Reverse genetics in invertebrates

Introduction

The main goal of transgenesis is germ-line transmission of the inserted sequences. Therefore, the altered gene must recombine with the cell's genome. In bacteria and yeast, this occurs with high frequency using the cell's own recombination machinery (see lesson 1 in yeast). In more complex organisms, the procedure of genome integration is more complicated due to a lack of efficient recombination mechanisms. In this lesson, we will discuss the techniques applied to integrate transgenes into the genome in invertebrates and will focus on methods that were originally developed for *Drososphila*, but have been used for other model systems as well.

Integration of the transgenic DNA: Transposons

Insect transgenesis, in general, has been dominated by transposon-mediated integration. Transposons are mobile DNA sequences (also called transposable elements, TE) that can relocate from one genomic location to another. Mobile elements were first discovered by Barbara McClintock in maize in 1951. She realized that mobile genetic elements can inactivate maize genes in which they reside, causing chromosome breaks and transpose to new locations within the genome, resulting in different phenotypes. She also suggested that these mysterious mobile elements of the genome might play some kind of regulatory role, determining which genes are turned on and when this activation takes place. The early speculations of McClintock were largely dismissed by the scientific community. Only recently have biologists begun to accept the possibility that this so-called "junk" DNA might not be junk after all. In fact, scientists now believe that TEs make up more than 40% of the human genome. It is also widely believed that TEs might carry out some biological function, most likely a regulatory one - just as McClintock speculated. Like all scientific hypotheses, however, data from multiple experiments were required to convince the scientific community of this possibility.

Eventually, McClintock's work was awarded the Nobel prize in 1983 (for her discovery made more than 30 years before).

Transposable elements have been found to exist in many organisms and are fairly frequent in their genomes; in maize, approximately 90% of the maize genome is made up of transposable elements, as is 44% of the human genome.

There are two broad classes of transposable element: DNA transposons and retrotransposons. Most DNA transposons move as pieces of DNA, cutting and pasting themselves into new genomic locations. The transposition of DNA transposons is catalyzed by an enzyme called transposase (see figure 3-1A). In contrast, retrotransposons relocate by a copy-and-paste mechanism: first, they are transcribed into RNA, which is reverse-transcribed to DNA. This copied DNA is then re-inserted in the genome at a different position. Thus, the original transposon is maintained at its original position (see figure 3-1B).

Retrotransposons are further divided into two groups: long terminal repeat (LTR) retroelements and non-LTR retroelements. LTR retrotransposons contain two long terminal repeats (LTRs; black arrows) and encode the genes for the enzymes reverse transcriptase and integrase, both crucial for retrotransposition. Non-LTR retrotransposons lack LTRs and encode genes for reverse transcriptase and an endonuclease that cuts the target DNA. Each group of transposable elements contains autonomous and non-autonomous elements. Autonomous elements encode the products required for transposition. Non-autonomous elements that are able to transpose have no significant coding capacity but retain the sequences necessary for transposition.

