Protected *P*-Element Termini Suggest a Role for Inverted-Repeat-Binding Protein in Transposase-Induced Gap Repair in *Drosophila melanogaster*

Brian E. Staveley,[1] Tim R. Heslip,[2] Ross B. Hodgetts[1] and John B. Bell[1]

Department of Genetics, University of Alberta, Edmonton, Alberta, Canada T6G 2E9

Manuscript received September 19, 1994
Accepted for publication December 1, 1994

ABSTRACT

P-element transposition is thought to occur by a cut-and-paste mechanism that generates a double-strand break at the donor site, the repair of which can lead to internally deleted elements. We have generated a series of both phenotypically stronger and weaker allelic derivatives of vg^{21}, a *vestigial* mutant caused by a *P*-element insertion in the 5′ region of the gene. Virtually all of the new alleles arose by internal deletion of the parental element in vg^{21}, and we have characterized a number of these internally deleted *P* elements. Depending upon the selection scheme used, we see a very different spectrum of amount and source of *P*-element sequences in the resultant derivatives. Strikingly, most of the breakpoints occur within the inverted-repeats such that the last 15–17 bp of the termini are retained. This sequence is known to bind the inverted-repeat-binding protein (IRBP). We propose that the IRBP may act to preserve the *P*-element ends when transposition produces a double-strand gap. This allows the terminus to serve as a template upon which DNA synthesis can act to repair the gap. Filler sequences found at the breakpoints of the internally deleted *P* elements resemble short stretches, often in tandem arrays, of these terminal sequences. The structure of the filler sequences suggests replication slippage may occur during the process of gap repair.

THE *P* element of *Drosophila melanogaster* is the transposable element responsible for the phenomenon of P-M hybrid dysgenesis (reviewed in ENGELS 1989; RIO 1990). The complete *P* element is 2907 bp in length and has terminal 31-bp inverted-repeats (O'HARE and RUBIN 1983). The element encodes both a transposase (KARESS and RUBIN 1984) and a somatic repressor (MISRA and RIO 1990). Transposition usually generates an 8-bp duplication at the target site (see ENGELS 1989; KAUFMAN and RIO 1992), and it requires the presence of a functional *P*-element transposase gene (KARASS and RUBIN 1984) and *P*-element ends (MULLINS *et al.* 1989). The process is one of cut-and-paste (KAUFMAN and RIO 1992), and the donor site can remain unaltered, revert to wild type, become internally deleted (ENGELS *et al.* 1990) or recruit sequences from an ectopic site (GLOOR *et al.* 1991). Of interest to us is the relationship between the *P*-element activities that lead to internal deletion (ENGELS *et al.* 1990; TAKASU-ISHIKAWA *et al.* 1992) and those specialized activities that include targeted gene replacement (GLOOR *et al.* 1991) and targeted transposition (HESLIP and HODGETTS 1994).

Corresponding author: Brian E. Staveley, Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2E9. E-mail: bstavele@gpu.srv.ualberta.ca

[1] *Present address*: Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2E9.

[2] *Present address*: Developmental Biology Center, University of California, Irvine, CA 92717.

Internally deleted versions make up the majority of *P* elements within the genomes of *P* strains (O'HARE and RUBIN 1983; O'HARE *et al.* 1992). A model of *P*-element transposition, which accounts for internally deleted elements, has been proposed that invokes double-strand gap repair at the break after excision (ENGELS *et al.* 1990) and synthesis-dependent strand annealing (SDSA) to allow efficient copying of sequences to the gap (NASSIF *et al.* 1994). Complete excision of a *P* element leaves behind a double-strand break that is repaired using the homologue, the sister chromatid, or an ectopic copy as a template (GLOOR *et al.* 1991). According to this model (NASSIF *et al.* 1994) each side of the double-strand break independently invades a template and initiates DNA synthesis from this template. The newly synthesized single strands then anneal to each other and further DNA synthesis completes the repair. This model accounts for a number of the incomplete *P* elements recovered (see O'HARE *et al.* 1992). It is further supported by the analysis of two recombinant *P* elements, one bearing a *white-apricot* construct (containing the *gypsy* retrotransposon) and another containing two direct copies of the flip recombinase target sequence (FRT) flanking a white minigene (KURKULOS *et al.* 1994). In both cases intervening material between the 276-bp terminal repeats of *gypsy* and between the 599-bp FRT sites was preferentially deleted.

The appearance of filler sequences at the deletion breakpoints of internally deleted *P* elements is common in naturally-occurring *P* elements (O'HARE and RUBIN

1983; SEARLES et al. 1986; WILLIAMS et al. 1988; O'HARE et al. 1992) and in internally deleted derivatives of recombinant *P* elements (TAKASU-ISHIKAWA et al. 1992). DNA replication and template slippage was suggested as a possible source of some filler sequences (O'HARE and RUBIN 1983). A different model, based upon a hairpin repair mechanism, has also been proposed (TAKASU-ISHIKAWA et al. 1992). In their model the free ends generated by the excision of the *P* element are ligated to form two hairpin structures. Nicking and linearization of the hairpins followed by ligation and DNA synthesis would then introduce extra sequences into the deletion breakpoint. We have found that filler sequences at the breakpoints of internally deleted *P* elements often resemble short stretches of the *P*-element terminus, sometimes in tandem arrays. We favor a replication-slippage model to explain their creation during gap repair.

Herein we describe a set of internally deleted *P* elements generated, in the presence of a stable source of transposase, from an extant *P*-element insertion allele. The vg^{21} allele (WILLIAMS and BELL 1988) has a wild-type phenotype as a homozygote but a strap wing phenotype in combination with most strong alleles of *vestigial*. We refer to this as a cryptic phenotype. In phenotypic selection experiments we have isolated both strong alleles of *vestigial* and revertants of the haplo-insufficient vg^{21} hemizygous phenotype. All of the internal deletions generated have a breakpoint within either one or both of the terminal 31-bp inverted-repeats and most retain the last 16 ± 1 bp of the end. This terminal sequence has been demonstrated to bind the inverted-repeat-binding protein (IRBP) (RIO and RUBIN 1988). The above model of double-strand gap repair (ENGELS et al. 1990; NASSIF et al. 1994) does not adequately account for our recovery of internally deleted *P* elements with breakpoints within the terminal inverted-repeats. We propose that the IRBP acts to maintain the integrity of the *P*-element terminus that can then serve as a template for DNA synthesis initiated at the other side of the double-strand gap and, therefore, repair the break.

MATERIALS AND METHODS

Drosophila strains: The $P[lacZ;ry^+]^{c49}$; *b cn* vg^{21}; *ry* line was constructed from *b cn* vg^{21}, which possesses an internally deleted *P* element inserted in the 5′ untranscribed region of *vestigial* (WILLIAMS and BELL 1988), and $P[lacZ;ry]^{c49}$, an *X*-linked enhancer trap *P* element (O'KANE and GEHRING 1987). The $P[Ddc^+](35C)D\ dc^{ts2}$ *pr* vg^{21} line was produced as previously described (HESLIP and HODGETTS 1994). The *w*; *Sb* $P[\Delta 2\text{-}3,ry^+]$ *e*/ *TM6,e* stock provided the stable source of transposase (ROBERTSON et al. 1988). The other Drosophila strains are described in LINDSLEY and ZIMM (1992). All flies were maintained at 22° on a standard yeast/glucose medium.

