Corces 2001; Capelson and Corces 2004). Here, we attempt to analyze these two properties simultaneously in an effort to understand how insulators might mediate genome organization and how this organization may influence retroviral selection of integration sites through the genome. We have previously developed an assay to show that the 5′ regulatory region of *ovo* is able to recruit *gypsy* insertions independently of its position in the genome (Labrador and Corces 2001). We have also provided genetic evidence suggesting that the Ovo protein is directly implicated in such recruitment. We now take advantage of the ability of inducing two consecutive *gypsy* insertions into a *yellow* reporter gene to analyze the role that interactions between a *gypsy* provirus and the *gypsy* preintegration complex may play in the selection of retrovirus integration sites and the effect of the relative orientation of interacting proviruses on the enhancer-blocking activity of the *gypsy* retrovirus.

MATERIALS AND METHODS

Cloning DNA sequences containing Ovo binding sites: The three constructs ovo^{WT}, ovo^{EN}, and ovo^{DIM} were obtained using polymerase chain reaction (PCR) following standard procedures (94° for 1.0 min, 60° for 1.5 min, and 72° for 2.0 min). Each sequence was cloned into the *Not*I site of the Casper-*yellow* vector as described (Labrador and Corces 2001). ovo^{WT} was constructed by PCR amplification using genomic DNA from the Oregon R wild-type strain as template. The primers OVO-F (5′-GATGGGTCGCGGCCGCTTGAAAACAGCAGAAAAATA-3′) and OVO-R (3′-ATACTAATTAAGCATAAATTCGCCGGCGTAT TGTAC-5′) were used to amplify ovo^{WT}. ovo^{EN} and ovo^{DIM} were each built from a set of 50–60 overlapping oligonucleotides as follows.

OVO-dim1: 3′-CTACCCAGCGCCGGCGAACTTTTGTCGTCT TTTTATAT-5′
OVO-dim2: 5′-CAGCAGAAAAATATGCCAATTTGTTTTGAAT TTAACAGATTTT-AATAAATAGTTTTAACTTAATGG-3′
OVO-dim3: 3′-CAAAATTGAATTACCAAGTCGTGTAAAATTT GAAGGGAT-TCATATACGTATCGGGA-5′
OVO-dim4: 5′-ATATGCATAGCCCTGTTTTAATTTTTAATTT CAAGTTAATAA-CTTTTATTCACATCATACT-3′
OVO-dim5: 3′-ATAAGTGTAGTATGAGTCTATCATATAAAAT CACAT-GAGAAAACTCTATTAAAGTAGTG-5′
OVO-dim6: 5′-AGATAATTTCATCACTTGTTGATTCGTCGTT GGCAACTCTGC-ACCCACGATTCCAAATAGGATTA-3′
OVO-dim7: 3′-AAGGTTTATCCTAATCTTACTGTGTGTGGG TACCTTAACTTC-TTACATTCCTTCTTTCC-5′
OVO-dim8: 5′-GTAAGGAAGAAAGGGAGTGTGATCGAAAGT CCGTTCC-TTTGCTCAAATAGATAGCAATCGTC-3′
OVO-dim9: 3′-CTATCGTTAGCAGGCTCGCTTGCCTGTCTG TTTAAAGACTCTT-AGCGTGAAGAAACGAAGAGAG-5′
OVO-dim10: 5′-TTCTTTGCTTCTCTCATTTTCGGTGATTTTAG AATGCTTGC-TTATTGTGTGTGCACTCGAAAGTTCTATT-3′
OVO-dim11: 3′-GAGCTTTCAAGATAATCCAAGGTGTCCCAA AAATATGT-ATACTAATTAAGCATAAATTCGCCGGCGTC TTGTAC-5′
OVO-dim12: 5′-TTTATACATATGATTAATTCGTATTTAAGCG GCCGCAGCAGA-TGCTACTTAACGT-3′
OVO-en1: 5′-AAATGACGATGGGTCGCGGCCGCTTGAAAA CAGCAGAAAAA-TAAAGCCGTTAAAATT-3′
OVO-en2: 3′-ATTTCGGCAATTTTAACTTTTTCACGTCAAAT TTACATTGAC-AATTATACTCGTCATATAGT-5′
OVO-en3: 5′-TGAGCAGTATATCACGACTACAGTTAGAATTA GCTCTACGG-ATCCT-3′
OVO-en4: 3′-GAGATGCCTAGGAAAAATGTCAATGTATCGTC TCAGGAACC-TAAAAGGCAACGAAAAAA-5′
OVO-en5: 5′-TTCCGTTGCTTTTTTATTGAGGCTGTGTGCA AAATAAAGCCG-TTAAAATTGAAGGTTCCAC-3′
OVO-en6: 3′-AACTTCCAAGGTGCCAATTTACATTGACAATTA TACCGA-TAAATATGTCGTCCGAGGA-5′
OVO-en7: 5′-CAGCAGGCTCCTTTTTACAGTTACATATGATTA ATTCGTAT-TTAAGCGGCCGCAGCAGATGCTACTTAACGT-3′
OVO-en8: 3′-TTGTACAATAAATGATA-5′

We used a two-step PCR amplification as previously described (Dillon and Rosen 1990). In the first PCR step, the ovo^{EN} template was assembled using 1.0 μl (100 ng/μl) of each of the OVO-en1–OVO-en8 overlapping oligonucleotides plus 0.5 μl Ex Taq polymerase (TaKaRa Biomedicals). This PCR step consisted of seven cycles at 94° for 1.0 min, 50° for 1.5 min, and 72° for 2.0 min. The ovo^{DIM} template was assembled in identical fashion using OVO-dim1–OVO-dim12. Once the ovo^{EN} and ovo^{DIM} templates were assembled, they were subjected to the second round of PCR amplification. The second-step PCR amplification was identical to that used to amplify ovo^{WT} from genomic DNA, using the same end primers containing *Not*I restriction sites. After the second-step PCR amplification, cloning of ovo^{EN} and ovo^{DIM} into the Casper-*yellow* vector was identical to the procedure used to clone ovo^{WT} into the *Not*I sites of Casper *yellow*. The Casper-*yellow* plasmids containing ovo^{WT}, ovo^{EN}, and ovo^{DIM} were sequenced to confirm that all three correspond to the expected sequences shown in Figure 1. The orientation of ovo^{WT}, ovo^{EN}, and ovo^{DIM} in the Casper-*yellow* vector is opposite to the direction of transcription of the *yellow* gene. Nucleotide numbers indicating the location of insertions throughout the manuscript are as described by Geyer and Corces (1987). To designate transgenes carrying the Casper-*yellow* plasmids containing *ovo* we use the terminology $P[y^{ovo}; w^{+}]$. The super index ovo^{WT}, ovo^{EN}, and ovo^{DIM} was added to denote the particular *ovo* sequences cloned into *yellow*.

Induction of *gypsy* mobilization using *flam* females: Stocks were maintained on standard cornmeal medium at 25° in a humidity-controlled incubator. The *flam* flies were maintained as a *y v f mal flam/FM3* stock. This stock segregates homozygous *flam* flies and heterozygous *flam/FM3* females. Only heterozygous females were used to maintain the stock each generation. Virgin homozygous *y v f mal flam* were collected from the stock for crosses requiring homozygous *flam* flies. ovo^{D1} mutants were maintained in a stock by crossing ovo^{D1} males to females carrying attached *X* chromosomes. Flies carrying $P[y^{ovo}; w^{+}]$ transgenes display a *yellow* wild-type phenotype and were maintained in a $y\ w^{67c}$ background (Labrador and Corces 2001). To induce *gypsy* integrations into the $P[y^{ovo}; w^{+}]$ transgenes, virgin females homozygous for *y v f mal flam* were crossed to $P[y^{ovo}; w^{+}]$ transgenic males. Since a single insertion in the germ line can give rise to several mutant flies, single female crosses were performed instead of mass crosses. Virgin