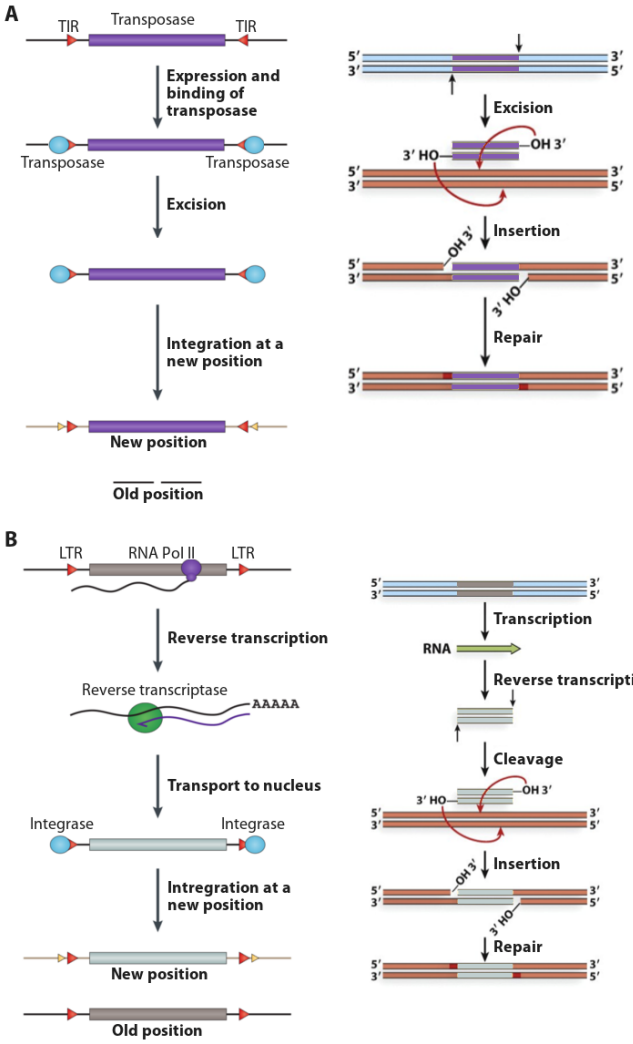

Figure 3-1 The two classes of transposons and their mechanisms of transpositions. (A) DNA transposons. Many DNA transposons are flanked by terminal inverted repeats (TIRs; red arrows), encode a transposase (blue circles), and mobilize by a 'cut-and-paste' mechanism. The transposase binds at or near the TIRs, excises the transposon from its existing genomic location, and pastes it into a new genomic location (brown bar). In the 'cut-and-paste' mechanism, the transposase makes a double-stranded cut in the donor DNA at the ends of the transposon and cleaves the target DNA such that the cleavages of the two strands are staggered. Each end of the donor DNA is then joined to an overhanging end of the recipient DNA. The 5′ end in the target DNA undergoes a nucleophilic attack from the transposon transferred strands 3′-OH. Each end of the donor DNA is then joined to an overhanging end of the recipient DNA. **(B) Retrotransposons.** Retrotransposons transpose via a 'copy-and-paste' mechanism in which mRNA transcribed from the element by RNA polymerase II (RNA Pol II) is converted into a cDNA by reverse transcription and then integrated by an integrase enzyme (blue circles) at a new position in the genome. Retrotransposons are further divided into long terminal repeat (LTR) retroelements and non-LTR retroelements, which differ in the mechanism of integration. Here, the mechanism for LTR retroelements is shown. LTR retrotransposons contain two long terminal repeats (LTRs; red arrows) and encode all proteins crucial for retrotransposition. The 5′ LTR contains a promoter that is recognized by the host RNA polymerase II and produces the mRNA of the retrotransposon (black wave). The reverse transcriptase (green circle) copies the TE mRNA into a full-length dsDNA. In the second step, integrase (blue circles) inserts the cDNA (now indicated as grey box) into the new target site. Similarly to the transposases of DNA transposons, retrotransposon integrases create staggered cuts at the target sites for integration of the recipient DNA. (adapted D. Lisch, *Nat. Rev. Genet.*, 2013)

Researchers have realized the potential of transposons to alter DNA inside a living organism very early. For example, in *Drosophila* and *C. elegans*, libraries of individuals who carry transposable elements at different sites in the genome are used for forward-genetic studies.

Transposon-mediated transgenesis in *Drosophila*

In *Drosophila*, transgenesis mainly relies on the P element transposon, the introduction of which was one of the most important breakthroughs in germ-line transgenesis in *Drosophila*. *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 location of integration cannot be controlled.

P elements are transposable elements, which were originally identified within the fly's own genome. They were discovered when researchers tried to mate females from laboratory strains with males derived from wild populations. The progeny of these crosses showed sterility, a high mutation rate and a high frequency of chromosomal aberrations. These flies are dysgenic, or biologically deficient, and the phenomenon is therefore called hybrid dysgenesis. Investigators hypothesized that the dysgenic mutants are caused by the insertion of transposable elements into genes, thereby rendering them inactive and causing the mutation. Indeed, it was found that many mutations in the dysgenic flies were caused by insertions of transposable elements into genes. This transposable element was the P

element. Interestingly, P elements were inserted frequently in wild-type strains of *Drosophila*, but not in laboratory strains. Why do the P elements present in wild-type strains not cause trouble, e.g., are inserted into genes and cause mutations? It is now thought that the transposase genes in P elements are silenced in wild-type strains, thus, P elements cannot relocate in wild-type flies.

P elements, like other transposons, contain two terminal repeats, including inverted-repeat sequences of 31 base pairs and other internally located sequence motifs absolutely required for their mobilization or transposition. Mobile (also autonomous) P element transposons encode an enzyme called P transposase that catalyzes transposition of the transposon. To use P elements for transgenesis, the P transposase was separated from the P element transposon backbone (called non-autonomous P element, see figure 3-2). A plasmid that encodes P transposase, a so-called helper plasmid, is provided together with the transgene that contains the transposon backbone necessary for integration, the sequence of interest and a marker. 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.

