Transgenesis upgrades for *Drosophila melanogaster*

Koen J. T. Venken¹ and Hugo J. Bellen¹,²,³,⁴

*Drosophila melanogaster* is a highly attractive model system for the study of numerous biological questions pertaining to development, genetics, cell biology, neuroscience and disease. Until recently, our ability to manipulate flies genetically relied heavily on the transposon-mediated integration of DNA into fly embryos. However, in recent years significant improvements have been made to the transgenic techniques available in this organism, particularly with respect to integrating DNA at specific sites in the genome. These new approaches will greatly facilitate the structure-function analyses of *Drosophila* genes, will enhance the ease and speed with which flies can be manipulated, and should advance our understanding of biological processes during normal development and disease.

**Introduction**

During the past few years, *Drosophila melanogaster* has gained in popularity because of the availability of its genome sequence (Adams et al., 2000), its rapid life cycle, the relative ease with which it can be handled and the multitude of genetic tools that are available for its study (Greenspan, 2004). The fly’s genome permits the most sophisticated manipulations of any of the known eukaryotes. Indeed, the number of existing and recently developed technological improvements, such as genome-wide transposon tagging and gene targeting (Venken and Bellen, 2005) and the availability of numerous resources, including online databases such as FlyBase and stocks from fly stock centers (see Box 1 for links to some of these resources) (Matthews et al., 2005), greatly facilitate research in the field and move it forward at a relentless pace. These technologies and resources further the study of various aspects of developmental biology, genetics, cell biology, neuroscience and behavior. Indeed, the identification of novel genes and their functional characterization in vivo greatly depends on these available tools. Moreover, as most human disease genes have a counterpart in the *Drosophila* genome, including those involved in genetic disorders and cancer (Bier, 2005; Vidal and Cagan, 2006), the fly is also becoming increasingly popular for studying the molecular mechanisms of human disease. Much of this research relies on an efficient and reliable transgenesis system.

Transgenesis in general can be defined as a group of technologies that allow DNA to be introduced into an organism of choice. The main goal of transgenesis is to integrate a foreign piece of DNA—a transgene—into an organism’s genome to result in germ line transmission (see Fig. 1), in order to study gene function. Insect transgenesis, in general, has been dominated by transposon-mediated integration (Handler and James, 2000). In *Drosophila*, transgenesis mainly relies on the *P* element transposon and this has been the foundation for most of the innovative developments within the fly field (Ryder and Russell, 2003). However, various improvements in fly transgenic techniques have been recently reported that predominantly employ the site-specific integration of transgenes at specific genomic docking sites (see glossary, Box 2) via the use of different recombinases and integrases (Groth et al., 2004; OBERSTEIN et al., 2005; Horn and Handler, 2005; Bateman et al., 2006; Venken et al., 2006; Bischof et al., 2007). Many of these advances have their origins in mouse molecular genetics (Seibler and Bode, 1997; BETHKE and SAUER, 1997; BOUHASSIRA et al., 1997; GROTH et al., 2000; THYAGARAJAN et al., 2001) and have been very useful for developing fly transgenic techniques, as discussed below.

Here, we summarize many of the current methods that are used to generate transgenic flies. We first review classical transposon-mediated transgenesis and site-specific integration methods, before describing a plethora of recent improvements that have their basis in site-specific integration systems.

**Transposon-mediated transgenesis in *Drosophila***

Transgenesis can be performed through various techniques. In *Drosophila*, transgenesis mainly relies on the *P* element transposon, the introduction of which (Rubin and Spradling, 1982) has been one of the most important breakthroughs in germ line transgenesis in *Drosophila*. As such, *Drosophila* research has been highly dependent on *P* element-mediated transgenesis, even though it has two major drawbacks: the size of the DNA that can be integrated is limited and the insertion sites cannot be controlled.

*P* elements are transposable elements, or transposons, which were originally identified within the fly’s own genome (Castro and Carareto, 2004). *P* elements, like other transposons, contain two terminal repeats, including inverted repeat sequences and other internally located sequence motifs absolutely required for their integration (Handler and James, 2000). However, various improvements in fly transgenic techniques have been recently reported that predominantly employ the site-specific integration of transgenes at specific genomic docking sites (see glossary, Box 2) via the use of different recombinases and integrases (Groth et al., 2004; OBERSTEIN et al., 2005; Horn and Handler, 2005; Bateman et al., 2006; Venken et al., 2006; Bischof et al., 2007). Many of these advances have their origins in mouse molecular genetics (Seibler and Bode, 1997; BETHKE and SAUER, 1997; BOUHASSIRA et al., 1997; GROTH et al., 2000; THYAGARAJAN et al., 2001) and have been very useful for developing fly transgenic techniques, as discussed below.

Here, we summarize many of the current methods that are used to generate transgenic flies. We first review classical transposon-mediated transgenesis and site-specific integration methods, before describing a plethora of recent improvements that have their basis in site-specific integration systems.