F_1 females with genotype *y v f mal flam/y* w^{67c}*; P[*y^{ovo}*; w*$^+$*]* were individually crossed to *y* w^{67c} males. To detect *gypsy* insertions into the *P[*y^{ovo}*; w*$^+$*]* transgene, the offspring of this cross (F_2 generation) was systematically screened for individuals with a mutant *yellow* phenotype as follows. Individual *y v f mal flam/y* w^{67c}*; P[*y^{ovo}*; w*$^+$*]* females crossed to several *y* w^{67c} males were transferred to new vials containing fresh food every 3–4 days. A minimum of three vials was screened per cross and crosses with low numbers of offspring were discarded. The yellow phenotype of the F_2 progeny containing the *P[*y^{ovo}*; w*$^+$*]* transgene was screened under a dissecting microscope and 90–105 flies were counted and screened per cross. This process was repeated two to three times per vial to allow new flies to eclose and the number of flies counted each time was annotated in the vial. Only independent mutant flies (from different mothers) were recorded and crossed in new vials for further analysis. As a positive control, to test whether *flam* permissive females produced active virus particles, virgin *y v f mal flam* females were crossed to ovo^{D1} males and the offspring were screened for fertile females as described previously (Labrador and Corces 2001). We used the ovo^{D1} reversion assay (Prud'homme *et al.* 1995) to determine the *gypsy* mobilization activity of the *y v f mal flam* stock. Five to ten *y v f mal flam* virgin females were crossed with ovo^{D1} males and kept in the same vial for 3–5 days. After this time, parental flies were removed from the vials and the offspring larvae were allowed to develop. Groups of 5–10 ovo^{D1}/ovo^{+} females from the offspring were transferred to new vials and crossed to wild-type males. ovo^{D1}/ovo^{+} females are sterile unless the ovo^{D1} allele carries a new insertion of the *gypsy* element integrated into the ovo^{D1} regulatory region. After several days, vials were examined for the presence of eggs or larvae. All the mothers in vials containing eggs were dissected under the microscope and females displaying functional ovaries were considered fertile and counted as a *gypsy* integration event into the ovo^{D1} allele.

Analysis of *gypsy* insertions by PCR and DNA sequencing: Genomic DNA was extracted from 10–20 flies for the detection of *gypsy* integrations into *P[*y^{ovo}*; w*$^+$*]* transgenes. DNA extraction was carried out using the potassium acetate quick prep as described (Dej *et al.* 1998). The primers used to amplify *gypsy* insertions were *gypsy*-P3 CTTTGCCGAAAATATGCAATG and *gypsy*-P1 CAACATGACCG-AGGAGCGGTCATAAAC located outside the LTRs in the 5′ and 3′ ends of *gypsy*, respectively, and yP0030 GCCCGATTACCACATTGAG, yP1400 GTTGCA CAAAA-TTACCGGC, yP1450 CTGTGGGTGCAATGATTAG, yP1120 TCATTGCCGCAAGC-TCTG, and yP2900 CGCCA CGGTCCACAGAAGAG, which are located at different points in the regulatory sequences of the *yellow* gene and therefore could be present adjacent to new *gypsy* insertions. A combination of P1 or P3 with a *yellow* primer will detect the insertion of *gypsy* in either orientation. Alternative combinations of P3 or P1 with alternative primers in the *yellow* gene can detect the same insertion at the opposite end of *gypsy* and was used to confirm the results obtained with the first combination of primers. Approximately 50 ng of genomic DNA were used per PCR amplification. Amplified DNA fragments containing *gypsy* insertion sites were directly sequenced using *gypsy*-LTR3 AAATCGCTATCGCCACAAGGC or *gypsy*-LTR5 GCAGCGT GAAG-CAACACTCCC as sequencing primers.

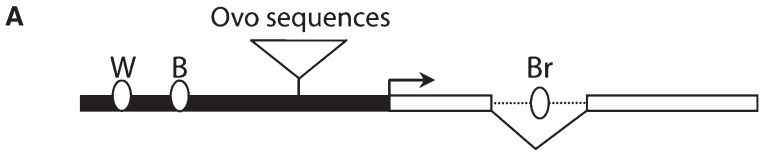
RESULTS

Ovo protein binding sites promote *gypsy* insertion into a *yellow* transgene: We have previously suggested that binding of the Ovo protein to the regulatory region of the *ovo* gene could account for the specificity of integration into this region of the *gypsy* retrotransposon (Labrador and Corces 2001). In previous experiments, a DNA fragment containing ~1.3 kb from the regulatory region of the *ovo* gene, including the promoter, was inserted into the Casper-*yellow* transformation vector. Transgenes containing these *ovo* gene sequences turned into highly specific targets for *gypsy* insertions, which occurred at a rate similar to that observed at the endogenous *ovo* gene (Dej *et al.* 1998; Labrador and Corces 2001).

To provide molecular evidence supporting the hypothesis that Ovo proteins are capable of targeting *gypsy* insertions into Ovo binding sites, we first asked whether a subset of the *ovo* sequences used in previous experiments, containing binding sites for the Ovo protein, would also induce a high rate of *gypsy* insertions. To this end we used PCR to specifically amplify a 508-bp DNA fragment (ovo^{WT}) containing nucleotides 512–1020 from the *ovo* gene regulatory sequence (Mevel-Ninio *et al.* 1996). This fragment lacks the promoter region of the Ovo-A isoform but still contains the promoter region of the Ovo-B isoform and includes five Ovo protein binding sites (Figure 1B). We then cloned the ovo^{WT} fragment into the *Not*I site at nucleotide 1980 of the Casper-*yellow* vector (Figure 1A), which is located between the wing blade and body enhancers and the promoter of the *yellow* gene (Labrador and Corces 2001). To determine whether the binding of Ovo proteins was necessary to specifically recruit *gypsy* insertions into Casper-*yellow* transgenes, we generated a synthetic DNA sequence identical to ovo^{WT} but containing specific point mutations into the five existing Ovo binding sites. The Ovo binding sites in Figure 1 were identified following the descriptions by Lu *et al.* (1998).

The induced point mutations result in sequences that strongly deviate from Ovo protein consensus binding sites, whereas the flanking sequences remain identical to those in ovo^{WT} (Figure 1C). We refer to this sequence as ovo^{DIM}. The ovo^{DIM} fragment was also cloned into the Casper-*yellow* vector and plasmids containing ovo^{WT} and ovo^{DIM} were independently microinjected into *y w;* Δ*2-3 Sb/TM6* embryos. Transgenic lines obtained from both constructs show a wild-type *yellow* phenotype in a y^1 mutant background, indicating that the *yellow* ovo^{WT} and ovo^{DIM} transgenes (*P[*y^{ovoWT}*; w*$^+$*]* and *P[*y^{ovoDIM}*; w*$^+$*]*, respectively) are able to express the Yellow protein in wing blades, body cuticle, and bristle tissues (see Figure 2B).

To test the role of Ovo binding sites in the specification of *gypsy* insertion, two independent transgenic lines from each construct were selected and transgenic males homozygous for the insertion were crossed to *flam* females. *Gypsy* elements are highly transcribed in the follicle cells surrounding the developing oocytes of these females, producing virus particles that will infect the oocyte and subsequently will integrate into the germ line DNA of the offspring (Pelisson *et al.* 1994; Song *et al.* 1994, 1997). Insertions of *gypsy* into *P[*y^{ovo}*; w*$^+$*]*



B
GATGGGTCGCGGCCGCTTGAAAACAGCAGAAAAATA**AAGCCGTTAAAA**TTGAATTTAACAG
ATTTTAATAAATAGTTTTAACTTAATGGTTCAGCACATTTTAAA**TGTAACTGTTA**ATATGC
ATAGCCCTGTTTTAATTTTTAATTTCAAGTTAATAACTTTTATTCACATCATACTCAGATA
GTATATTTTAGTGTACTCTTTTGAGATAATTTCATCACTTGTTGATTCGTCGTTGGCAACT
CTGCACCCACGA**CTACAGTTAGA**ATTAGAATGACACACACCCATGGAATTGAAGAATGTAA
GGAAGAAAGGGAGTGTGATCGAAAGTCCGTTCCTTT**TTACAGTTACA**TAGCAATCGTCCGA
GCGAACGGACAGACAAATTTCTGAGAATCGCACTTCTTTGCTTCTCTCATTTTCGGTGATT
TT**CCGTTGCTTTT**TTATTGTGTGTGCACTCGAAAGTTCTATTAGGTTCCACAGGGTTTTTA
TACATATGATTAATTCGTATTTAAGCGGCCGCATAACATG