Selection for *vestigial* phenotype: The extreme *vestigial* alleles were generated during experiments to study targeted transposition (HESLIP et al. 1992; HESLIP and HODGETTS 1994; STAVELEY et al. 1994). $P[lacZ;ry^+]^{c49}$; *b cn* vg^{21}; *ry* females were crossed to *w*; *Sb* $P[\Delta 2\text{-}3,ry^+]$ *e*/ *TM6,e* males to produce dysgenic males of the genotype $P[lacZ;ry^+]^{c49}$; *b cn* vg^{21}/+; *Sb* $P[\Delta 2\text{-}3,ry^+]$ *e*/ *ry*. About 10 dysgenic males were mated *en masse* to 20–25 *b cn* vg^{21} females in pint bottles. Individuals heterozygous for vg^{21} and a strong *vg* allele have nicked or scalloped wings, but vg^{21} homozygotes display a wild-type phenotype. Selection was based upon the assumption that recruitment to or deletion of sequences at vg^{21} would result in a noncryptic allele of *vestigial* (HESLIP et al. 1992). Offspring displaying any wing-margin defects were crossed to $Df(2R)vg^B$/ *SM5* individuals, or *Sp Bl* L^m/ *CyO* and lines were established for further analysis. The lines were inspected for extreme *vestigial* wing phenotypes. The allele designation reflects the number of the *en masse* mating vessel, such that the lines labeled $vg^{21\text{-}n.1}$ and $vg^{21\text{-}n.2}$ (where n = 1 to 300) represent different isolates from a single *en masse* mating and therefore may be the result of a single premeiotic dysgenic event. The $vg^{21\text{-}f}$ series were generated in a similar manner in that $P[Ddc^+](35C)\ Ddc^{ts2}$ *pr* vg^{21}/ *CyO*; *Sb* $P[\Delta 2\text{-}3,ry^+]$ *e*/+ dysgenic males were crossed to Ddc^{ts2} *pr* vg^{79d5}/ *CyO* females, and offspring displaying a phenotype more severe than vg^{21}/ vg^{79d5} were selected (HESLIP and HODGETTS 1994).

Selection for phenotypic revertants of *vestigial*: Revertants of the haplo-insufficient property of vg^{21} (the scalloped to strapped wing phenotype observed in vg^{21} hemizygotes) were generated using three different chromosomal constitutions. Dysgenic males bearing the transposase source $P[\Delta 2\text{-}3](99B)$ on chromosome *3* and the vg^{21}-bearing second chromosome possessed one of three second chromosome homologues: (1) a standard sequence chromosome with a wild-type copy of *vestigial*, (2) the multiply inverted balancer chromosome (*CyO*) with a wild-type copy of *vestigial* or (3) the deficiency-containing chromosome $Df(2R)vg^B$, deleted for *vestigial* and several surrounding genes (LASKO and PARDUE 1988). The revertant allele designation reflects the chromosome configuration from which it was selected such that $vg^{21\text{-}IR1}$ is the first of a series of revertant alleles descended from a dygenic male that, in addition to the vg^{21} chromosome, possessed a standard sequence chromosome.

Selection for altered size of PCR fragment: To verify that the specificity in the positioning of the deletion breakpoints that were isolated in the above experimental procedures (see RESULTS) were not merely the result of an inherent bias of the above methods of selection, a small scale experiment that did not discriminate by phenotype was undertaken. A similar experiment has been carried out at *white* (JOHNSON-SCHLITZ and ENGELS 1993). Dysgenic males of the genetic constitution vg^{21}/+; *Sb* $P[\Delta 2\text{-}3,ry^+]$ *e*/+ (with and without the presence of the $P[lacZ;ry^+]^{c49}$ transposon) were crossed to $Df(2R)vg^B$/ *CyO* females; *Cy* Sb^+ progeny were crossed to $Df(2R)vg^B$/ *CyO*, and lines were established. From these stocks single flies possessing one of the dysgenic male's second chromosomes in hemizygous combination with the *vestigial* deficiency chromosome were evaluated for PCR fragment size (see *DNA amplification*). Those different from either vg^+ or vg^{21} were subjected to DNA sequencing (see *Sequencing of PCR products*) and the nature of the *vestigial* phenotype was evaluated.

DNA amplifications: DNA extracts from single flies were generated in a manner similar to the method described by GLOOR et al. (1993). However, in some instances the crude DNA preparations were further extracted with equal volumes of phenol/chloroform/isoamyl alcohol, precipitated with 3 volumes of 95% ethanol and resuspended in distilled water. The primers used in DNA amplification flank the *P*-element insertion of vg^{21} located at the 5′ end of the gene and are described in HESLIP et al. (1992) as primer 1, 5′-AATCAA-

TABLE 1

Analysis of deletion breakpoints in derivatives of vg^{21} selected for strong *vestigial* phenotype

Allele	5′ break	3′ break	Sequence of breakpoint junction	Phenotype
vg^{21}	380	2602 (306)	. . . acctagtaaa/aaattagaat . . .	Cryptic
$vg^{21\text{-}1.3}$	105	2892 (16)	. . . gttgtgtgcgg/atgttatttcatcatg	Strap
$vg^{21\text{-}6.1}$	2714	2892 (16)	. . . ctttaaaaata/atgttatttcatcatg	Strap
$vg^{21\text{-}33.1}$	294	2889 (19)	. . . tgatacccact/cttatgttatttcatcatg	Extreme
$vg^{21\text{-}56.1}$	110	2891 (17)	. . . gtgcggacgaa/tatgttatttcatcatg	Extreme
$vg^{21\text{-}60.1}$	118	2891 (17)	. . . gaatttttttt/tatgttatttcatcatg	Extreme
$vg^{21\text{-}74.2}$	2703	2892 (16)	. . . taaaaataaaa/atgttatttcatcatg	Strap
$vg^{21\text{-}78.1}$	148	2891 (17)	. . . gtggaataaaa/TGTTTG/tatgttatttcatcatg	Extreme
$vg^{21\text{-}81.1}$	142	2892 (16)	. . . ccttacgtgga/atgttatttatcatg[a]	Extreme
$vg^{21\text{-}83\text{-}1}$	130	2892 (16)	. . . gaaacattaa/CATTAACATTAACATTAACATTA/atgttatttcatcatg	Strap
$vg^{21\text{-}158.1}$	142	2892 (16)	. . . ccttacgtgga/TAGTAAAATGTTTGT/atgttatttcatcatg	Extreme
$vg^{21\text{-}161.1}$	105	2892 (16)	. . . gttgtgtgcgg/atgttatttcatcatg	Strap
$vg^{21\text{-}197.5}$	148	2891 (17)	. . . gtggaataaaa/TGTTTG/tatgttatttcatcatg	Extreme
$vg^{21\text{-}212.2}$	251	2899 (9)	. . . agcagagcctt/ttcatcatg	Extreme
$vg^{21\text{-}213.1}$	105	2892 (16)	. . . gttgtgtgcgg/atgttatttcatcatg	Strap
$vg^{21\text{-}213.4}$	2697	2892 (16)	. . . aattattaaaa/atgttatttcatcatg	Strap
$vg^{21\text{-}241}$	205	2897 (11)	. . . aaattaattca/GTTAATTA/atttcatcatg	Extreme
$vg^{21\text{-}f106}$	119	2893 (15)	. . . aatttttttt/tgttatttcatcatg	Extreme