**Box 1. Relevant websites**

*Drosophila* Genomics Resource Center: plasmid resource center for fly transgenesis.
http://dgrc.cgb.indiana.edu

DrosDel: docking site stock center for FLP remobilization.
http://www.drosdel.org.uk

FlyBase: general online fly resource.
http://www.flybase.org

FlyC31: *Φ*C31 integrase system, plasmids and fly stocks.
http://www.frontiers-in-genetics.org/flyc31

P(acman): recombinering and the *Φ*C31 integrase system.
http://flypush.imgen.bcm.tmc.edu/lab/pacman.html

ΦC31 fly stocks at the Bloomington *Drosophila* Stock Center:
http://flystocks.bio.indiana.edu/Browse/misc-browse/phiC31.htm

ΦC31 RMCE website: *Φ*C31 integrase-mediated RMCE, plasmids and fly stocks.
http://genepath.med.harvard.edu/WuLab/RMCE

Recombineering website: resource for public available recombinering reagents.
http://recombineering.ncifcrf.gov

Gensat Database: resource for RecA assisted modification.
http://www.gensat.org

Vienna *Drosophila* RNAi Center: transgenic RNAi fly lines.
http://www.vdrc.at

National Institute of Genetics (Japan) RNAi Fly Stocks: transgenic RNAi fly lines.
http://www.shigen.nig.ac.jp/fly/nigfly

¹Program in Developmental Biology, ²Department of Molecular and Human Genetics, ³Department of Neuroscience and ⁴Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX 77030, USA.

e-mails: kv134369@bcm.tmc.edu; hbellen@bcm.tmc.edu

mobilization or transposition (see Fig. 2 and Table 1) (Beall and Rio, 1997). Mobile or autonomous P element transposons encode a functional enzymatic protein called P transposase that catalyzes transposition through both terminal repeats of the transposon. P element-mediated transgenesis requires the separation of the P transposition through both terminal repeats of the transposon. A plasmid that encodes P transposase, a so-called helper plasmid, is provided in trans with another plasmid (the transgene) that contains the transposon backbone, the sequence of interest and a marker (see Fig. 2B) (Karess and Rubin, 1984). In vitro synthesized mRNA that encodes the transposase or purified transposase protein itself (Kaufman and Rio, 1991) can also be co-injected with modified P elements. Co-injections limit transposase activity, which is often advantageous. Alternatively, the transposase can be expressed from a genomic source (Cooley et al., 1988), allowing the injection of a P element without a helper plasmid. Transgene expression can be rendered constitutive or inducible through the inclusion of a heat-shock promoter. A hyperactive form of P transposase has been isolated that results in increased transposition rates (Beall et al., 2002). In general, transposons are injected into fly strains that are devoid of the same transposon, avoiding unwanted mobilization events of transposons present in the genome, thereby ensuring the stable integration and maintenance of the injected transgene. The unbiased identification of integration events is crucial for transgenesis and predominantly relies on the incorporation of dominant markers, which are identified through screening or selection (see Table 2 and Box 3 for more information).

Transposition occurs by the excision or replication of the transposon from the injected plasmid and its insertion into the host genome. Different transposons have unique insertion site characteristics. Integration events of P elements are strongly biased towards the 5′ end of genes. Hot spots – insertion sites that attract P elements at a much higher frequency than others – also exist within the Drosophila genome (Spradling et al., 1995; Bellen et al.,

![Fig. 1. Drosophila transgenesis. White transgene DNA (red) is injected into generation zero Drosophila embryos (G0) of less than 1 hour old, which have been obtained from a parental (P) generation. The early developmental stages of Drosophila embryos are characterized by rapid nuclear divisions that occur without accompanying cell divisions, creating a syncytium. Prior to cellularization, pole cells (black) bud off at the posterior end. For germ line transmission to occur, the transgenic DNA must be taken up into the pole cells that are fated to become germ cells. Transgenic DNA integrated into a pole cell (red pole cell) can be transmitted from one generation (G0) to the next (G1 progeny). The resulting integration events are identified using an appropriate marker, such as as white+. When used in a mutant white− strain, this transgene marks transgenic flies by giving them a darker eye color (see Table 2 and Box 3 for more information on the markers used in fly transgenesis).](image-url)

**Box 2. Glossary of specialized terms**

**Acceptor site:** A genomic site that receives in vivo DNA mobilized from a different location – the donor site. This occurs through FLP remobilization or P element replacement.

**Docking site:** Alternatively called a landing site. A genomic site that receives injected DNA during embryo microinjection.

**Donor site:** A genomic site that contains DNA sequence that will be donated for integration at another location, through FLP remobilization, P element replacement or gene targeting.

**Episomal fragment:** An independent DNA element, such as a plasmid, that can exist extrachromosomally or that can be maintained by integrating into the genome of the host.

**Gal4/UAS system:** Based on the yeast transcriptional activator GAL4 and its high-affinity binding site, the upstream activating sequence (UAS), this system is generally used to ectopically express a gene of interest. When a tissue-specific GAL4 line is crossed to an effector line that carries the UAS fused to a gene of interest, progeny with both the GAL4 and UAS components express the gene of interest in an activator (and often tissue)-specific manner.

**Insulator:** A DNA sequence that blocks the interaction between cis-acting regulatory elements. These sites are sometimes used to protect transgenes from genomic position effects.

**MARC3:** MARCM (mosaic analysis with a represessible cell marker) allows mutant clones generated by mitotic recombination to be identified in an otherwise wild-type unlabeled background (Lee and Luo, 1999). A recent modification of the MARCM system, dual-expression-control MARCM, has added another level of sophistication to this technique (La and Lee, 2006).

**Mitotic recombination:** A cross-over between two homologous double-stranded DNA molecules. This recombination occurs frequently during meiosis, but is relatively rare during mitosis.

**Position effects:** The effect of the local chromosomal environment on the level or pattern of transgene expression, owing to local chromatin configuration or nearby cis-acting regulatory elements.

**Position effect variegation:** A phenomenon discovered in Drosophila that occurs when genes placed close to large heterochromatic regions are repressed. This repression is metastable in that the silenced state can be occasionally released, giving rise to derepressed cells and a variegated phenotype.