C
GATGGGTCGCGGCCGCTTGAAAACAGCAGAAAAATA**TGCCAATTTGTT**TTGAATTTAACAG
ATTTTAATAAATAGTTTTAACTTAATGGTTCAGCACATTTTAAA**CTTCCCTAAGT**ATATGC
ATAGCCCTGTTTTAATTTTTAATTTCAAGTTAATAACTTTTATTCACATCATACTCAGATA
GTATATTTTAGTGTACTCTTTTGAGATAATTTCATCACTTGTTGATTCGTCGTTGGCAACT
CTGCACCCACGA**TTCCAAATAGG**ATTAGAATGACACACACCCATGGAATTGAAGAATGTAA
GGAAGAAAGGGAGTGTGATCGAAAGTCCGTTCCTTT**TTACAGATAGA**TAGCAATCGTCCGA
GCGAACGGACAGACAAATTTCTGAGAATCGCACTTCTTTGCTTCTCTCATTTTCGGTGATT
TT**AGAATGCTTGC**TTATTGTGTGTGCACTCGAAAGTTCTATTAGGTTCCACAGGGTTTTTA
TACATATGATTAATTCGTATTTAAGCGGCCGCATAACATG

D
GATGGGTCGCGGCCGCTTGAAAACAGCAGAAAAATA**AAGCCGTTAAAA**TTGAAAAAGTGCAG
TTTAAA**TGTAACTGTTA**ATATGAGCAGTATATCACGA**CTACAGTTAGA**ATTAGCTCTACGGA
TCCTTT**TTACAGTTACA**TAGCAAGAGTCCTTGATTTT**CCGTTGCTTTT**TTATTGAGGCTGTG
TGCAAAATA**AAGCCGTTAAAA**TTGAAGGTTCCACGGTTAAA**TGTAACTGTTA**ATATGGCTAT
TTATACAGCAGGCTCCTTT**TTACAGTTACA**TATGATTAATTCGTATTTAAGCGGCCGCATAA
CATG

FIGURE 1.—Ovo DNA sequences cloned into the *yellow* regulatory region of the different *P[y^{ovo}; w$^+$]* transgenes. (A) Schematic drawing of the *yellow* gene with the location of *ovo* sequences. The black box indicates the upstream regulatory region and the white boxes indicate exons 1 and 2 of the *yellow* gene. Tissue-specific wing (W), body (B), and bristle (Br) enhancers are indicated as ovals. The open triangle indicates the Ovo sequences inserted between wing and body enhancers and the *yellow* promoter at nucleotide 1980 of the *yellow* gene. (B) DNA fragment 508 bp long from the wild-type regulatory region of the *ovo* gene containing five Ovo binding sites. (C) Synthetic sequence identical to the one shown in B but with the Ovo binding sites mutated to prevent binding of Ovo proteins. (D) Synthetic 314-bp sequence containing eight Ovo binding sites separated by random DNA sequences.

transgenes can be detected in the F_2 generation by the changes in the *yellow* phenotype induced by the *gypsy* insulator after integration of *gypsy* into the regulatory region of the *P[y^{ovo}; w$^+$]* transgene (LABRADOR and CORCES 2001). For each experimental cross, controls using the *ovo^{D1}* reversion assay (MEVEL-NINIO *et al.* 1989; PRUD'HOMME *et al.* 1995) were performed to ensure that *gypsy* was actively transposing in the *flam* stock (Table 1). We analyzed an average of 100 F_2 flies per each *y v f mal flam/y w; P[y^{ovo}; w$^+$]* F_1 parent. The actual number of F_2 descendants counted for each cross was 95–105 and this variability was randomized by using a large number of F_1 progenitors and discarding low yield crosses equally among all samples. Statistical significance was calculated using the χ^2 contingency tables, assuming that the probability of *gypsy* insertions into transgenes was uniform across vials within each sample.

Results from these experiments indicate that flies containing *P[y^{ovoWT}; w$^+$]* transgenes show a high frequency of *gyspy* insertions into *yellow*, whereas flies containing *P[y^{ovoDIM}; w$^+$]* transgenes did not yield a single *gypsy* insertion. Table 1 indicates the number of flies analyzed and the frequencies of *gypsy* insertions in the *P[y^{ovo}; w$^+$]* transgenes from different transgenic lines. From 395 F_1 females analyzed, a total of 43 new insertions occurred into *P[y^{ovoWT}; w$^+$]* trangenes (10.85%), whereas analysis of 253 F_1 *P[y^{ovoDIM}; w$^+$]* females did not reveal any new *gypsy* insertion (Table 1). Therefore, these results suggest that removal of all Ovo binding sites from the *ovo^{WT}* sequence eliminates the ability of *gypsy* to specifically integrate into *P[y^{ovo}; w$^+$]* transgenes, suggesting that the frequency and specificity of *gypsy* insertions into the regulatory region of the *ovo* gene depends on the presence of the Ovo protein.

A synthetic DNA sequence containing an array of eight Ovo binding sites flanked by random intervening sequences actively promotes *gypsy* insertion into a *yellow* transgene: Although the experiments described above strongly suggest a requirement of Ovo proteins to specifically recruit *gypsy* insertions, we could not discard a role of the flanking sequences in this process. It is possible that these sequences could recruit additional proteins with an active role in targeting *gypsy* for integration. Therefore, we asked whether the presence of Ovo binding sites flanked by random DNA sequences in a synthetic sequence will also specifically induce a high rate of *gypsy* insertion. We recreated the Ovo binding sites previously identified in the ovo regulatory region (LU *et al.* 1998) to generate a 314-bp synthetic DNA sequence, named *ovo^{EN}* (Figure 1D). The *ovo^{EN}* sequence has eight Ovo binding sites identical to those found in wild-type ovo sequences (Figure 1A) except that they are flanked by random DNA sequences (Figure 1D). The *ovo^{EN}* sequence was inserted into the *Not*I site of the same Casper-*yellow* transformation vector used in the previous experiments (Figure 2A) and the resulting Casper-*yellow^{ovoEN}* DNA was microinjected into *y w ; Δ2-3 St/TM6* Drosophila embryos.

Two transgenic stocks carrying *yellow^{ovoEN}* transgenes (*P[y^{ovoEN}; w$^+$]*) were subsequently analyzed to test the ability of the *ovo^{EN}* sequence to specifically promote new *gypsy* integrations (Table 1). The phenotype of trans-