The deletion breakpoints are numbered according to the published sequence of the complete *P* element (O'HARE and RUBIN 1983), although the *P* element of vg^{21} is internally deleted (687 bp) (WILLIAMS *et al.* 1988). The numbers in the 5′- and 3′-break columns indicate the nucleotide positions at either side of the internal deletion. The parenthetical number after the 3′ breakpoint indicates the number of base pairs of the *P* element retained at this end. The DNA sequence of the junction is presented such that a slash separates sequences assigned to either side of the breakpoint. Lower case letters represent known *P*-element sequences and upper case represent filler sequences present at the breakpoint in a number of alleles.

[a] c at position 2901 is not present.

GTGGGCGGTGCTTG-3′, and primer 7, 5′-ATCCCGCGC-GGCGGTGAGAG-3′. The amplification reactions were carried out in a total volume of 30 μl starting with 3 μl of the above DNA extract (~50 ng of genomic DNA) and a final concentration of 50 mM Tris-Cl (pH 9.2), 1.5 mM $MgCl_2$, 0.01% dithiothreitol, 0.1 μg/μl bovine serum albumin (Boehringer Mannheim), 200 μM of dATP, dCTP, dGTP and dTTP, 0.3 ng/μl of primer and 1 U of *Taq* DNA polymerase (BRL). The amplifications were produced using a Stratagene Robocylcer 40 with a program of 1 cycle of 5 min at 95°, 1 min at 60° and 3 min at 73°, followed by 30 cycles of 1 min at 92°, 1 min at 60° and 3 min at 73°. Products of the amplification were fractionated by electrophoresis on 1–2% agarose gels in TAE (40 mM Tris-acetate, 1 mM EDTA).

Sequencing of PCR products: Some DNA sequencing was performed using a Sequenase II DNA sequencing kit as follows. Initially, PCR products amplified from several alleles were cloned into pT7T3-19U and subjected to double-stranded plasmid sequencing according to manufacturer's instructions (New England Biolabs). PCR products were directly sequenced in the following manner. Approximately 0.3–0.5 μg of amplified DNA isolated utilizing the GenecleanII procedure from 1–2% agarose in TAE, was mixed with 100 ng of primer, both dissolved in water, boiled for 8 min and immediately plunged into liquid nitrogen until ready to carry out the sequencing reactions. In other cases DNA amplification products were purified from agarose gels with GenecleanII and sequenced using the ABI Taq DyeDeoxy Terminator Cycle Sequencing Kit according to the manufacturer's instructions. The sequencing reactions were fractionated and analyzed on an ABI 370A DNA sequencing apparatus.

RESULTS

In the first set of experiments, strong *vestigial* alleles were generated from the cryptic vg^{21} allele as described in MATERIALS AND METHODS. With the exception of targeted events, all of the strong *vestigial* alleles that were recovered contained deletions. In one screen 16 independent alleles were isolated from approximately 50,000 chromosomes scored: 12 had internal deletions, three had excisions of regions outside of the *P* element (and removed an adjacent *Eco*RI restriction site) and one was a targeted event (STAVELEY *et al.* 1994). All of the internally deleted *vestigial* alleles share one striking feature, a breakpoint within the 31-bp terminal inverted repeat at the 3′ end (Table 1). To differentiate these from other nonautonomous *P* elements, we refer to deletions that have breakpoints within one or both inverted-repeats but that retain the extreme terminal sequences as terminal internal deletions. The number of base pairs retained at the 3′ end of these alleles ranges from 9 to 19. Most (15/17) retained the terminal 15–19 bp of the inverted-repeat. In contrast the position of the 5′ breakpoint of these internal deletions varies from nucleotide 105 ($vg^{21\text{-}1.3}$) to nucleotide 2714 ($vg^{21\text{-}6.1}$). Most of the 5′ breakpoints are clustered in three regions of the vg^{21} *P* element: nucleotides 105 to 119, 142 to 148 and 2697 to 2714 (numbered according

TABLE 2

Analysis of deletion breakpoints in derivatives of vg^{21} selected for reversion

Allele	5′ break	3′ break	Sequence of breakpoint junction	Phenotype
vg^{21}	380	2602 (306)	. . . acctagtaaa/aaattagaat . . .	Cryptic
$vg^{21\text{-}IR2}$	17	2892 (16)	catgatgaaataacata/TAATATATGTATATATTATAAATTATA/atgttatttcatcatg	Revertant
$vg^{21\text{-}IR3}$	17	2891 (17)	catgatgaaataacata/TA/tatgttatttcatcatg	Revertant
$vg^{21\text{-}IR6}$	16	2890 (18)	catgatgaaataacat/CATGTTATGTTATGTTATG/ttatgttatttcatcatg	Revertant
$vg^{21\text{-}IR7}$	16	2892 (16)	catgatgaaataac at gttatttcatcatg	Revertant
$vg^{21\text{-}IR8}$	17	352 (334)	catgatgaaataacata/ACTCAAT/gactcaacgca . . .	Revertant
$vg^{21\text{-}IIR2}$	16	2895 (13)	catgatgaaataaca t tatttcatcatg	Revertant
$vg^{21\text{-}IIR8}$	16	2893 (15)	catgatgaaataaca t gttatttcatcatg	Revertant
$vg^{21\text{-}IIIR1}$	17	2893 (15)	catgatgaaataacata/TATGTAACATAACATAACATAACATC/tgttatttcatcatg	Revertant
$vg^{21\text{-}IIIR2}$	16	2897 (11)	catgatgaaataac at ttcatcatg	Revertant
$vg^{21\text{-}IIIR3}$	16	2836 (72)	catgatgaaataacat/tgcacttattt . . .	Revertant
$vg^{21\text{-}IIIR4}$	19	2859 (49)	catgatgaaataaca taag tggatgtctc . . .	Revertant
$vg^{21\text{-}IIIR5}$	17	2891 (17)	catgatgaaataacata/TATGTAACATAACATAACATAACA/tatgttatttcatcatg	Revertant
$vg^{21\text{-}IIIR7}$	13	2891 (17)	catgatgaaataa/TAATGATAA/tatgttatttcatcatg	Revertant
$vg^{21\text{-}IIIR8}$	16	2892 (16)	catgatgaaataac at gttatttcatcatg	Revertant