**Rescue:** A condition achieved by introducing a wild-type DNA fragment that can complement a genomic mutation by producing the functional or missing protein.
Insertion site duplication because elements do not remobilize efficiently (Lozovsky et al., 2002), and al., 2003; Thibault et al., 2004; Metaxakis et al., 2005). As insertional specificity to Drosophila hydei (Franz and Savakis, 1991) and (hAT) family members Tam3 Drosophila melanogaster, respectively (McGinnis et al., 1983). These include piggyBac (Handler and Harrell, 1999), identified in the cabbage looper moth Trichoplusia ni (Cary et al., 1989; Smith et al., 1993), isolated from the cabbage looper moth Trichoplusia ni (Loukeris et al., 1995) and Mariner (Lidholm et al., 1993), isolated from Drosophila hydei (Fritz and Savakis, 1991) and Drosophila mauritiana, respectively (Jacobson et al., 1986); and the hobo, Ac, Tam3 (hAT) family members Hermes (O’Brochta et al., 1996) and hobo (Blackman et al., 1989; Smith et al., 1993), isolated from the house fly Musca domestica (Handler et al., 1999) and Drosophila melanogaster, respectively (McGinnis et al., 1983). These transposons function in a variety of organisms, but their use in Drosophila transgenesis has been limited (O’Brochta and Atkinson, 1996; Ryder and Russell, 2003). piggyBac and Minos have been used as alternative mutagens because they have a different insertional specificity to P elements (Hacker et al., 2003; Horn et al., 2003; Thibault et al., 2004; Metaxakis et al., 2005). As Mariner elements do not remobilize efficiently (Lozovsky et al., 2002), and because hobo is present in most laboratory stocks, neither is commonly used. Finally, hobo and Hermes have been shown to cross-mobilize (Sundararajan et al., 1999). These features have limited the use of these transposable elements.

Applications of transposon-mediated transgenesis

The transposon-mediated integration of transgenes has been used for numerous experiments in the fly field. These experiments can be broadly subdivided into two main groups: gene disruption methods and transgenic technologies. Gene disruption occurs when a transposon insertion interferes with the function of a gene. Transgenic technologies usually involve introducing the different components of novel techniques (see below) or performing rescue experiments.

Almost all technological progress in flies depends on our ability to transform them. Indeed, P element-mediated enhancer detection (O’Kane and Gehring, 1987; Bellen et al., 1989; Bier et al., 1989), the use of the FLP/FRT system to create mutant clones by inducing mitotic recombination (see glossary, Box 2) (Xu and Rubin, 1993), the gene knockout methods in flies (Rong and Golic, 2000; Gogol and Golic, 2003), the creation of molecularly defined deletions throughout the genome (Thibault et al., 2004; Ryder et al., 2004), the generation of marked mutant clones by MARCM (see glossary, Box 2) (Lee and Luo, 1999), and many other technological advances have relied on transgenesis. The recent availability of a genomewide library of RNAi transgenic insertions that allows the knockdown of most fly genes (Dietzl et al., 2007) will also provide an invaluable tool to study gene function.

In addition, transposon-mediated phenotypic rescue of a mutation is considered to be the best and most convincing evidence that a piece of DNA contains a gene of interest. Unfortunately, traditional high-copy-number plasmids, including the P element-containing plasmids, have a limited cargo capacity of ~20-25 kb of DNA owing to plasmid instability in bacteria. To circumvent this, P elements were engineered in a medium-copy-number cosmid backbone (Haenlin et al., 1985; Steller and Pirrotta, 1985), providing a higher cargo capacity of up to 40-50 kb. Unfortunately, the difficulties associated with obtaining integration of 30-50 kb P element-based cosmids did not promote the use of this methodology. As a result, a transgenic cDNA rescue based on the GAL4/UAS system (see glossary, Box 2) (Fischer et al., 1988; Brand and Perrimon, 1993) or heat-shock induction (Basler and Hafen, 1989) became more popular.

Neutralization of position effects

One of the major drawbacks of P element-mediated transgenesis is that P elements most often integrate into the 5’ regulatory regions of genes (Bellen et al., 2004), thereby causing two unwanted consequences. First, the insertion often disrupts another gene that may or may not be relevant (e.g. within the same pathway) to the gene that is being studied (Norga et al., 2003). Second, the gene within the transposon may be subject to unwanted position effects or to position effect variegation (see glossary, Box 2) dictated by the surrounding

### Table 1. Transposons for Drosophila transgenesis

<table>
<thead>
<tr>
<th>Transposon</th>
<th>Inverted repeats (bp)</th>
<th>Insertion site preference</th>
<th>Target site duplication (bp)</th>
<th>Species compatibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>P element</td>
<td>31</td>
<td>5’ end of genes</td>
<td>8</td>
<td>Drosophilidae only</td>
</tr>
<tr>
<td>piggyBac</td>
<td>13</td>
<td>TA</td>
<td>4</td>
<td>Broad</td>
</tr>
<tr>
<td>Minos</td>
<td>255</td>
<td>TAA</td>
<td>2</td>
<td>Broad</td>
</tr>
<tr>
<td>Mariner</td>
<td>28</td>
<td>TA</td>
<td>2</td>
<td>Broad</td>
</tr>
<tr>
<td>Hermes</td>
<td>17</td>
<td>Low sequence specificity</td>
<td>8</td>
<td>Broad</td>
</tr>
<tr>
<td>hobo</td>
<td>12</td>
<td>Low sequence specificity</td>
<td>8</td>
<td>Broad</td>
</tr>
</tbody>
</table>
Table 2. Dominant marker genes for *Drosophila* transgenesis