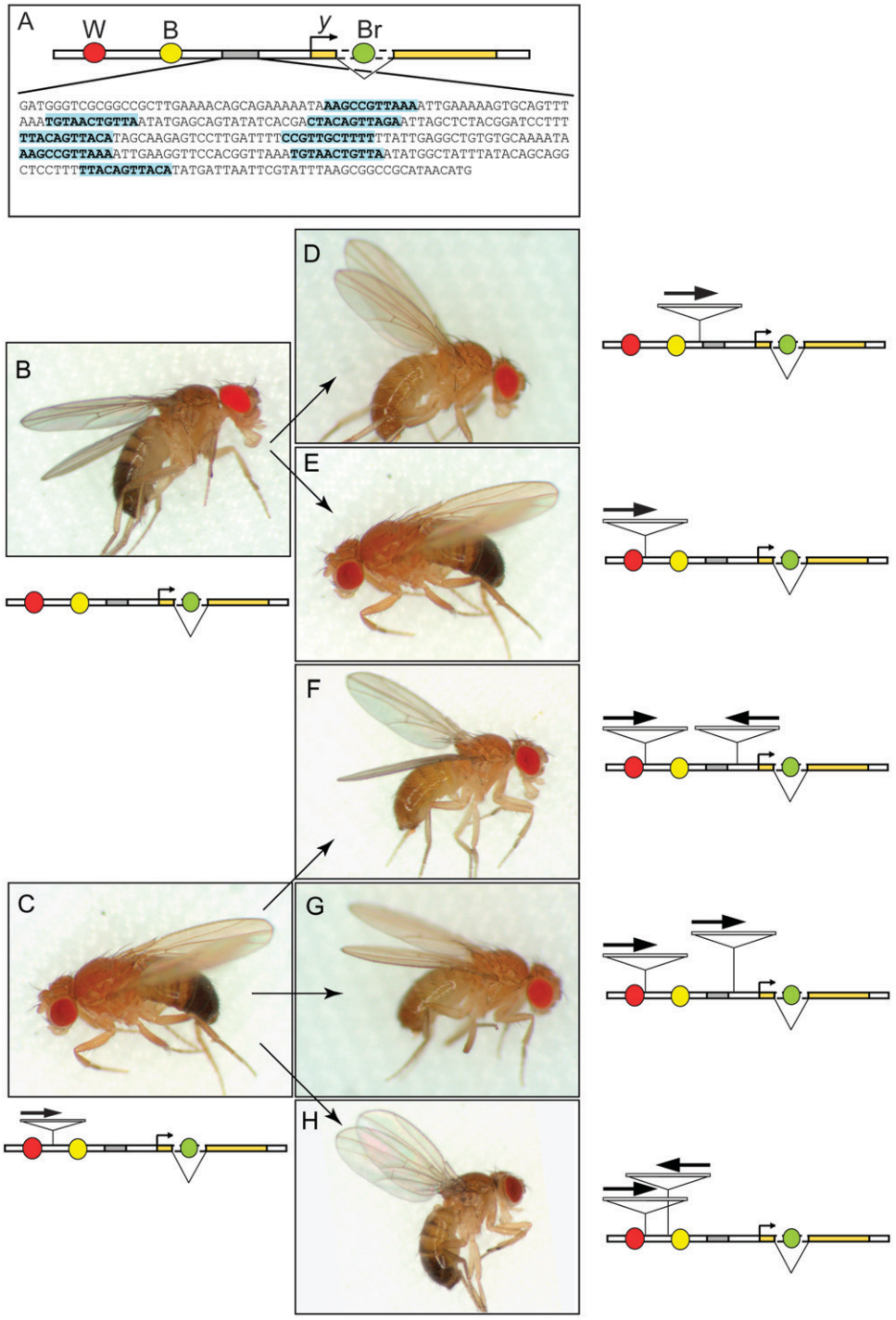


FIGURE 2.—*Gypsy*-induced *yellow* phenotypes depend on the integration sites as well as the relative orientation of *gypsy*. (A) The *yellow* reporter gene in the Casper-*yellow*^{ovo} transformation vector showing the synthetic DNA sequence containing eight Ovo binding sites. (B) *y w* male containing the *P[y*^{ovoEN}*; w*$^{+}$*]* 2.4 transgene shows black wing blade, black body, and black bristles. (C and E) Integration of *gypsy* between the wing and body enhancers results in yellow wings and black body phenotype. (D) Integration of *gypsy* between the body enhancer and the promoter results in yellow wings and body phenotype. (F) Integration of a second *gypsy* between the body enhancer and promoter, in opposite orientation to the preexisting one (C and E), results in black wings and yellow body, indicating that the wing enhancer is capable of bypassing the enhancer-blocking activity of the two insulators. (G) Integration of a second *gypsy* between the body enhancer and promoter, in the same orientation as the preexisting *gypsy* insertion (C and E), results in yellow wings and yellow body, suggesting that both enhancers, wing and body, are prevented from activating the *yellow* promoter by the enhancer-blocking activity of the two insulators. (H) Two insertions in opposite orientations with integration sites between the wing and body enhancers result in wild-type *yellow* phenotype.

genic flies carrying *P[y*^{ovoEN}*; w*$^{+}$*]* transgenes was indistinguishable from that of wild type, indicating that the synthetic *ovo*^{EN} sequence has no effect on the expression of the *yellow* transgene (Figure 2B). *Gypsy* insertions into the *P[y*^{ovoEN}*; w*$^{+}$*]* transgene were induced after activation of *gypsy* transposition in crosses with *flam* mutant females. Insertions were identified by their effect on the *yellow* phenotype in adult flies. *Gypsy* insertions into the *P[y*^{ovoEN}*; w*$^{+}$*]* transgene were identified in 24 of 488 (4.92%) analyzed females, further supporting the hypothesis that the presence of Ovo proteins is responsible for recruiting new *gypsy* insertions into the transgene. The lower frequency of *gypsy* insertion into *ovo*^{EN} *vs.* *ovo*^{WT} sequences may perhaps reflect a role of the sequences flanking the Ovo binding sites in the binding affinity of the Ovo proteins.

***Gypsy* insertions also occur into sequences flanking Ovo binding sites:** Analysis of the phenotypes emerging from new *gypsy* insertions into *P[y*^{ovo}*; w*$^{+}$*]* transgenes unexpectedly revealed that some insertions did not take place into *ovo* sequences. For example, flies from the strain *P[y*^{ovoEN}*; w*$^{+}$*]* 2.4 *y*^{wing} have wild-type black pigmen-

TABLE 1

Frequency of *gypsy* insertions into *P[y^{ovo}; w^{+}]* transgenes

Lines	Females tested	New insertions[a]	Frequency (%)
ovo^{D1} controls[b]	2134	141	6.61
P[y^{ovoWT}; w^{+}] 2.1	214	24	11.21
P[y^{ovoWT}; w^{+}] 1.1	181	19	10.50
P[y^{ovoDIM}; w^{+}] 2.1	132	0	0
P[y^{ovoDIM}; w^{+}] 1.2	121	0	0
P[y^{ovoEN}; w^{+}] 2.4	250	14	5.60
P[y^{ovoEN}; w^{+}] 4.2	238	10	4.20
P[y^{ovoEN}; w^{+}] 2.4 *y^{wing}*	128	17	13.28

P[y^{ovoEN}; w^{+}] 2.4 *vs. P[y^{ovoEN}; w^{+}]* 2.4 *y^{wing}* $\chi^2 = 6.63$; $P = 0.01$. *P[y^{ovoEN}; w^{+}]* 4.2 *vs. P[y^{ovoEN}; w^{+}]* 2.4 *y^{wing}* $\chi^2 = 8.76$; $P = 0.0031$.

[a] Number of females that produced at least one progeny with a new *gypsy* insertion into a *yellow* transgene.

[b] Controls using the *ovo^{D1}* assay were performed in parallel with experiments using transgenic lines to determine the activity of *gypsy* in *flam* flies.

tation in the body but *yellow* wing blades. This phenotype could be explained if a *gypsy* insertion occurred between the wing and body enhancers, with the *gypsy* insulator preventing the wing enhancer from activating transcription in wing tissues, but allowing the body enhancer to activate transcription in the body cuticle (Figure 2E). The insertion of *gypsy* between the wing and body enhancers would be unexpected, since the *ovo^{EN}* sequences are placed between the body enhancer and the promoter of *yellow*. The location of the *gypsy* insertion site in the *P[y^{ovoEN}; w^{+}]* 2.4 *y^{wing}* transgene was determined using PCR. *Gypsy*-P1 and -P3 primers (see MATERIALS AND METHODS) were selected from 5′ and 3′ non-LTR *gypsy* sequences, respectively, and primers *yellow*-Pn were selected from sequences present in the 5′ regulatory region of the *yellow* gene. PCR products were obtained using sets of two primers such that one would match *yellow* gene sequences and the second would match *gypsy* sequences. Using this procedure, each combination of primers successfully amplifying a specific DNA fragment provides information on the orientation and the approximate insertion site of the *gypsy* insertion in the *P[y^{ovo}; w^{+}]* transgene. The precise location of *gypsy* integration sites was then obtained by sequencing of the amplified DNA using *gypsy* LTR-P3 or *gypsy* LTR-P1 primers. The *gypsy* insertion in the *P[y^{ovoEN}; w^{+}]* 2.4 *y^{wing}* transgene occurred between the wing and body enhancers, as predicted by the phenotype (Figure 2E), at nucleotide 1111 in the upstream regulatory sequences of the *yellow* gene and outside of the *ovo^{EN}* sequences (GEYER and CORCES 1987).