The deletion breakpoints are numbered according to the published sequence of the complete *P* element (O'HARE and RUBIN 1983), and the sequence of the junctions are presented as in Table 1. Underlined nucleotides represent sequences that may have arisen from either side of the breakpoint.

to the published sequence of the complete *P* element; O'HARE and RUBIN 1983). These are located at the 5′ side of very AT-rich regions of the *P* element: the sequence from position 105 to 130 is 5′-GACGAATTT-TTTTTTGAAAACATTAA-3′ (80% AT), 142 to 171 is 5′-AATAAAAAAAAATGAAATATTGCAAATTTT-3′ (90% AT) and 2686 to 2719 is 5′-AATTATTAAAAA-TAAAACTTTAAAAATAATTT-3′ (97% AT). Alleles $vg^{21\text{-}1.3}$, $vg^{21\text{-}161.1}$ and $vg^{21\text{-}213.1}$ are identical but were isolated at different times and must have arisen independently. Apparently, deletions of the 3′ region of the *P* element result in loss of the cryptic phenotype of vg^{21}.

Although 12 of the 17 events examined are simple deletions, the other five have AT-rich filler sequences included at the breakpoint (Table 1). The $vg^{21\text{-}83.1}$ allele has four tandem copies of the hexameric repeat of 5′-CATTAA-3′ as filler. Alleles $vg^{21\text{-}78.1}$ and $vg^{21\text{-}197.5}$ appear to be identical and have the sequence 5′-TGT-TTG-3′ at the breakpoint. These alleles were isolated from different dysgenic fathers and may have resulted from either two independent events or two separate isolations of a preexisting variant in the $P[lacZ;ry^{+}]^{c49}$; *b cn* vg^{21}; *ry* stock. In support of the former, $vg^{21\text{-}81.1}$ and $vg^{21\text{-}7}$ (WILLIAMS *et al.* 1988), which were isolated under dissimilar experimental conditions separated by several years, share exactly the same deletion breakpoints and apparently only differ by the absence of the C at position 2901 in the $vg^{21\text{-}81.1}$ allele.

In the second screen a series of phenotypic revertant alleles was generated in which the scalloped to strapped wing phenotype of vg^{21} hemizygotes was reverted to wild type. Revertants represented approximately 3–5% of progeny. Two classes of internally deleted *P* elements were isolated (Table 2). Eleven alleles were isolated that had breakpoints in both inverted-repeats, and three alleles were isolated that had a breakpoint in the 5′ inverted-repeat and another elsewhere within the *P* element. The positions of the 5′ breakpoints range from the 13th to the 19th nucleotide. Most deletions (12 of 14) retain either 16 or 17 base pairs of the 5′ inverted-repeat. Of the 11 revertant alleles that also have 3′ breakpoints in the inverted-repeat, 11–18 bp are retained, and nine possess 16 ± 1 bp of the terminus. Deletions of the 5′ region of the *P* element result in a revertant phenotype as does deletion of both the 5′ and 3′ regions of the *P* element. The extreme alleles, by contrast, all retain at least approximately 100 bp of the 5′ region of the *P* element but are missing 3′ *P* element sequences.

Of the 14 revertant alleles seven possess filler sequences at the breakpoints. As with the strong *vestigial* alleles, these filler sequences are also AT rich. Most interesting is the inclusion of short tandem repeats of simple sequence in several of the mutants. For example, $vg^{21\text{-}IIIR1}$ and $vg^{21\text{-}IIIR5}$ have repeats of 5′-AACAT-3′, and $vg^{21\text{-}IR6}$ has repeats of the complement 5′-ATGTT-3′. Other filler sequences, although not composed of simple tandem repeats, do have a somewhat repetitive structure such as the filler sequence of $vg^{21\text{-}IR2}$ (5′-TAATATATGTATATATTATAAATTATA-3′). The sequences often appear to consist of tandem combinations of sequences present in the inverted-repeat. Exact assignment of the breakpoint in some alleles is impossible due to short regions of sequence identity (1, 2 or 4 bp) on either side of the breakpoint. For example, $vg^{21\text{-}IIIR8}$ could have retained 16 bp of the 5′ end or 16 at the 3′ end but the entire length is only 30 bp.

TABLE 3

Analysis of deletion breakpoints in derivatives of vg^{21} selected for altered PCR fragment length

Allele	5′ break	3′ break	Sequence of breakpoint junction	Phenotype
vg^{21}	380	2602 (306)	. . . acctagtaaa/aaattagaat . . .	Cryptic
$vg^{21\text{-}G3.11}$	15	2895 (13)	catgatgaaataaca/ttatttcatcatg	Revertant
$vg^{21\text{-}G4C6}$	16	2898 (10)	catgatgaaataacat/tttcatcatg	Revertant
$vg^{21\text{-}G5B1}$	16	2892 (16)	catgatgaaataac at gttatttcatcatg	Revertant
$vg^{21\text{-}G6B3}$	17	2891 (17)	catgatgaaataaca ta tgttatttcatcatg	Revertant
$vg^{21\text{-}G7B9}$	16	2895 (13)	catgatgaaataacat/ttatttcatcatg	Revertant

The deletion breakpoints are numbered according to the published sequence of the complete *P* element (O'HARE and RUBIN 1983), and the sequence of the junctions are presented as in Table 1 and Table 2.

Our observation that the majority of the deletion breakpoints are nested within the terminal inverted-repeats led us to question the existence of a bias inherent in the selection schemes. To examine this, derivatives of vg^{21} were isolated solely by alterations in the size of the *P* element. Five alleles were isolated that produced PCR-generated fragments of altered length (Table 3). All five alleles had breakpoints in both the 5′ and 3′ inverted-repeats and, therefore, resemble the majority of the alleles generated in our revertant screen. The 5′ breakpoints range from bases 15 to 17, and the 3′ breakpoints vary from 10 to 17 bp retained. All are phenotypic revertants. Although loss of only the 3′ region of the *P* element results in strong *vestigial* alleles and loss of 5′ *P*-element sequences leads to phenotypic reversion, the positioning of the breakpoint within the inverted-repeat at approximately the 16th bp from the end does not seem to result from a bias based upon phenotypic selection.

An examination of the distribution of the breakpoints that occurred within the inverted-repeat from our experiments is presented in Figure 1. We have also included data from the sequences of internally deleted *P* element derived from an insertion of pUC*hsneo* at position 3C (*Xneo3*) (TAKASU-ISHIKAWA *et al.* 1992). Analysis of the distribution of breakpoints demonstrates that a very high percentage of the disrupted *P* element termini (70% combined) have a breakpoint at 16 ± 1 bp from an end. The conservation of the terminal 16 bp of the *P* element suggests that these sequences are protected during the process that leads to terminal internal deletion. A very good candidate for this activity is the IRBP isolated by RIO and RUBIN (1988).