<table>
<thead>
<tr>
<th>Dominant marker*</th>
<th>Screening/selection (compound)</th>
<th>Mutant line required</th>
</tr>
</thead>
<tbody>
<tr>
<td>white</td>
<td>Screening</td>
<td>Yes</td>
</tr>
<tr>
<td>yellow</td>
<td>Screening</td>
<td>Yes</td>
</tr>
<tr>
<td>rosy</td>
<td>Screening</td>
<td>Yes</td>
</tr>
<tr>
<td>rough</td>
<td>Screening</td>
<td>Yes</td>
</tr>
<tr>
<td>vermillion</td>
<td>Screening</td>
<td>Yes</td>
</tr>
<tr>
<td>3xP3 ‘fluorescent protein’</td>
<td>Screening</td>
<td>No</td>
</tr>
<tr>
<td>hs-neo</td>
<td>Selection (G418)</td>
<td>No</td>
</tr>
<tr>
<td>hs-opd</td>
<td>Selection (paraoxon)</td>
<td>No</td>
</tr>
<tr>
<td>Resistant to deldrin (Rdl)</td>
<td>Selection (deldrin)</td>
<td>No</td>
</tr>
<tr>
<td>Alcohol dehydrogenase (Adh)</td>
<td>Selection (ethanol)</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*white, rosy, rough and vermillion are eye color markers, whereas yellow is a body color marker. hs-neo encodes a heat shock–inducible neomycin-selectable marker (Steller and Pirotta, 1985), hs-opd encodes a heat shock–inducible insecticide-degrading enzyme (Benedict et al., 1995), Resistant to deldrin (Rdl) encodes the GABA-A receptor (Stilwell et al., 1995) and Alcohol dehydrogenase (Adh) encodes an enzyme involved in ethanol catabolism (Goldberg et al., 1983).

Position effects can be partially neutralized through the incorporation of insulator sequences (Roseman et al., 1995). Insulators (see glossary, Box 2) tend to shield the transgene from regulatory influences imposed by the surrounding genome. Insulators, such as gypsy, have been used in some *P* element vectors because they are more mutagenic than other *P* elements that do not contain insulators (Roseman et al., 1995). They were also incorporated into *P* element reporter transposons developed to analyze gene regulatory sequences (Barolo et al., 2000; Barolo et al., 2004). Insulators allow for a better comparison of different transgene insertions at different loci. Yet insulators may also influence the expression of the gene that they flank within the construct and are still somewhat subject to position effects in the genome.

There are at least four alternative genetic strategies to neutralize position effects when different transgenes are being compared at the same locus. The simplest method is transgene coexpression (Siegel and Hartl, 1996), which allows any two transgenes, such as a rescue fragment and its mutant version, to be compared in the same orientation at the same locus (Fig. 3). Both transgenes are integrated into a *P* element that contains the site-specific recognition sites *FRT* and *loxP*, the targets of FLP and Cre recombinases, respectively (see Box 4 for more information on these recombinases). After integration of the *P* element, FLP can remove one transgene and Cre can remove the other. Recognition sites are oriented such that either recombination event results in an identical configuration for either transgene. This method also introduced the use of Cre recombinase into the *Drosophila* field (Siegal and Hartl, 1996; Siegal and Hartl, 2000). One drawback of the technique is that only two transgenes can be compared at the same locus.

A second method is based on FLP recombinase-mediated transgene remodeling (Golic et al., 1997) (Fig. 4A). First, a ‘donor’ *P* element (see glossary, Box 2), containing a transgene together with the *white* marker flanked by *FRT* sites, is integrated into the fly genome using *P* transposition, resulting in a donor site. Second, the transgene with the *white* marker, flanked by *FRT* sites, is remobilized through FLP excision. This episomal fragment (see glossary, Box 2) can integrate into a second single *FRT*-containing ‘acceptor’ transposon (see glossary, Box 2), which also carries another dominant marker and is located elsewhere in the genome. Successful mobilization events can be identified through screening, as relocalization usually results in changes in *white* marker expression owing to position effects. This strategy is facilitated if a split *white* marker strategy is integrated into the system (Fig. 4B): the *white* marker is separated into 5' and 3' fragments, and only becomes functional after the reconstitution of these fragments through site-specific recombination within an intron located between both fragments (Golic et al., 1997). In FLP recombinase-mediated transgene remodeling, *white* expression is only obtained after correct mobilization and site-specific integration (Fig. 4C), facilitating the screening procedure of integration events. Interestingly, thousands of *P* element insertions obtained by the...
DrosDel project (see Box 1) were generated by the mobilization of the P[RS5] and P[RS3] transposons (Golic and Golic, 1996) and were subsequently used for the generation of precise deletions (Ryder et al., 2004). Both transposons can be used as acceptor elements for in vivo FLP-mediated DNA mobilization using the split white+ marker strategy, and they provide numerous docking sites that are dispersed all over the fly genome (Fig. 4D,E). A drawback of FLP recombinase-mediated transgene remobilization in general is that a second round of crossings for remobilization and screening has to be performed after an initial P element-mediated transformation to obtain the required integration events of donor elements.

A third genetic trick to neutralize position effects is P element replacement or targeted transposition (Gloor et al., 1991; Lankenau and Gloor, 1998). An ‘acceptor’ P element inserted at one location is replaced by a second ‘donor’ P element, integrated at another location, through in vivo gap repair (Fig. 5). The technique requires homologous recombination between the 10-20 bp footprints of the 31 bp inverted terminal repeat that remain after excision of the acceptor element and the homologous counterpart of the donor element. The homologous recombination event is promoted owing to a double-stranded gap that is generated after the excision of the acceptor element. Various donor P elements can be targeted to the same locus, allowing different transgenes to be directly compared with each other. A drawback of the technique is that replacement can occur in both directions, requiring additional molecular verification. The technique has not been used to perform structure-function analysis of differently mutagenized transgenes but has been proven useful for converting existing lacZ enhancer-detector

**Fig. 3. Transgene compleition.** Two inserts (1 and 2, red), each containing a cloned fragment, such as a genomic rescue fragment, are integrated into a P element that contains appropriately positioned loxP (yellow) and FRT (pink) sites. Cre recombination results in the removal of insert 2, whereas FLP recombination results in the removal of insert 1, positioning either insert in the same orientation at the same locus (indicated in blue), thereby neutralizing position effects. In each case, recombination events are identified by the loss of a dominant marker (green).