Results from previous work had suggested that *gypsy* insertions occur only within the DNA fragment containing the binding sites for the Ovo protein, indicating that somehow targeting of the *gypy* preintegration complex and integration into DNA were part of the same process (DEJ *et al.* 1998; LABRADOR and CORCES 2001). Interestingly, integration of *gypsy* in the *P[y^{ovoEN}; w^{+}]* 2.4 *y^{wing}* transgene in this case took place 869 bp away from the insertion site of the *ovo^{EN}* sequences, suggesting that targeting of *gypsy* preintegration complexes and *gypsy* integration itself might take place in two independent events during the process of *gypsy* integration into genomic DNA. To obtain additional insights into the mechanisms determining the specificity of *gypsy* integration, we sequenced the insertion sites of a total of 85 insertion events (Figure 3 and Table 2). Results confirmed the previous observation and showed that insertions into *P[y^{ovo}; w^{+}]* transgenes occurred up to 869 bp upstream and up to 441 bp downstream of the actual *ovo* sequence (Figure 3). Interestingly, whereas insertions in *y^{ovoWT}* occur both into the *ovo* sequences or into the flanking *yellow* sequences, insertions induced by the *ovo^{EN}* sequence fall mostly into flanking *yellow* DNA.

In addition to the Ovo binding sites, there may be other sequences present in *ovo^{WT}* but not in *ovo^{EN}* that might be responsible for the observed difference in the distribution of *gypsy* integration sites when transgenes carrying these two sequences are compared. For example, the Ovo-B promoter is present in the o*vo^{WT}* sequence and is missing in the *ovo^{EN}* sequence. Although there is no experimental evidence to support this explanation, it is possible that the ability to activate transcription by the Ovo-B promoter in the *ovo^{WT}* sequence could involve local changes in chromatin structure in a manner that favors *gypsy* integration. These data suggest that, without other regulatory or promoter regions, Ovo proteins may recruit *gypsy* preintegration complexes to the vicinity of Ovo binding sites. The specific integration site may be selected later by the *gypsy* preintegration complexes within a few hundred nucleotides from the Ovo binding sites by a still unknown mechanism.

It has been shown previously that *gypsy* has a preference for integration into the sequences TATATA and TACATA, later defined as YRYRYR, in which Y is a pyrimidine and R is a purine (DEJ *et al.* 1998). Given that integration sites into this sequence are significantly more abundant than in any other sequence, our results confirm the preference of *gypsy* for integration into YRYRYR sequences, with 61 of the 85 sequences analyzed conforming to the consensus (Table 2). However, a number of other sequences are also capable of functioning as target sites for *gypsy* integration, suggesting that, although important, DNA sequence is not the only factor determining integration site selection. Interestingly, the pattern of insertion sites is significantly different at each side of the *ovo* sequences (Figure 3). Clusters of insertions into sequences distal to the promoter occur at intervals of 145, 159, 246, and 319 nucleotides, suggesting that these intervals may correspond to nucleosomal DNA that is protected and not accessible to the *gypsy* integrase. On the other hand, insertions into sequences proximal to the *yellow* promoter occur in a more arbitrary fashion. A large cluster

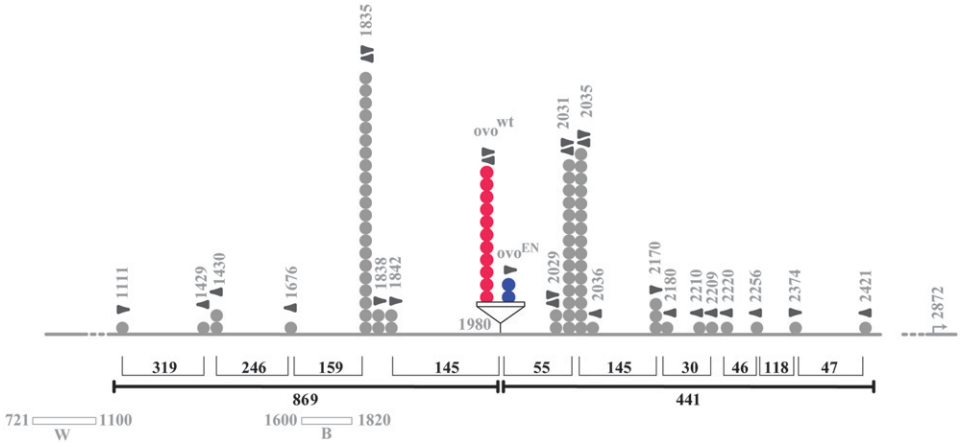


FIGURE 3.—*gypsy* integration sites and orientation in the upstream regulatory region of *P[y^ovo; w^+]* transgenes. A total of 85 integration sites were characterized at the sequence level (see Table 2). Filled circles correspond to integration events. Arrowheads on top of circles indicate orientation of the insertions with the arrowhead pointing toward the 3′ end of *gypsy*. Ovo sequences were cloned at nucleotide 1980 in the upstream regulatory sequence of *yellow*. Integration sites within the *ovo* sequences are colored. Boldface segments below indicate distances in nucleotides between the first and the last insertion before and after *ovo* DNA sequences. Numbers in brackets indicate distance between adjacent integration sites. White boxes illustrate the position of wing (W) and body (B) enhancers. The *yellow* promoter is indicated by a broken arrow. Drawing is not to scale.

of multiple integrations is observed at 55 bp from the Ovo binding sequences and small clusters or single insertion sites are found at intervals of only 145, 30, 46, 118, and 47 nucleotides. Except for the 145-bp interval between insertions at nucleotides 2036 and 2170, which may suggest the presence of a nucleosome, the remaining insertion sites are randomly scattered, suggesting that positioned nucleosomes may be absent from this region at the time of retroviral integration. Therefore, our results suggest that in addition to the recruitment of *gypsy* preintegration complexes by Ovo proteins other factors such as the specific DNA sequence of the insertion site, nucleosomal arrangement, or other DNA binding proteins may also play an important role in the final determination of the integration site.

TABLE 2

Target site sequences and integration sites into *P[y^ovo; w^+]* transgenes

Target site duplication	Integration site	No. of integrations
TATATG[a]	1835	21
TGTATG[a]	2035	15
TATATG[a]	2031	14
TATAAA	491 in *ovo^WT*	6
TGTATA[a]	489 in *ovo^WT*	4
TATACA[a]	2170	3
TATATC	1838	2
TGTATA[a]	2029	2
TATACT	1842	2
TGTTTG	1430	2
CTGTAT	252 in *ovo^EN*	2
TATAAA	1111	1
TATTAA	2374	1
TATAAC	2220	1
CGTTAT	2210	1
TGTAGA	1676	1
TGTATT	2256	1
TATACA[a]	2180	1
ACGTTA	2209	1
TATGCT	2036	1
GTGTTT	1429	1
GCACGA	2421	1
TGCACA[a]	454 in *ovo^WT*	1

[a] Target site duplications conform to the preferred consensus YRYRYR for *gypsy* integration.

The presence of a *gypsy* provirus results in high frequency of secondary *gypsy* insertions: Interactions between Ovo proteins and components of the *gypsy* preintegration complex are likely responsible for targeting *gypsy* insertions to the vicinity of Ovo binding sites. Since *gypsy* insulator proteins mediate interactions between individual insulators to create chromatin loops (GERASIMOVA *et al.* 2000; CAI and SHEN 2001; MURAVYOVA *et al.* 2001; BYRD and CORCES 2003; KUHN *et al.* 2003), it is possible that the presence of a *gypsy* provirus in a *P[y^ovo; w^+]* transgene will increase the frequency of secondary *gypsy* insertions. To test this possibility, we compared the insertion rates into *yellow-ovo* sequences present in transgenes containing or lacking a copy of a *gypsy* provirus. We used the transgenic line *P[y^ovoEN; w^+]2.4 y^wing*, which contains a *gypsy* insertion in the *yellow* transgene regulatory region between the wing and body enhancers and displays wild-type black pigmentation in the body but yellow wing blades. Starting with this transgenic line we then induced *gypsy* mobilization using the same *flam* assay described above. We set up 128 F_1 *y w f ml flam/ P[y^ovoEN; w^+]2.4 y^wing* single female crosses and, as in previous assays, we analyzed an average of 100 offspring from each single female cross (range 95–105), screening for changes in the *yellow* phenotype (Table 1).