DISCUSSION

These experiments amount to an *in vivo* analysis of transposase-mediated deletions of the *P* element of vg^{21} (Figure 2). Deletions of the 3′ region of the *P* element, as small as the 177-bp deletion between positions 2714 and 2892 of $vg^{21\text{-}6.1}$, result in a strong *vestigial* phenotype. Deletions of the 5′ region of the *P* element, as well as deletions that remove everything but the very 5′ and 3′ termini, result in revertant phenotypes. The site where polyadenylation of the *P*-element transcript occurs (LASKI *et al.* 1986) is absent in all the extreme alleles generated. All of the phenotypic revertants have 5′ deletions (with a breakpoint at ~16 bp of the 5′ inverted-repeat), and all the strong *vestigial* alleles retain 5′ sequences to at least the 105th bp (*i.e.*, $vg^{21\text{-}1.3}$).

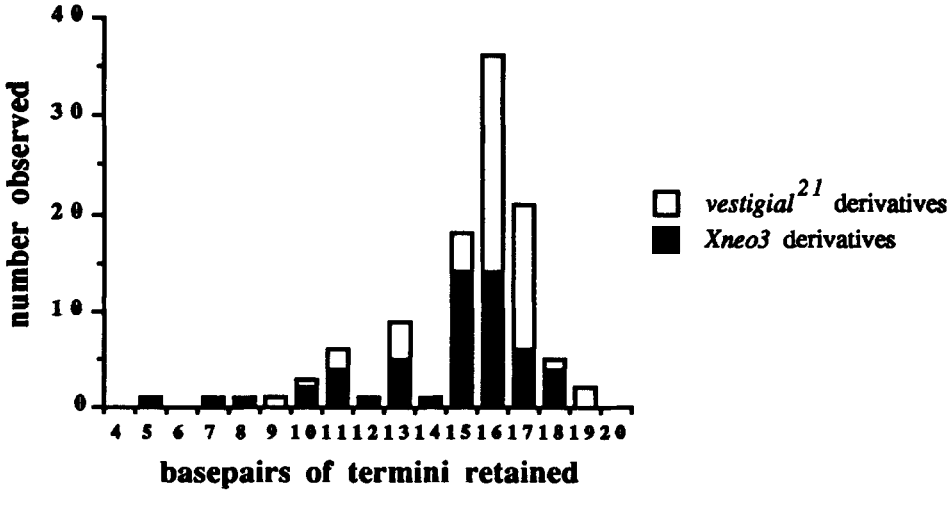

FIGURE 1.—Distribution of the breakpoint position in terminal internal deletions. The number of *P*-element terminal sequences (5–19 bp) of both $vestigial^{21}$ (this study) and *Xneo3* derivatives (TAKASU-ISHIKAWA *et al.* 1992) is presented. Note the strong tendency to retain 16 ± 1 bp of the terminus.

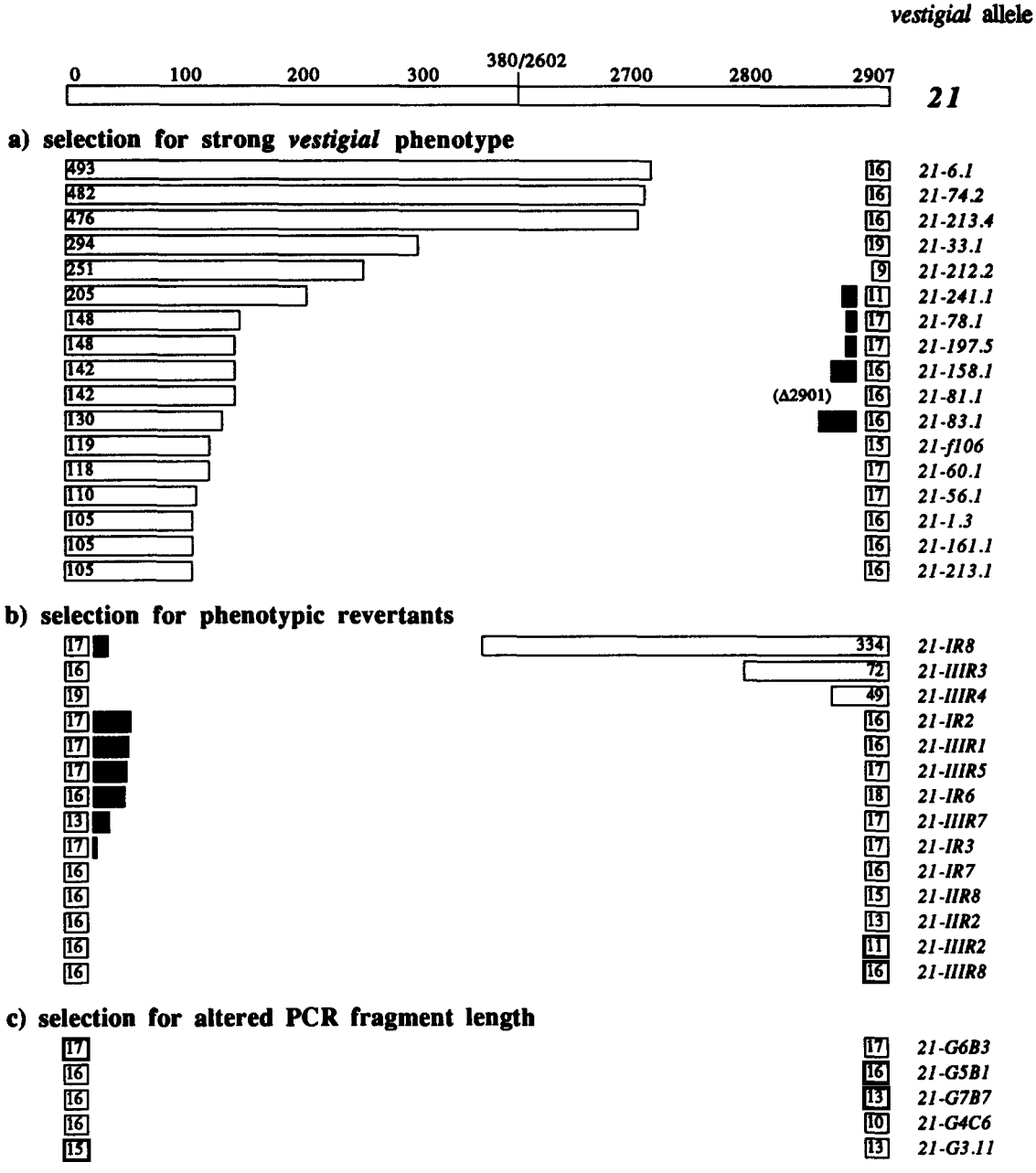

FIGURE 2.—Terminal internal deletion derivatives of *vestigial*^{21}. □ at the top represents the sequence of the *P* element located in the 5′ region of the transcription unit in the *vg*^{21} allele. The line in the centre of this box indicates the internal deletion breakpoint of this *P* element. □ below represent the *P*-element sequences present in the allele indicated at the right. The numbers within the open boxes correspond to the number of base pairs of *P*-element sequences, according to the published sequence (O'HARE and RUBIN 1983), remaining on that side of the breakpoint. The areas between the open boxes represents internally deleted *P*-element sequences. ■, filler sequences located at the deletion breakpoints.