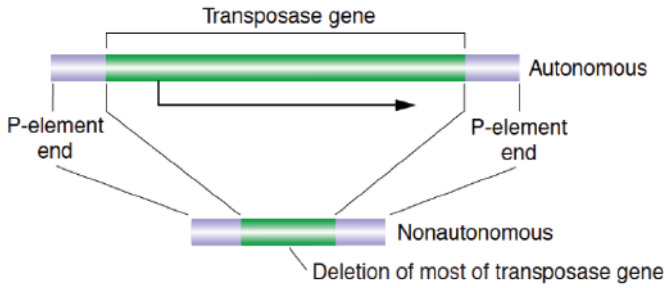

Figure 3-2 Structure of a P element. In the original version, the P element encodes a protein called transposase that is necessary for P elements to move to new positions in the genome. This type is called "autonomous", because it can be transposed through the action of its own transposase. The second type is called "non-autonomous", because the transposase gene is deleted; but these elements can be mobilized by externally supplying transposase. The P-element ends contain the inverted repeats required for insertion into the genome.

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. To circumvent these limitations, several other transposons with a different insertional specificity have been identified that are suitable for germ-line transformation in *Drosophila*. These include the PiggyBac transposon identified in the cabbage looper moth *Trichoplusia ni.*, which can be used in a variety of hosts, ranging from insect to mammals.

Applications of transposon-mediated transgenesis in *Drosophila*

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 or to insert exogenous genes into the genome. Almost all technological progress in flies depends on the ability to transform them, and P element-mediated transformation is the basis

for all of them. For example, the use of the Flp/FRT system to create mutant clones by inducing mitotic recombination, the gene-knockout methods in flies, or the generation of a genome-wide RNAi library of transgenic insertions to knock down most fly genes would not have been possible with the use of P elements. In addition, many forward-genetic screens using special P elements (EP elements) to overexpress genes rely on these elements.

Enhancer traps

One transgenic technology that relies on P element transposition is the so-called enhancer-trap assay. An enhancer trap is a transgenic construct that is used to identify genes that are expressed in a specific tissue. When the construct inserts near a tissue-specific enhancer, the weak promoter present on the construct comes under the control of this enhancer, resulting in tissue-specific expression of the reporter gene.

Thus, an enhancer trap allows hijacking of an enhancer from another gene, and so, identification of which enhancers regulate the expression of which gene. The enhancer trap construct contains a transposable element (that allows random integration into the genome) and a reporter gene (that allows visualization of the spatial regulation by the enhancer). For selection, the construct usually includes a genetic marker, e.g., w^+, producing red-colored eyes in *Drosophila*. The most common enhancer trap constructs are P[lacW] that carries a *lacZ* gene fused to the the 5' end of the P element (see figure 3-3). As P elements frequently insert near the promoter of a gene, *lacZ* will often be expressed in the same spatial and temporal pattern as the gene into which the element is inserted.

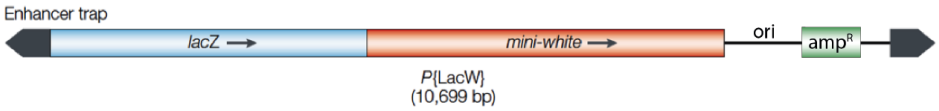

Figure 3-3 The enhancer trap construct P[lacW]. This construct consists of a P element that contains the *lacZ* gene and a *white+* gene (here called *mini white*). Further, it contains elements that are necessary for propagation in bacteria (ori and ampicillin resistance). The black arrows represent the P element ends containing the inverted repeats.

By mobilizing the transposon to other random chromosomal sites, one can detect diverse enhancers at those sites and study their tissue-specific gene control or identify flanking genes displaying interesting expression patterns. Enhancer-trap transposons have been engineered to permit cloning of genomic DNA adjacent to the site of any particular insertion, enabling the isolation of the enhancer and its target gene from genomic DNA.

As you can see in figure 3-3, P[lacW] also carries sequences that are necessary for propagation as circular plasmids in bacteria, such as the ori (origin of replication) and the ampicillin-resistance gene. These elements are useful for a method called "plasmid rescue" that serves to clone DNA that flanks a transgenic construct. This is useful in all cases where the insertion site of the P element is not known, e.g., to identify the original insertion site of a P element or after a P element has been mobilized to integrate into another unknown location in the genome. Genomic DNA isolated from flies containing a P[lacW] insertion can be cut with restriction enzymes, the cut sequence is ligated, and the circular DNA is transformed into bacteria. After selection for ampicillin resistance, the plasmid DNA is recovered and sequenced to identify the gene. Like this, the gene that is associated can easily be identified (see figure 3-4).