Transgenes with new insertions resulted in three new phenotypes when starting from flies with an initial *y^wing* phenotype (Figure 2): flies with a *y^2* phenotype (yellow wing blades, yellow body, and black bristles), flies with wild-type phenotype (black wing blades, black body, and

black bristles), and y^{body} flies (black wing blades, yellow body, and black bristles). We obtained 17 such events with a final frequency of 13.28%. A Fisher exact test indicates that the number of new *gypsy* insertions in $P[y^{ovoEN}; w^+]2.4\ y^{wing}$ is significantly higher than the same number in the parental transgenic stock $P[y^{ovoEN}; w^+]2.4$ (Table 1). Since the initial $P[y^{ovoEN}; w^+]$ transgene is inserted in the same genomic location in both stocks, the observed differences are not due to chromosomal position effects. The results suggest that the higher frequency of second *gypsy* integrations may be due to interactions between the preexisting *gypsy* element in the genomic copy of the transgene and proteins present in the *gypsy* preintegration complex.

Insulator function depends on the orientation of *gypsy* provirus sequences: To confirm that the new phenotypes described in the previous section are due to new *gypsy* integration events, we mapped the integration sites and orientation of the putative new insertions. All new phenotypes contain a single additional new *gypsy* insertion. Interestingly, each phenotype correlates either with the orientation or with a specific integration site of the second *gypsy* insertion (Figure 2). We obtained a total of 10 independent new insertions resulting in a y^{body} phenotype and all of them originated from a second insertion, in an orientation opposite to the first one, between the body cuticle enhancer and the *yellow* promoter (Figure 2F). In these flies the body cuticle enhancer is flanked by two *gypsy* insertions in opposite orientations and they display wild-type coloration of all tissues except for the body cuticle. We obtained only one event of a y^+ phenotype, which originated from a second *gypsy* insertion between the body cuticle enhancer and the wing enhancer but integrated in an orientation opposite to the first one (Figure 2H). Finally, we obtained 6 independent new insertions resulting in y^2 phenotypes (Figure 2G). Interestingly, all 6 independent new lines carried a second *gypsy* insertion between the body enhancer and the promoter, the same arrangement observed in y^{body} strains.

The only difference between flies with y^{body} phenotypes and y^2 phenotypes is the relative orientation of the two *gypsy* insertions. In flies displaying a y^{body} phenotype the two copies of *gypsy* are in opposite orientations (Figure 2F) whereas in flies with a y^2 phenotype the two copies of *gypsy* are in the same orientation (Figure 2G). These results show that two copies of the *gypsy* provirus inserted in the regulatory region of the *yellow* gene have the same properties as two paired insulators in experimental transgenes (Cai and Shen 2001; Muravyova *et al.* 2001; Kuhn *et al.* 2003). The presence of a single *gypsy* provirus prevents the wing and body cuticle enhancers from activating the *yellow* promoter due to the presence of the *gypsy* insulator (Figure 2D). However when a second insertion occurs in the opposite orientation, distal enhancers can bypass both *gypsy* insertions and their respective insulators (Figure 2, F and H).

Although the specific mechanism explaining this property is still unknown, current models suggest that this phenomenon is the result of molecular interactions between the two adjacent insulators and that these interactions loop out the intervening DNA. Enhancers outside the resulting chromatin loop are capable of activating transcription of distal promoters, but enhancers trapped within the chromatin loop are blocked by the interacting insulators. Our results confirm this model when the two copies of *gypsy* are in opposite orientation and suggest that such interactions can occur at distances much longer than the distances so far reported (usually a few hundred nucleotides). When the wing enhancer bypasses two *gypsy* insertions in opposite orientation as in Figure 2F and once the distance between the two insertion sites and the size of the two *gypsy* sequences (7.4 kb) is taken into account, the distance between the two *gypsy* insulators is >14 kb. Interestingly, these interactions between insulators seem to be abrogated when the two copies of the *gypsy* provirus are in the same orientation. Since the only genetic difference between *gypsy* insertion arrangements originating the y^{body} and y^2 phenotypes is the relative orientation of the two *gypsy* elements, the results suggest that the relative orientation of *gypsy* sequences may affect the ability of the two insulators to pair and thus influence the enhancer blocking properties of *gypsy* insulators.

DISCUSSION

The *gypsy* retrovirus of Drosophila may offer valuable clues as to how retroviruses develop strategies to specifically select integration sites into the genome. Results shown here suggest that interactions between the *gypsy* preintegration complex and, most likely, Ovo proteins are sufficient to promote site-specific integration of *gypsy* into the *ovo* locus of Drosophila. Alteration of Ovo binding sites from a wild-type *ovo* gene fragment abolishes the ability of *gypsy* to specifically integrate into adjacent sequences. In addition, a synthetic DNA sequence carrying eight Ovo binding sites flanked by random DNA sequences is sufficient to function as a highly specific target for integration of the *gypsy* retrovirus. Although direct interactions between Ovo proteins and the *gypsy* preintegration complex have not been substantiated, the data point to a mechanism by which Ovo proteins may tether the *gypsy* preintegration complexes to their binding sites.

It is tempting to speculate that the *gypsy* integrase may actually interact with the Ovo proteins and that such interaction may target integration to genome sites enriched in Ovo. Evidence demonstrating tethering of integrases as a mechanism capable of targeting retroviruses to specific DNA binding sites has been shown in experiments using fusion proteins in which the DNA binding domain of phage λ-repressor was fused to the integrase of the HIV retrovirus and successfully showed preferential integra-

tion into target DNA near λ-repressor-binding sites (Bushman 1994). Similarly, experiments with yeast retrotransposons have shown that such interactions may occur between the retrotransposon integrase and proteins that target the integration to their cognate chromosomal DNA binding sites (Gai and Voytas 1998; Xie *et al.* 2001; Sandmeyer 2003; Zhu *et al.* 2003; Gao *et al.* 2008).

It has been reasoned that retrotransposons in small genomes such as that of yeast may develop tethering mechanisms of site-specific integration by stimulating interactions between the integrase encoded by the retrotransposon and endogenous proteins, thus minimizing the chances of deleterious mutations induced by retrotransposon integration events. In larger genomes such mechanisms appear infrequently, probably due to the lack of selective pressure from the host genome. In humans for example, sequences related to interspersed retroviruses occupy >50% of the genome and only relatively low frequencies of integration events in specific target spots have been reported (Farnet and Bushman 1997; Suzuki *et al.* 2004; Bushman *et al.* 2005; Shun *et al.* 2007). In Drosophila only a number of specialized non-LTR retrotransposons have acquired specificity of integration associated with specialized chromosomal regions such as telomeres or ribosomal DNA (Jakubczak *et al.* 1992; Casacuberta and Pardue 2005). Nonvertebrate retroviruses such as ZAM, Idefix, and *gypsy* appear to be an exception when compared with their vertebrate counterparts, since some degree of sequence specificity and targeted site integration has been described in all three (Mevel-Ninio *et al.* 1989; Desset *et al.* 1999; Leblanc *et al.* 1999; Conte *et al.* 2000). The high rate of insertion of *gypsy* into the *ovo* gene and the role apparently played by the Ovo proteins provide an excellent tool to study the integration mechanism and how retroviruses may acquire integration site specificity *in vivo*.