The region between the 5′ inverted-repeat and position 105 contains both the promoter and the transcriptional start site of the *P*-element transcription unit (KARESS and RUBIN 1984). We suggest that transcription initiated at the *P* element promoter that is not terminated within the *P* element interferes with *vestigial* gene function. The termination of transcription within the *P* element, as in *vg*^{21}, only moderately effects *vestigial* expression and results in the cryptic phenotype. Removal of the *P* transcriptional start site seems to return *vestigial* expression to wild-type levels even if several hundred base pairs of the 3′ end of the *P* element, as in *vg*$^{21\text{-}IR8}$, remain at the insertion site.

The tendency for deletion derivatives of *P*-element alleles in *Drosophila melanogaster* to have an endpoint within the 31-bp terminal inverted-repeat has been reported previously (SEARLES *et al.* 1986; WILLIAMS *et al.* 1988; JOHNSON-SCHLITZ and ENGELS 1993) and is evi-

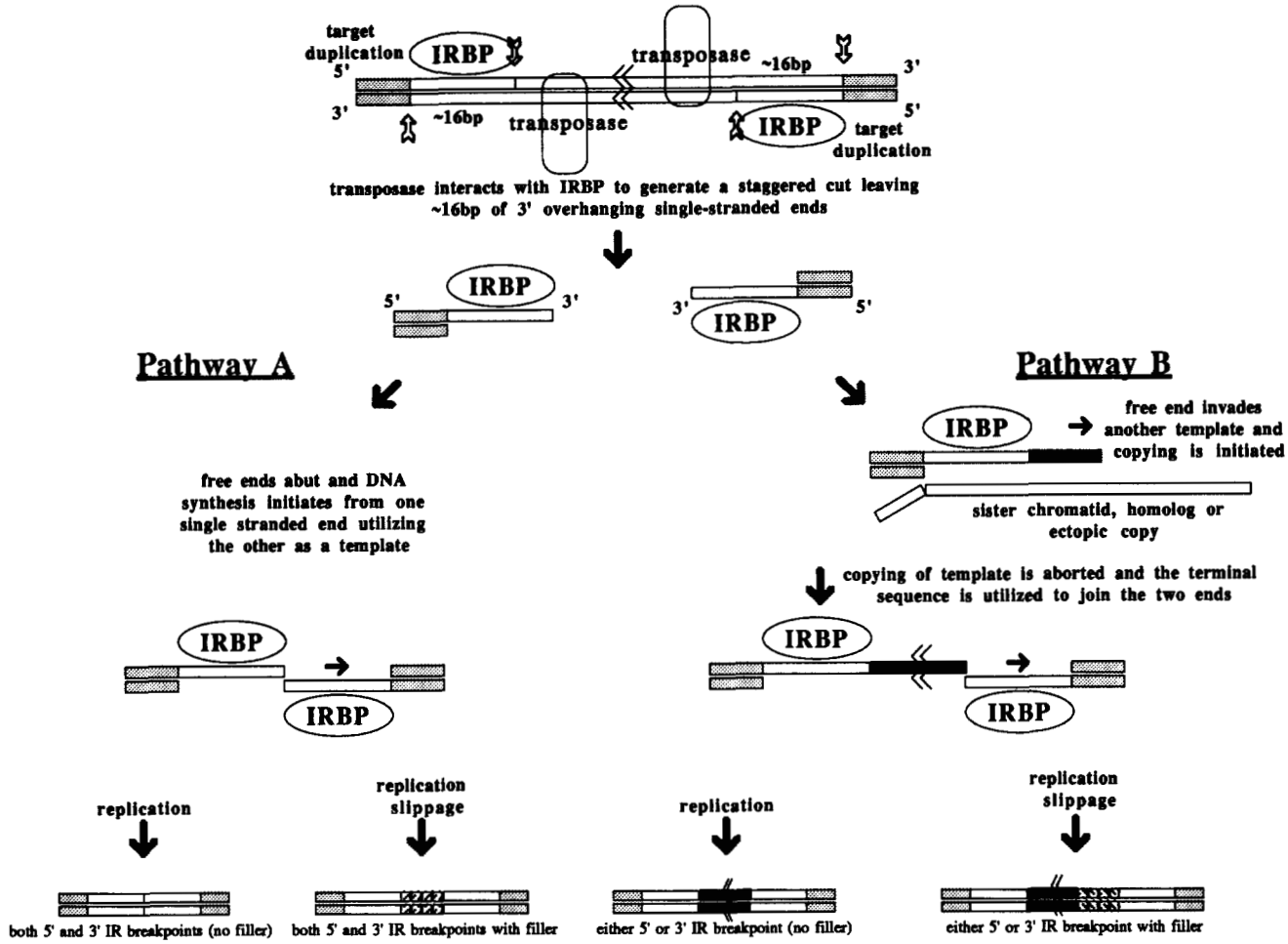

FIGURE 3.—Possible role for IRBP in transposase-induced double-strand gap repair. The sequence analysis of *P*-element terminal internal deletions suggest that the IRBP may maintain the terminal ~16 bp of the *P* element after excision of the *P* element at the donor site. The protected sequences can then be joined by a synthesis-dependent reaction by which DNA synthesis results in bridging the gap between free ends (pathway A). Invasion of another template (pathway B) before this results in either recovery of at least some of the internal *P*-element sequences (if an identical *P* element is copied) or of different *P*-element sequences, targeted transposition (if another *P* element is copied). Generation of filler sequences can be explained by rounds of replication and slippage during the synthesis-dependent joining of the ends. The diagrams of pathway B account for the 3′ *P*-element deletions of Figure 2a. Deletions of the 5′ *P*-element sequences would occur when the 3′ terminus invades another template.