**Fig. 4. FLP remobilization.** (A) FLP remobilization technique. A donor transposon contains a transgenic insert (red) together with a marker (1) flanked by two FRT sites. An acceptor transposon, at a desired locus, contains a second marker (2) and one FRT site. Remobilization of the donor transposon by FLP results in the excision of its transgene and its potential integration into the FRT site of the acceptor transposon. This remobilization can be followed through changes in expression of marker 1, such as white+, that occur because of changes in position effects (from yellow in the original site to orange in the acceptor site). Different donor transposons, each containing different transgenes, can be targeted to the same acceptor, thereby neutralizing position effects. (B) Split white+ marker strategy. The white+ marker is divided into two parts: 5'-white+ (5’) and 3’-white+ (3’). Neither part can produce eye pigmentation alone (indicated in gray). Recombination between appropriately localized recombination sites, FRT in this case, results in white+ reconstitution and its expression (orange). (C) Integration of the split white+ marker strategy into the FLP remobilization technique. The correct remobilization and integration of the transgene (red) are identified by white+ reconstitution (orange). Marker 2 (yellow) identifies donor transgenes. (D,E) DrosDel elements P[RS5] and P[RS3]. FLP-mediated recombination at (D) P[RS5] and (E) P[RS3] results in chromosomal remnants, P[RS5r] and P[RS3r], respectively. Each contains one part of the white+ marker. Both remnants can be reconstituted through FLP remobilization of an appropriately designed donor transposon (see C).
Box 4. Site-specific recombinases and integrases

Site-specific recombinases and integrases (SSRIs) often require only two components: a site-specific enzyme, which, preferentially, functions without additional proteins, and a pair of DNA recombination sites (RSSs) (Sorrell and Kolb, 2005). SSRIs are subdivided into the tyrosine and serine recombinases, which use a conserved tyrosine and serine residue during recombination, respectively. Commonly used tyrosine recombinases are and Cre and FLP. Cre (causes recombination of the bacteriophage P1 genome) recognizes minimal loxP [ locus of crossing-over (X)] in P1] RSSs of 34 bp, which consist of two 13 bp perfect inverted repeats (red) flanking an 8 bp asymmetric spacer (black) that confers directionality (Hoess et al., 1982) (see figure). Likewise, FLP (flips DNA) recognizes minimal FRT (FLP recombinase target) RSSs of 34 bp and has a similar configuration to loxP but with a different sequence (McLeod et al., 1986) (see figure). A genuine FRT RSS, absolutely required for site-specific integration, consists of 48 bp, containing an additional isolated base pair and a third 13 bp direct repeat (see figure).

A commonly used serine recombinase is the integrase from the Streptomyces bacteriophage ɸC31 (see figure) (Thorpe and Smith, 1998). ɸC31 integrase recognizes a minimal high-efficiency attP RSS (attachment site in the phage genome) of 39 bp and a minimal high-efficiency attB RSS (attachment site in the bacterial genome) of 34 bp (Groth et al., 2000). attP and attB sites contain imperfect inverted repeats (red) flanking a short recombination core (TTG, black) that provides directionality (see figure).

Recombination between two RSSs can lead to an inversion, integration/excision or recombinase-mediated cassette exchange (RMCE), depending on the orientation and types of RSS (see figure). A translocation can also occur if the RSSs are on different chromosomes (not shown). The presence of two compatible RSSs results in a recombination event, which in the case of FRT or loxP leads to the reformation of a still functional RSS, potentially resulting in additional recombination events. This problem can be overcome using RSS inverted repeat variants, such as lox71 and lox66, which contain mutations in the left and right inverted repeat, respectively. Recombination between lox71 and lox66 results in wild-type attP and a double-mutant lox72, two sites that do not recombine with each other (Albert et al., 1995). Similar variants exist for FRT sites (Senecco et al., 1988). In the case of attPattB, new attL (att Left) and attR (att Right) sites are created (see figure), which are no longer substrates for the integrase, ensuring that recombination is irreversible.

Integration using a single RSS results in the integration of the entire plasmid, including the vector backbone. This can be avoided through RMCE (Schlake and Bode, 1994; Baer and Bode, 2001). RMCE uses a double-reciprocal cross-over reaction between two cassettes, one integrated into the genome, the other episomal (circle in figure, see Box 2), to mediate transgene integration, while avoiding the integration of vector backbone (see figure). Because the use of two sets of loxP or FRT RSSs favors deletion over RMCE, double reciprocal cross-over requires the use of RSS spacer variants: recombination sites that react preferentially with each other but not with other variants. RSS spacer variants include lox5171 (Hoess et al., 1986), lox5171 and lox2272 (Lee and Saito, 1998); and m2, m3, m7 and m11 (Langer et al., 2002). lox2272 and m2 seem to be the least leaky and therefore most useful for RMCE in conjunction with a wild-type loxP RS. An expansion of the loxP collection has been recently reported (Missirlis et al., 2006). Similar spacer variants also exist for FRT. F3 and F5 are useful for RMCE reactions together with a wild-type FRT site (Schlake and Bode, 1994). Alternatively, inverted RSS eliminate deletion but cause cassette inversion when FRT or loxP are used (Feng et al., 1999) (see figure). Because integration can occur in two orientations, additional screening is required to determine the exact nature of the integration event.