Interestingly, even though Ovo proteins appear to have a role in targeting *gypsy* to their binding sites, *gypsy* insertion sites do not necessarily occur into the Ovo binding sequences themselves. From a total of 85 sequenced insertion sites, only 13 (15%) occurred into the DNA fragment containing Ovo binding sites; the remaining integration sites fall within an interval of >1300 bp flanking the DNA containing the Ovo binding sites. The analysis of insertion sites suggests that the targeting and the integration mechanisms are uncoupled, with the precise integration sites distributed in a nonrandom manner. Results shown here confirm previous observations suggesting that *gypsy* has a preference for integration into YRYRYR sequences. However, a variety of other sequences appear to be able to function as integration sites. This disparity makes it difficult to draw a clear conclusion as to what is the mechanism ultimately involved in selecting target sites; however, it is tempting to speculate a role for nucleosome positioning as one of the factors determining the selection of insertion sites by the *gypsy* retrovirus.

Several indirect lines of evidence suggest such a role. For example, integration frequencies are significantly higher between the promoter of the *yellow* gene and the Ovo binding sites, indicating a preference that probably reflects a difference in chromatin structure. Forty-two independent integration events occurred into a fragment of 441 bp located between Ovo binding sequences and the promoter of the *yellow* gene, whereas only 30 integrations occurred in a 869-bp DNA fragment located distal to the promoter and upstream of the Ovo binding sites. This asymmetry does not appear to reflect sequence differences or viability effects, suggesting an epigenetic basis for integration site selection. In addition, insertion sites found distally to the *yellow* promoter and upstream of the Ovo binding DNA sequences appear to be spatially distributed in four intervals >140 bp, whereas insertions proximal to the promoter are distributed in a random manner, with only one large gap of 145 bp and apparently lacking meaningful spacing intervals. A possible interpretation of these results is that *gypsy* integration preferentially occurs at specific points of either the nucleosome or the linker DNA. In the distal interval, one spacing >140 bp could reflect nucleosome positioning, whereas nucleosomes may be absent or not positioned in the promoter proximal region of the transgene.

Results presented here also suggest that a preexisting *gypsy* insertion significantly increases the chances of new *gypsy* insertions into adjacent sequences by more than twofold. We speculate that such enhancement of insertion frequency might be the result of a tethering mechanism mediated by protein–protein interactions between the *gypsy* element located in the chromosomal DNA and the *gypsy* preintegration complex during the normal process by which a new copy of the retrovirus is inserted into the chromosome. Since two copies of the *gypsy* insulator have been shown to be able to interact with each other (Gerasimova *et al.* 2000; Byrd and Corces 2003), it is tempting to speculate that interactions between *gypsy* insulators are responsible for the increased frequency of *gypsy* insertions. Nevertheless, we cannot rule out the possibility that *gypsy* sequences other than the insulator or proteins associated with the *gypsy* element itself are responsible for the observed interactions. If the high frequency of secondary insertions is due to interactions between *gypsy* insulator proteins present in the provirus and in the preintegration complex, the results would lend support to proposed models suggesting that individual insulators located in different regions of a chromosome can interact to form chromatin loops.

The analysis of phenotypes resulting from double insertions allows further elaboration of this model and offers additional insights into the mechanisms by which insulators affect enhancer–promoter interactions. For example, we have shown that a wing enhancer distal to two adjacent *gypsy* insertions is capable of bypassing the

activity of the two insulators when the two copies of the *gypsy* provirus are inserted in opposite orientation. These results demonstrate that interactions such as the ones determined genetically in transgenes, involving pairs of 400-bp *gypsy* insulators (Cai and Shen 2001; Muravyova *et al.* 2001; Kuhn *et al.* 2003), also occur between pairs of *gypsy* insulators embedded in the *gypsy* provirus and suggest that establishing such interactions is part of the normal life cycle strategy used by the retrovirus. Interestingly, when two *gypsy* insertions occur in the same orientation, distal enhancers are unable to bypass the two insulators and are blocked from activating the promoter, contrary to what it has been observed with direct repeats of insulator sequences (Cai and Shen 2001; Muravyova *et al.* 2001; Kuhn *et al.* 2003). The main difference between the two sets of experiments is the presence of additional DNA sequences in the *gypsy* provirus. These sequences may be able to form a stem-loop structure when the two copies of *gypsy* are arranged in opposite, but not when they are in the same, orientation. A similar role has been suggested for the relative orientation of insulator sequences between interacting Mcp insulators (Kyrchanova *et al.* 2007). The stem-loop structure would allow interactions between insulator proteins present in the two copies of *gypsy* with opposite orientations but a direct tandem arrangement of the two copies of the provirus would preclude such interactions. These observations support the hypothesis that interactions between paired insulators are required to bypass insulator function and allow enhancer–promoter communication.

This work was supported by U.S. Public Health Service Award GM35463 from the National Institutes of Health.

LITERATURE CITED

Andrews, J., D. Garcia-Estefania, I. Delon, J. Lu, M. Mevel-Ninio *et al.*, 2000 OVO transcription factors function antagonistically in the Drosophila female germline. Development **127:** 881–892.

Bachman, N., M. E. Gelbart, T. Tsukiyama and J. D. Boeke, 2005 TFIIIB subunit Bdp1p is required for periodic integration of the Ty1 retrotransposon and targeting of Isw2p to S. cerevisiae tDNAs. Genes Dev. **19:** 955–964.

Bowen, N. J., I. K. Jordan, J. A. Epstein, V. Wood and H. L. Levin, 2003 Retrotransposons and their recognition of pol II promoters: a comprehensive survey of the transposable elements from the complete genome sequence of Schizosaccharomyces pombe. Genome Res. **13:** 1984–1997.

Brady, T. L., C. L. Schmidt and D. F. Voytas, 2008 Targeting integration of the Saccharomyces Ty5 retrotransposon. Methods Mol. Biol. **435:** 153–163.

Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis *et al.*, 2007 Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell **128:** 1089–1103.

Bucheton, A., 1995 The relationship between the flamenco gene and gypsy in Drosophila: how to tame a retrovirus. Trends Genet. **11:** 349–353.

Bushman, F., M. Lewinski, A. Ciuffi, S. Barr, J. Leipzig *et al.*, 2005 Genome-wide analysis of retroviral DNA integration. Nat. Rev. Microbiol. **3:** 848–858.

Bushman, F. D., 1994 Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. Proc. Natl. Acad. Sci. USA **91:** 9233–9237.

Bushman, F. D., 2003 Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. Cell **115:** 135–138.

Byrd, K., and V. G. Corces, 2003 Visualization of chromatin domains created by the gypsy insulator of Drosophila. J. Cell Biol. **162:** 565–574.

Cai, H. N., and P. Shen, 2001 Effects of cis arrangement of chromatin insulators on enhancer-blocking activity. Science **291:** 493–495.

Capelson, M., and V. G. Corces, 2004 Boundary elements and nuclear organization. Biol. Cell **96:** 617–629.

Casacuberta, E., and M. L. Pardue, 2005 HeT-A and TART, two Drosophila retrotransposons with a bona fide role in chromosome structure for more than 60 million years. Cytogenet. Genome Res. **110:** 152–159.

Christensen, S. M., and T. H. Eickbush, 2005 R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. Mol. Cell. Biol. **25:** 6617–6628.

Conte, C., V. Calco, S. Desset, P. Leblanc, B. Dastugue *et al.*, 2000 Impact of multiple insertions of two retroelements, ZAM and Idefix at an euchromatic locus. Genetica **109:** 53–59.

Dej, K. J., T. Gerasimova, V. G. Corces and J. D. Boeke, 1998 A hotspot for the Drosophila gypsy retroelement in the ovo locus. Nucleic Acids Res. **26:** 4019–4025.

Desset, S., C. Conte, P. Dimitri, V. Calco, B. Dastugue *et al.*, 1999 Mobilization of two retroelements, ZAM and Idefix, in a novel unstable line of Drosophila melanogaster. Mol. Biol. Evol. **16:** 54–66.

Dillon, P. J., and C. A. Rosen, 1990 A rapid method for the construction of synthetic genes using the polymerase chain reaction. Biotechniques **9**(298): 300.

Farnet, C. M., and F. D. Bushman, 1997 HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. Cell **88:** 483–492.

Gai, X., and D. F. Voytas, 1998 A single amino acid change in the yeast retrotransposon Ty5 abolishes targeting to silent chromatin. Mol. Cell **1:** 1051–1055.