dent in the extensive data set of TAKASU-ISHIKAWA *et al.* (1992). We wish to propose a model, a derivative of the SDSA model (NASSIF *et al.* 1994), to explain the formation of terminal internal deletions (Figure 3). The transposase, which binds at sites just inside the inverted-repeats (KAUFMAN *et al.* 1989), must cut at both ends of the transposon (5′-CATG-3′) to allow transposition. To satisfy the requirements of the cut-and-paste transposition model to generate target site duplications (SHAPIRO 1979; KAUFMAN and RIO 1992) and assuming that excision is an initial event in a procedure that may or may not be followed by integration, a mechanism that generates either 3′ overhanging or blunt ends must exist. A 16-nucleotide 3′ overhang could be specified by the binding of the IRBP through interaction with the transposase. A staggered cut of this length is much longer than the four-to-seven-nucleotide overhang that the recovery of plasmid borne excision products from injected embryos would seem to suggest (O'BROCHTA *et al.* 1991). However, *P* elements excised from plasmids transiently placed in the embryonic soma may not be subjected to the same constraints as *P* elements integrated into the germline. We propose that the IRBP has a role in the stabilization of the 16-nucleotide 3′ overhang and thus participates in repair of the double-strand break. In the case of the very smallest *P* elements recovered, those that retain only 16 or so bp of each terminus fused together, the simplest model to

explain their origin involves the two 3′ overhanging ends abutting or overlapping by a basepair or two, followed by DNA synthesis to cross the gap (Figure 3, Pathway A). The IRBP has the potential to mediate this activity by maintaining the single-stranded 3′ overhang thus preserving a template to be copied. Larger *P* elements that have a breakpoint in only one of the inverted-repeats seem to have had one end invade a homologous sequence (Figure 3, Pathway B). However, before the entire element is copied, DNA synthesis switches to copy the terminal sequences. The template switching often seems to occur in the 5′ regions of AT-rich sequences. Either end can initiate synthesis because we have isolated terminal internal deletions that break in either the 5′ or the 3′ repeat. This can also explain the recovery, at *white*, of truncated *yellow* sequences fused to 14 or 17 bp of the *P*-element terminus (NASSIF *et al.* 1994). In addition, the recovery of chimeric *P* elements in which flanking DNA sequences were recruited from a donor site (TSUBOTA and DANG-VU 1991; HESLIP *et al.* 1992) may be explained in a similar manner, except that DNA synthesis has apparently continued through the ectopic *P* element and into adjacent sequences (NASSIF *et al.* 1994) and finally resolved by switching to the inverted-repeat template.

An examination of the filler sequences included at the breakpoint reveals an interesting property. Short tandem repeats of simple sequence of 5′-AACAT-3′, 5′-ATGTT-3′ or 5′-TAACAT-3′ are present in some alleles. If we compare these sequences to the first 18 bp of the terminal inverted-repeat (5′-catgatgaaataacataa-3′), the repeats are identical to sequences at positions 10–18 (or its complement). This observation suggests that, at least in some cases of filler sequences found at the deletion breakpoint, a mechanism exists to copy at this junction short sequences identical to regions of the inverted-repeat. For example, the filler sequence from $vg^{21\text{-}78.1}$ is 5′-TGTTTGT-3′ and could represent bases 2893–2896 (5′-TGTT-3′) and 2893–2895 (5′-TGT-3′) joined in tandem. In the specific case of the $vg^{21\text{-}83.1}$ allele, the 5′-CATTAA-3′ repeat could have been generated either as described above or through a similar mechanism copying a sequence located between position 123 and 131 (5′-AACATTAAC-3′). A similar doublet 5′-ATCATTA-3′, found in the cloned *P* element λπ51 (π13B) (O'HARE and RUBIN 1983; O'HARE *et al.* 1992), could also have been copied from the adjacent 5′-ATCATTATC-3′ between positions 2114 and 2122. In both cases the alternate repeats are very similar and may have arisen under similar circumstances. To generate filler sequences, we favor a model based on the process of replication-slippage-replication originally proposed by O'HARE and RUBIN (1983) and supported by KURKULOS *et al.* (1994). This model is very different from the hairpin model proposed by TAKASU-ISHIKAWA *et al.* (1992). The generation of the short tandem repeats of simple sequence in $vg^{21\text{-}IIIR1}$, $vg^{21\text{-}IIIR5}$, $vg^{21\text{-}IR6}$ and $vg^{21\text{-}83.1}$ can readily be attributed to several rounds of replication, slippage and replication utilizing part of the inverted-repeat sequence (5′-catgatgaaataacataa-3′ or a similar sequence) or its complement as a template. Most of the other filler sequences are highly AT rich, and some of the less structured filler sequences could have arisen in a similar manner. The region of the inverted-repeat that appears to be utilized as template in these proposed rounds of replication and slippage correspond to the proximal sequences protected by the IRBP in the DNA protection assays (RIO and RUBIN 1988).

The IRBP is a host-derived protein and its normal activity is unclear. One correlation, which may have some biological significance, should be mentioned. In studies of *Drosophila melanogaster* satellite sequences, LOHE and BRUTLAG (1987) described stretches of short repetitive sequences that form a smooth transition from one repeat to another without losing the periodicity of the repeated structure. For example, a series of $(AAXAY)^n$ and $(AAXAZ)^m$ would show the junction as . . . AAXAY AAXAY AAXAZ AAXAZ . . . (X, Y and Z may represent A, C, G or T). They proposed a model of template amplification to explain the origin of smooth junctions between two different but similar satellite sequences. According to this model, during the synthesis of an array of satellite DNA repeats, each repeat is generated using the preceding unit as a template. Precise replication results in long homogeneous tandem arrays, whereas errors in replication would result in misreading of the repeat length or the introduction of a base change in the array followed by perpetuation of the error. This mechanism would produce the smooth junctions observed. The repeats observed in some of the filler sequences resemble satellite sequences; 5′-AACAT-3′ fits the general structure of most pentameric tandem repeats of 5′-AAXAY-3′ but not one previously described (LOHE *et al.* 1993). If the IRBP has a role in the generation of filler sequences, as we propose, then perhaps the protein is also involved in processes such as template amplification.

Recently, the IRBP has been identified as the Drosophila homologue (RIO, reported by PLASTERK 1993) of the 86-kDa subunit of the human Ku autoimmune antigen, a DNA-binding protein (YANEVA *et al.* 1989). If the two proteins share similar activities, it is possible that the Ku antigen may be also involved in the maintenance of some microsatellite sequences. The instability of particular microsatellites is believed to result in some genetic diseases including fragile *X* syndrome, Kennedy's disease, myotonic dystrophy, Huntington's disease and spinocerebellar ataxia type 1 (KUNKEL 1993). Such instability is believed to result from strand slippage during DNA synthesis.

The authors are indebted to GREG GLOOR and JOHN LOCKE for helpful discussions and their comments on the manuscript. Also, we gratefully acknowledge GREG GLOOR for suggesting the experiment that tested for a possible selection bias in the data presented. We thank BILL CLARK, GARY RITZEL, MICHAEL VEEMAN and TOVE REECE for technical support. B.E.S. was the recipient of a University of Alberta Graduate Teaching Assistantship. T.R.H. was the recipient of a Studentship from the Alberta Heritage Foundation for Medical Research. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada to R.B.H. and J.B.B.

LITERATURE CITED

ENGELS, W. R., 1989 *P* elements in *Drosophila melanogaster*, pp. 437–484 *Mobile DNA*, edited by D. E. BERG and M. M. HOWE. American Society for Microbiology, Washington DC.