P elements into GAL4 drivers (Sepp and Auld, 1999; de Navas et al., 2006) that allow more versatile reporter analysis using the GAL4/UAS system (Fischer et al., 1988; Brand and Perrimon, 1993).

The best but most labor-intensive way to eliminate position effects is in vivo gene targeting through homologous recombination. Gene targeting in Drosophila can be performed using two strategies: ‘ends-in’ or insertional gene targeting (Rong and Golic, 2000) and ‘ends-out’ or replacement gene targeting (Gong and Golic, 2003) (Fig. 6). Insertional gene targeting results in the insertion of the entire targeting sequence into the region of homology. This results in a duplication that can be resolved during a second round of homologous recombination (Fig. 6A) (Rong et al., 2002). Replacement gene targeting results in the substitution of an endogenous DNA sequence with exogenous DNA through a double-reciprocal recombination event between two stretches of homologous sequence (Fig. 6B). Both strategies require the introduction of a ‘donor’ element, which contains the gene-targeting cassette, through transgenesis prior to in vivo homologous recombination, and require extensive screening. Although the techniques have not been used to compare the phenotypic outcome of different transgenes at the same locus, they are gaining in popularity for creating targeted mutations (O’Keefe et al., 2007).

Recent efforts have focused on making gene targeting more efficient in Drosophila through the use of site-specific zinc-finger-nuclease-stimulated gene targeting (Bibikova et al., 2003; Beumer et al., 2006). Zinc-finger nucleases are protein fusions between the FokI nuclease and (generally) three zinc-finger DNA-binding domains that introduce sequence specificity. Because each zinc finger recognizes 3 bp, zinc-finger nucleases can be designed to bind to a unique segment of 9 bp. As these nucleases need to dimerize at the target site before they can cut the target DNA, a recognition site of 18 bp is effectively required, a sequence that is
likely to be unique in the fly genome. Thus, cutting by zinc-finger nucleases can be directed to specific target sites to create a double-stranded break, resulting in increased gene-targeting efficiency when a linearized donor targeting element is introduced.

**Site-specific transgenesis for Drosophila**

Although all the strategies of site-specific integration described above are elegant and useful, they have not been used extensively. The main drawbacks are that they allow only a limited number of transgenes to be compared and are too labor-intensive, as they require transgenesis of a donor construct prior to extensive genetic screening to obtain the required site-specific transgenic insertion event. Hence, the primary goal of true-targeted transgenesis is to achieve efficient site-specific integration upon injection of the DNA without the need for further manipulations.

This strategy was pioneered in the fly field using the bacteriophage \(\Phi C31\) integrase, which can integrate transgenic constructs at defined docking sites (Groth et al., 2004). Moreover, \(\Phi C31\) integrase-mediated transgenesis allows large DNA fragments to be integrated into the fly genome, well beyond the fragment sizes that can be introduced by \(P\) element-mediated integration (Venken et al., 2006). As discussed in more detail below, this approach has also introduced a user-friendly DNA modification platform, called recombinase-mediated cassette exchange (RMCE) (Baer and Bode, 2001).

**Fig. 5.** \(P\) element replacement. (A) Two \(P\) elements, an acceptor element, containing marker 1 (orange), and a donor element, containing transgenic insert (red) and marker 2 (yellow), are brought together. (B) In the presence of \(P\) transposase, the acceptor element might excise. (C) This excision might promote double-stranded gap repair through homologous recombination between the 10-20 bp footprints (pink) of the 31 bp inverted terminal repeats at the acceptor site (blue) and the similar sequence at the donor site (green). (D) This results in the integration of the donor element into the acceptor locus.

It has been shown that the \(\Phi C31\) integrase can catalyze the site-specific integration of \(attB\)-containing plasmids into so-called \(attP\)-containing ‘docking’ or ‘landing’ sites that have been introduced into mammalian cell lines (Groth et al., 2000; Thyagarajan et al., 2001). Interestingly, \(attB\)-containing plasmids integrate more readily into \(attP\)-containing genomic docking sites than do \(attP\) sites in the reciprocal reaction, indicating that the integration reaction is asymmetric in nature (Thyagarajan et al., 2001; Belteki et al., 2003). This phenomenon was recently confirmed in Drosophila (Nimmo et al., 2006).

In Drosophila, recombination is mediated via \(\Phi C31\) integrase, provided through an mRNA source, between an \(attP\) docking site, previously integrated with a transposon into the fly genome, and an \(attB\) site present in an injected plasmid (Groth et al., 2004) (Fig. 7A). Three so-called pseudo-\(attP\) docking sites have been identified within the Drosophila genome. As one of these pseudo-sites is located in the endogenous transposable element copia, the true number of available pseudo-sites is likely to be high (Kaminker et al., 2002). Fortunately, these pseudo-sites were shown not to be receptive to \(attB\) plasmids, as all integration events were at the desired \(attP\) sites (Groth et al., 2004). However, rare non-specific integrations have been documented in Drosophila (Venken et al., 2006; Nimmo et al., 2006; Bischof et al., 2007). The \(\Phi C31\) integrase-mediated transformation technique has also recently been introduced successfully in the yellow fever mosquito Aedes aegypti (Nimmo et al., 2006).

After the original report describing two \(attP\) \(P\) element docking sites (Groth et al., 2004), numerous additional docking sites have been created. One set is embedded in a piggyBac backbone (Venken et al., 2006), whereas a second set is embedded in a Mariner backbone (Bischof et al., 2007). Venken et al. (Venken et al., 2006) observed that one out of seven docking sites tested was not receptive, suggesting that the genomic position of the docking site can affect integration efficiency. This was not observed for the 19 sites tested by Bischof et al. (Bischof et al., 2007). A detailed characterization and comparison of all the available docking sites will allow us to determine which ones are the most useful for specific purposes, such as cDNA overexpression, RNAi, genomic rescue or promoter/enhancer analysis.