Gao, X., Y. Hou, H. Ebina, H. L. Levin and D. F. Voytas, 2008 Chromodomains direct integration of retrotransposons to heterochromatin. Genome Res. **18:** 359–369.

Gause, M., P. Morcillo and D. Dorsett, 2001 Insulation of enhancer-promoter communication by a gypsy transposon insert in the Drosophila cut gene: cooperation between suppressor of hairy-wing and modifier of mdg4 proteins. Mol. Cell. Biol. **21:** 4807–4817.

George, J. A., P. G. DeBaryshe, K. L. Traverse, S. E. Celniker and M. L. Pardue, 2006 Genomic organization of the Drosophila telomere retrotransposable elements. Genome Res. **16:** 1231–1240.

Gerasimova, T. I., K. Byrd and V. G. Corces, 2000 A chromatin insulator determines the nuclear localization of DNA. Mol. Cell **6:** 1025–1035.

Gerasimova, T. I., and V. G. Corces, 1998 Polycomb and trithorax group proteins mediate the function of a chromatin insulator. Cell **92:** 511–521.

Gerasimova, T. I., and V. G. Corces, 2001 Chromatin insulators and boundaries: effects on transcription and nuclear organization. Annu. Rev. Genet. **35:** 193–208.

Geyer, P. K., and V. G. Corces, 1987 Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the yellow locus in Drosophila melanogaster. Genes Dev. **1:** 996–1004.

Geyer, P. K., C. Spana and V. G. Corces, 1986 On the molecular mechanism of gypsy-induced mutations at the yellow locus of Drosophila melanogaster. EMBO J. **5:** 2657–2662.

Ghosh, D., T. I. Gerasimova and V. G. Corces, 2001 Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function. EMBO J. **20:** 2518–2527.

Golovnin, A., A. Mazur, M. Kopantseva, M. Kurshakova, P. V. Gulak *et al.*, 2007 Integrity of the Mod(mdg4)-67.2 BTB domain is critical to insulator function in Drosophila melanogaster. Mol. Cell. Biol. **27:** 963–974.

Jakubczak, J. L., M. K. Zenni, R. C. Woodruff and T. H. Eickbush, 1992 Turnover of R1 (type I) and R2 (type II) retrotransposable elements in the ribosomal DNA of *Drosophila melanogaster*. Genetics **131:** 129–142.

Kuhn, E. J., M. M. Viering, K. M. Rhodes and P. K. Geyer, 2003 A test of insulator interactions in Drosophila. EMBO J. **22:** 2463–2471.

Kyrchanova, O., S. Toshchakov, A. Parshikov and P. Georgiev, 2007 Study of the functional interaction between Mcp insulators from the Drosophila bithorax complex: effects of insulator pairing on enhancer-promoter communication. Mol. Cell. Biol. **27:** 3035–3043.

Labrador, M., and V. G. Corces, 2001 Protein determinants of insertional specificity for the Drosophila gypsy retrovirus. Genetics **158:** 1101–1110.

Leblanc, P., B. Dastugue and C. Vaury, 1999 The integration machinery of ZAM, a retroelement from Drosophila melanogaster, acts as a sequence-specific endonuclease. J. Virol. **73:** 7061–7064.

Lewinski, M. K., D. Bisgrove, P. Shinn, H. Chen, C. Hoffmann *et al.*, 2005 Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. J. Virol. **79:** 6610–6619.

Lu, J., J. Andrews, D. Pauli and B. Oliver, 1998 Drosophila OVO zinc-finger protein regulates ovo and ovarian tumor target promoters. Dev. Genes Evol. **208:** 213–222.

Mevel-Ninio, M., E. Fouilloux, I. Guenal and A. Vincent, 1996 The three dominant female-sterile mutations of the Drosophila ovo gene are point mutations that create new translation-initiator AUG codons. Development **122:** 4131–4138.

Mevel-Ninio, M., M. C. Mariol and M. Gans, 1989 Mobilization of the gypsy and copia retrotransposons in Drosophila melanogaster induces reversion of the ovo dominant female-sterile mutations: molecular analysis of revertant alleles. EMBO J. **8:** 1549–1558.

Mou, Z., A. E. Kenny and M. J. Curcio, 2006 Hos2 and Set3 promote integration of Ty1 retrotransposons at tRNA genes in *Saccharomyces cerevisiae*. Genetics **172:** 2157–2167.

Muravyova, E., A. Golovnin, E. Gracheva, A. Parshikov, T. Belenkaya *et al.*, 2001 Loss of insulator activity by paired Su(Hw) chromatin insulators. Science **291:** 495–498.

Pelisson, A., S. U. Song, N. Prud'homme, P. A. Smith, A. Bucheton *et al.*, 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.

Prud'homme, N., M. Gans, M. Masson, C. Terzian and A. Bucheton, 1995 Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. Genetics **139:** 697–711.

Roseman, R. R., V. Pirrotta and P. K. Geyer, 1993 The su(Hw) protein insulates expression of the Drosophila melanogaster white gene from chromosomal position-effects. EMBO J. **12:** 435–442.

Sandmeyer, S., 2003 Integration by design. Proc. Natl. Acad. Sci. USA **100:** 5586–5588.

Scherdin, U., K. Rhodes and M. Breindl, 1990 Transcriptionally active genome regions are preferred targets for retrovirus integration. J. Virol. **64:** 907–912.

Schroder, A. R., P. Shinn, H. Chen, C. Berry, J. R. Ecker *et al.*, 2002 HIV-1 integration in the human genome favors active genes and local hotspots. Cell **110:** 521–529.

Shun, M. C., N. K. Raghavendra, N. Vandegraaff, J. E. Daigle, S. Hughes *et al.*, 2007 LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes Dev. **21:** 1767–1778.

Singleton, T. L., and H. L. Levin, 2002 A long terminal repeat retrotransposon of fission yeast has strong preferences for specific sites of insertion. Eukaryot. Cell **1:** 44–55.

Song, S. U., T. Gerasimova, M. Kurkulos, J. D. Boeke and V. G. Corces, 1994 An env-like protein encoded by a Drosophila retroelement: evidence that gypsy is an infectious retrovirus. Genes Dev. **8:** 2046–2057.

Song, S. U., M. Kurkulos, J. D. Boeke and V. G. Corces, 1997 Infection of the germ line by retroviral particles produced in the follicle cells: a possible mechanism for the mobilization of the gypsy retroelement of Drosophila. Development **124:** 2789–2798.

Spana, C., D. A. Harrison and V. G. Corces, 1988 The Drosophila melanogaster suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon. Genes Dev. **2:** 1414–1423.

Suzuki, Y., H. Yang and R. Craigie, 2004 LAP2alpha and BAF collaborate to organize the Moloney murine leukemia virus preintegration complex. EMBO J. **23:** 4670–4678.

Wu, X., Y. Li, B. Crise and S. M. Burgess, 2003 Transcription start regions in the human genome are favored targets for MLV integration. Science **300:** 1749–1751.

Xie, W., X. Gai, Y. Zhu, D. C. Zappulla, R. Sternglanz *et al.*, 2001 Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. Mol. Cell. Biol. **21:** 6606–6614.

Yang, J., H. S. Malik and T. H. Eickbush, 1999 Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. Proc. Natl. Acad. Sci. USA **96:** 7847–7852.

Yant, S. R., X. Wu, Y. Huang, B. Garrison, S. M. Burgess *et al.*, 2005 High-resolution genome-wide mapping of transposon integration in mammals. Mol. Cell. Biol. **25:** 2085–2094.

Yieh, L., G. Kassavetis, E. P. Geiduschek and S. B. Sandmeyer, 2000 The Brf and TATA-binding protein subunits of the RNA polymerase III transcription factor IIIB mediate position-specific integration of the gypsy-like element, Ty3. J. Biol. Chem. **275:** 29800–29807.

Zhu, Y., J. Dai, P. G. Fuerst and D. F. Voytas, 2003 Controlling integration specificity of a yeast retrotransposon. Proc. Natl. Acad. Sci. USA **100:** 5891–5895.

Communicating editor: M. J. Simmons