ENGELS, W. R., D. M. JOHNSON-SCHLITZ, W. B. EGGLESTON and J. SVED, 1990 High-frequency *P* element loss in Drosophila is homolog dependent. Cell **62:** 515–525.

GLOOR, G. B., N. A. NASSIF, D. M. JOHNSON-SCHLITZ, C. R. PRESTON and W. R. ENGELS, 1991 Targeted gene replacement in Drosophila via *P* element-induced gap repair. Science **253:** 1110–1117.

GLOOR, G. B., C. R. PRESTON, D. M. JOHNSON-SCHLITZ, N. A. NASSIF, R. W. PHILLIS *et al.*, 1993 Type I repressors of *P* element mobility. Genetics **135:** 81–95.

HESLIP, T. R., and R. B. HODGETTS, 1994 Targeted transposition at the *vestigial* locus of *Drosophila melanogaster*. Genetics **138:** 1127–1135.

HESLIP, T. R., J. A. WILLIAMS, J. B. BELL and R. B. HODGETTS, 1992 A *P* element chimera containing captured genomic sequences was recovered at the *vestigial* locus in Drosophila following targeted transposition. Genetics **131:** 917–927.

JOHNSON-SCHLITZ, D. M., and W. R. ENGELS, 1993 *P*-element-induced interallelic gene conversion of insertions and deletions in *Drosophila* melanogaster. Mol. Cell. Biol. **13:** 7006–7018.

KARESS, P. D., and G. M. RUBIN, 1984 Analysis of *P* transposable element functions in *Drosophila*. Cell **38:** 135–146.

KAUFMAN, R. E., and D. C. RIO, 1992 *P* element transposition *in vitro* proceeds by a cut and paste mechanism and uses GTP as a cofactor. Cell **69:** 27–39.

KAUFMAN, R. E., R. F. DOLL and D. C. RIO, 1989 Drosophila *P* element transposase recognizes internal *P* element DNA sequences. Cell **59:** 359–371.

KUNKEL, T. A., 1993 Slippery DNA and diseases. Nature **365:** 207–208.

KURKULOS, M., J. M. WEINBERG, D. ROY and S. M. MOUNT, 1994 *P* element-mediated *in vivo* deletion analysis of *white-apricot:* deletions between direct repeats are strongly favored. Genetics **136:** 1001–1011.

LASKI, F. A., D. C. RIO and G. M. RUBIN, 1986 Tissue specificity of Drosophila *P* element transposition is regulated at the level of mRNA splicing. Cell **44:** 7–19.

LASKO, P. F., and M. L. PARDUE, 1988 Studies of the genetic organization of the *vestigal* microregion of *Drosophila melanogaster*. Genetics **120:** 495–502.

LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press, San Diego.

LOHE, A. R., and D. L. BRUTLAG, 1987 Adjacent satellite DNA segments in Drosophila. Structure of junctions. J. Mol. Biol. **194:** 171–179.

LOHE, A. R., A. J. HILLIKER and P. A. ROBERTS, 1993 Mapping simple repeated DNA sequences in heterochromatin of *Drosophila melanogaster*. Genetics **134:** 1149–1174.

MISRA, S., and D. C. RIO, 1990 Cytotype control of Drosophila *P* element transposition: the 66kD protein is a repressor of transposase activity. Cell **62:** 269–384.

MULLINS, M. C., D. C. RIO and G. M. RUBIN, 1989 Cis-acting DNA sequence requirements for *P* element transposition. Genes Dev. **3:** 729–738.

NASSIF, N., J. PENNEY, S. PAL, W. R. ENGELS and G. B. GLOOR, 1994 Efficient copying of nonhomologous sequences from ectopic site via P-element-induced gap repair. Mol. Cell. Biol. **14:** 1613–1625.

O'BROCHTA, D. A., S. P. GOMEZ and A. M. HANDLER, 1991 *P* element excision in *Drosophila melanogaster* and related drosophilids. Mol. Gen. Genet. **225:** 387–394.

O'HARE, K., and G. M. RUBIN, 1983 Structure of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell **34:** 25–35.

O'HARE, K., A. DRIVER, S. MCGRATH and D. M. JOHNSON-SCHLITZ, 1992 Distribution and structure of cloned *P* elements from the *Drosophila melanogaster P* strain π_2. Genet. Res. **60:** 33–41.

O'KANE, C. J., and W. J. GEHRING, 1987 Detection *in situ* of genomic regulatory elements in Drosophila. Proc. Natl. Acad. Sci. USA **84:** 9123–9127.

PLASTERK, R. H. A., 1993 Molecular mechanisms of transposition and its control. Cell **74:** 781–786.

RIO, D. C. 1990 Molecular mechanisms regulating Drosophila *P* element transposition. Annu. Rev. Genet. **24:** 543–578.

RIO, D. C., and G. M. RUBIN, 1988 Identification and purification of a *Drosophila* protein that binds to the terminal 31-base-pair inverted repeats of the *P* transposable element. Proc. Natl. Acad. Sci. USA **85:** 8929–8933.

ROBERTSON, H. M., C. R. PRESTON, R. W. PHILLIS, D. M. JOHNSON-SCHLITZ, W. K. BENZ *et al.*, 1988 A stable genetic source of P-element transposase in *Drosophila melanogaster*. Genetics **118:** 461–470.

SEARLES, L. L., A. L. GREENLEAF, W. E. KEMP and R. A. VOELKER, 1986 Sites of *P* element insertion and structures of *P* element deletions in the 5′ region of *Drosophila melanogaster RpII215*. Mol. Cell. Biol. **6:** 3312–3319.

SHAPIRO, J. A., 1979 Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proc. Natl. Acad. Sci. USA **76:** 1933–1937.

STAVELEY, B. E., R. B. HODGETTS, S. L. O'KEEFE and J. B. BELL, 1994 Targeting of an enhancer trap to *vestigial*. Dev. Biol. **165:** 290–293.

TAKASU-ISHIKAWA, E., M. YOSHIHARA and Y. HOTTA, 1992 Extra sequences found at *P* element excision sites in *Drosophila melanogaster*. Mol. Gen. Genet. **232:** 17–23.

TSUBOTA, S. I., and H. DANG-VU, 1991 Capture of flanking DNA by a *P* element in *Drosophila melanogaster*: creation of a transposable element. Proc. Natl. Acad. Sci. USA **88:** 693–697.

WILLIAMS, J. A., and J. B. BELL, 1988 Molecular organization of the *vestigial* region in *Drosophila melanogaster*. EMBO J. **7:** 1355–1363.

WILLIAMS, J. A., S. S. PAPPU and J. B. BELL, 1988 Molecular analysis of hybrid dygenesis-induced derivatives of a P-element allele at the *vg* locus. Mol. Cell. Biol. **8:** 1489–1497.

YANEVA, M., J. WEN, A. AYALA and R. COOK, 1989 cDNA-derived amino acid sequence of the 86-kDa subunit of the Ku antigen. J. Biol. Chem. **264:** 13407–13411.

Communicating editor: M. J. SIMMONS