Although the first reports used mRNA-encoded \(\Phi C31\) integrase to integrate the DNA (Groth et al., 2004; Bateman et al., 2006; Venken et al., 2006), Bischof et al. (Bischof et al., 2007) recently reported an efficient germ line \(\Phi C31\) integrase source that is driven by nanos or vasa regulatory elements. Interestingly, through \(\Phi C31\) integrase-mediated transgenesis, different \(\Phi C31\) integrase sources have been incorporated at the same docking sites. Additionally, the same \(\Phi C31\) integrase source was integrated into different docking sites, allowing the most efficient genomic \(\Phi C31\) integrase source to be selected. In the same study, a Drosophila codon-optimized \(\Phi C31\) integrase was described that performs better than the non-optimized version (Bischof et al., 2007).

Site-specific integration using a single recombination site results in the integration of the vector backbone, which may interfere with transgene expression (Chen et al., 2003). This can be minimized through marker genes strategically positioned between transgene and vector backbone (Venken et al., 2006). Alternatively, appropriately engineered recombinase sites in both the docking site and integration plasmid can be used to remove unwanted vector backbone sequence after correct integration events are isolated (Bischof et al., 2007). Finally, the integration of the backbone can be directly avoided through recombinase-mediated cassette exchange (RMCE) (Baer and Bode, 2001).
In RMCE, both docking site and transgene are flanked by a recombination site (see Box 4 and Fig. 7). Double reciprocal crossover results in the integration of a transgene without its vector backbone. However, two sets of directly oriented $loxp$ or FRT sites will favor deletion over RMCE. This problem can be overcome with sites, called spacer variants, that support recombination between themselves but not with others (see Box 4 for more information). The use of RMCE with spacer variants was initially utilized in the mouse in Cre-(Bethke and Sauer, 1997; Bouhassira et al., 1997) and FLP- (Seibler et al., 1998) based genetic engineering. This approach has been recently exploited in Drosophila for both recombinases (Oberstein et al., 2005; Horn and Handler, 2005) (Fig. 7B). For example, RMCE has been used elegantly to perform structure-function analysis of the eve2 (eve–FlyBase) enhancer with a $lacZ$ reporter (Oberstein et al., 2005). An alternative way to ensure that RMCE avoids the deletion or integration of plasmid backbone when employing FLP or Cre, is to use inverted recombination sites (as shown for $PhiC31$ integrase in Fig. 7C), which was pioneered in the mouse using Cre (Feng et al., 1999). This strategy eliminates the deletion problem but causes inversions.

Interestingly, RMCE events can be locked using the $PhiC31$ integrase system, as these recombination reactions are unidirectional. Pioneered in the yeast Schizosaccharomyces pombe (Thomason et al., 2001) and in mouse (Belteki et al., 2003), a modification of this system using inverted att recombination sites was recently described for Drosophila (Bateman et al., 2006) (Fig. 7C). This study demonstrated that unmarked constructs can be integrated through RMCE, as site-specific integration events are identified by loss of the marker (Bateman et al., 2006).

**Recombineering: BAC transgenesis for Drosophila**

Transposons are generally characterized by a low cargo capacity, limiting the amount of DNA that can be integrated and mobilized. Transgene size limitation can be overcome by the incorporation of a site-specific integration system, such as $PhiC31$ integrase, as shown in a chicken cell culture system (Dafhnis-Calas et al., 2005), or by gene targeting at the $Hprt1$ locus in mouse embryonic stem (ES) cells (Heaney et al., 2004). Unfortunately, an intermediate cell culture system supporting both in vitro gene manipulation and subsequent germ line transmission, similar to mouse ES cells, is not available for Drosophila. Moreover, there

---

**Fig. 6. Gene targeting in Drosophila.** (A) Ends-in insertional gene targeting. The donor construct, within a $P$ element, contains a region of homology (the targeting construct, red) interrupted by a restriction recognition site for the meganuclease I-SceI and flanked by FRT recognition sites for FLP recombinase. It also contains a marker ($white^+$) and an appropriately located restriction recognition site for the meganuclease I-CreI for a second round of homologous recombination. After $P$ element transgenesis, a linearized episome is generated in vivo by FLP and I-SceI. Correct targeting results in $white^+$ expression and a tandem duplication of the locus. This duplication can be reduced to single copy using I-CreI, resulting in loss of $white^+$. (B) Ends-out replacement gene targeting. The donor construct, within a $P$ element, contains a region of homology interrupted by a $white^+$ marker and is flanked by restriction recognition sites for the meganuclease I-CreI and FRT recognition sites for FLP recombinase. After $P$ element transgenesis, identified by $white^+$, linearized targeting DNA is generated in vivo by FLP and I-SceI. Correct targeting results in a $white^+$ phenotype and replacement of part of the locus.
is a strong negative correlation between the upper size limit of the insert and plasmid copy number: large DNA fragments are unstable when present in high-copy-number vectors in bacteria. Therefore, low-copy-number plasmids, such as P1 bacteriophage (Sternberg, 1990), bacterial artificial chromosomes (BACs) (Shizuya et al., 1992) and P1 artificial chromosomes (PACs) (Ioannou et al., 1994), were developed to maintain large inserts. Unfortunately, these plasmids interfere with both cloning and linearization of the transgene between both homology arms using Cre or FLP is provided. Correct recombination events identified through absence of marker 1 and result in hybrid sites, that are no longer substrates for ΦC31 integrase. Therefore, low-copy plasmids has an upper size limit of ~30 and 80 kb, respectively (Lee et al., 2001). However, by maintaining the plasmid at low copy number, large fragments of up to 102 kb can be efficiently gap-repaired (Venken et al., 2006), as observed by others (Kotzamanis and Huxley, 2004). Interestingly, one 133 kb fragment that encodes one of the largest genes in the fly genome, Tenascin major, was reconstituted from two different BACs, each containing part of the gene (Venken et al., 2006). Gap-repaired DNA was induced to high copy number, isolated, and integrated into the fly genome using both attP and attB sites, that are no longer substrates for ΦC31 integrase. The integrated DNA can be in either orientation (arrows).
fragments, as the occurrence of unique cutting sites decreases with increasing plasmid size. Recombineering does not suffer from those limitations and allows BACs to be modified more rapidly using PCR products or oligonucleotides that contain the desired mutation as recombination templates (Copeland et al., 2001; Court et al., 2002) (Fig. 9A). This strategy can easily be combined with positive/negative selectable markers, such as galK and thyA (Warming et al., 2005; Wong et al., 2005). Positive/negative selectable markers are targeted to the desired site for mutagenesis during a first round of recombineering using selection, and are then replaced by the desired mutation or tag during a second round of recombineering using counterselection (Fig. 9B). An alternative way to modify DNA constructs uses the recombinogenic protein RecA, also known as RecA-assisted modification (Yang et al., 1997; Gong et al., 2002), a methodology that is somewhat different from traditional recombineering. In a first recombination step, a modifying plasmid is integrated into the target plasmid, resulting in a co-integrate that becomes resolved during a second round of recombination (Fig. 9C). The technique allows the integration and deletion of fragments within a genomic fragment (Misulovin et al., 2001) and has been applied on a high-throughput level in the mouse field to create an atlas of gene expression in the mouse central nervous system (Gong et al., 2003). BAC modification was pioneered in the mouse field because most mouse genes tend to have desired locations might allow the site-specific integration of any element could be used to convert many of the existing genome deletions by FLP/FRT recombination (Parks et al., 2004; Ryder et al., 2004).

Combinations of the different methodologies described here should also greatly enhance our ability to manipulate the fly genome. For example, P replacement with an attP-containing P element could be used to convert many of the existing P elements into a useful docking site for PhiC31 integrase-mediated transgenesis or RMCE. Alternatively, gene targeting of recombination sites at desired locations might allow the site-specific integration of any DNA fragment. Finally, PhiC31 integrase-mediated transgenesis can be used to insert the donor constructs that are required for gene targeting. These are just a few examples of possible future fly manipulations. In another vein, these technologies will also help to improve genome-wide studies of Drosophila. For example, one could try to identify optimal genomic sites for the integration of all RNAi

**Future applications**

The advent of site-specific integration combined with recombineering and other methodologies will impact the fly field in numerous ways. These techniques make it possible to carry out structure-function studies at a higher resolution with fewer transgenes, as position effects can be mitigated using some of these approaches. Moreover, we are no longer confined to the study of single genes but can now tackle entire gene complexes that play key roles in development. Through repeated rounds of mutagenesis via recombineering, one can dissect the in vivo role of each gene and each regulatory region within a gene complex. Similar manipulations are now also possible for larger genes and for loci that have previously had no available mutations to study. These loci can now be identified and studied through the introduction of small deletions by FLP/FRT recombination (Parks et al., 2004; Ryder et al., 2004).
constructs. These sites should permit the optimal expression of RNA hairpin loops in all tissues at all developmental stages to allow the efficient RNAi-mediated knockdown of any gene. This approach may alleviate some of the potential drawbacks that are associated with P transposase-mediated transgenesis of RNAi constructs, such as poor transgene expression or misregulation (Dietzl et al., 2007). Furthermore, many different genomic DNA fragments containing cis-regulatory elements that drive GAL4 expression could be integrated at the same docking site to allow the labeling and manipulation of specific cell populations. Finally, the integration of overlapping duplications of defined areas of the X chromosome into the same docking site would be a useful way to map essential genes on the X chromosome.

Each transgenesis technique might result in unwanted side effects. Transposases might cause multiple insertion and excision events before the final transgenic insertion is stabilized. Recombinases and integrases may recognize pseudo-sites localized within the genome, as previously identified for Cre in the mammalian genome (Thyagarajan et al., 2000). Moreover, at high doses, Cre has been demonstrated to result in undesired effects in both vertebrates (Schmidt et al., 2000; Loonstra et al., 2001) and Drosophila (Heidmann and Lehner, 2001). Detrimental side effects have been observed for PhiC31 integrase in mammalian cell culture (Ehrhardt et al., 2006; Liu et al., 2006), although ectopic expression in vivo in mice and Drosophila indicates that it has minimal side effects in these organisms (Belteki et al., 2003; Raymond and Soriano, 2007; Bischof et al., 2007). Even gene targeting in Drosophila has not been spared from artifacts, and second-site mutations have been shown to cause interference with phenotypic characterization (O’Keefe et al., 2007). Interestingly, so far no such observations have been documented for FLP, which is widely used within the fly community (Blair, 2003).

Unfortunately, the solution to an important problem such as efficient site-specific integration, immediately results in the creation of new challenges: the handling of thousands of fly strains associated with typical high-throughput projects, as well as the maintenance of thousands of new stocks. Recent methods for the automated microinjection of fly embryos for high-throughput in vivo RNAi experiments have been developed (Zappe et al., 2006), although no solution has yet been developed to maintain numerous additional fly stocks, except through the expansion of existing or new stock centers.

We apologize to those whose work could not be cited owing to space limitations. We thank Karen Schulze for comments on the manuscript. K.J.T.V. and H.J.B. are supported by the NIH and the Howard Hughes Medical Institute.
References