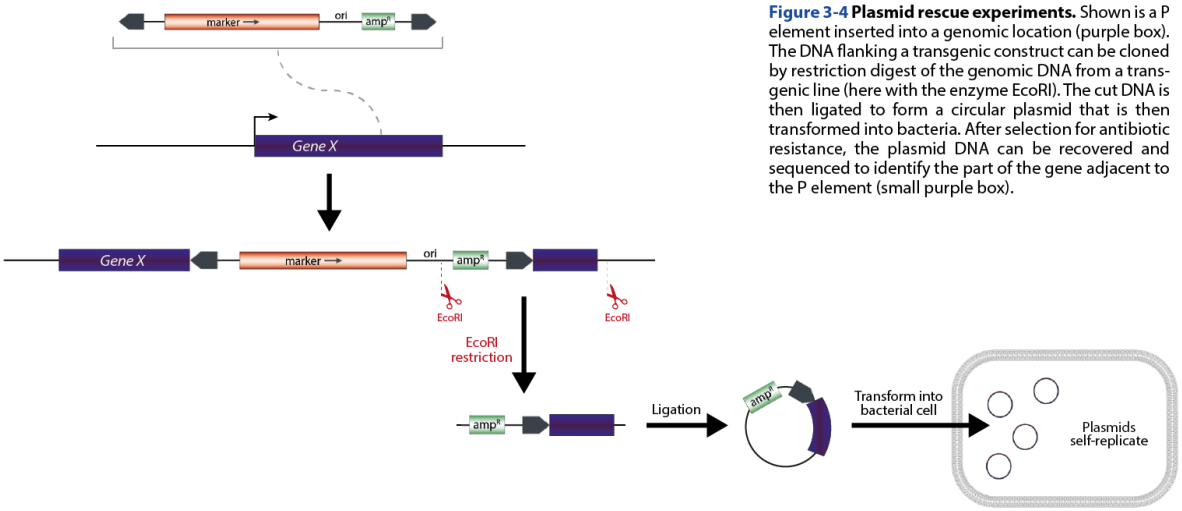

Figure 3-4 Plasmid rescue experiments. Shown is a P element inserted into a genomic location (purple box). The DNA flanking a transgenic construct can be cloned by restriction digest of the genomic DNA from a transgenic line (here with the enzyme EcoRI). The cut DNA is then ligated to form a circular plasmid that is then transformed into bacteria. After selection for antibiotic resistance, the plasmid DNA can be recovered and sequenced to identify the part of the gene adjacent to the P element (small purple box).

Gene-disruption methods using P elements

We will first discuss how transposons can be used to create mutations by insertion. P elements inserted into genes disrupt genes at random locations, creating mutants with different phenotypes. Several gene-disruption projects have generated thousands of stocks that each contain a single P element construct inserted at a known position of the genome. For many of these, the exact sequence at which the element is inserted has been determined. These stocks are available to all researchers and one can simply order the stocks that carry an insertion in the gene of interest. We will discuss here a few of these projects and how the stock collections they generated can be used for reverse-genetic approaches.

Gene misexpression by P elements

We have already discussed the use of special P elements, so called EP elements, that carry UAS sequences that allow activation by the transcription factor Gal4 (see lesson 2 in "Forward genetics"). If a P element is inserted near the 5' end of a gene, the expression of this gene can be induced by Gal4. There exist a few thousand P(EP) insertion lines that can be used for overexpression screens.

Gene disruption

Transposons are useful tools to disrupt genes if they are inserted such that they suppress the expression of a gene by disrupting its sequence. Thus, transposons can act as a mutagen to cause insertional mutations in proximally located genes (see figure 3-5). P elements have been used in different projects with the aim to introduce a disrupting insertion into every gene in *Drosophila*. These insertion mutants are extremely useful for many purposes. They are used for gene identification in complementation analysis for mutants identified in forward-genetic screens. Thus, they serve to validate a mutation discovered in forward-genetic screens by comparing the phenotypes of the insertional mutant with the one obtained in the forward screen. In reverse genetics, insertion mutations are still one of the first resources to explore if one wants to study the function of a certain gene. There are a few thousand different insertion mutants publically available that can be simply ordered and assayed for their phenotype. However, for many genes, insertion mutations are still lacking. There a several methods in which P elements can be used to generate new mutations close to the original insertion, since P elements have the tendency to transpose locally (within 200 kb). If there is a P element near the gene of interest it can be remobilized providing the transposase enzyme. The P element will often transpose within a few kilobases of the original location, hopefully affecting your gene of interest.

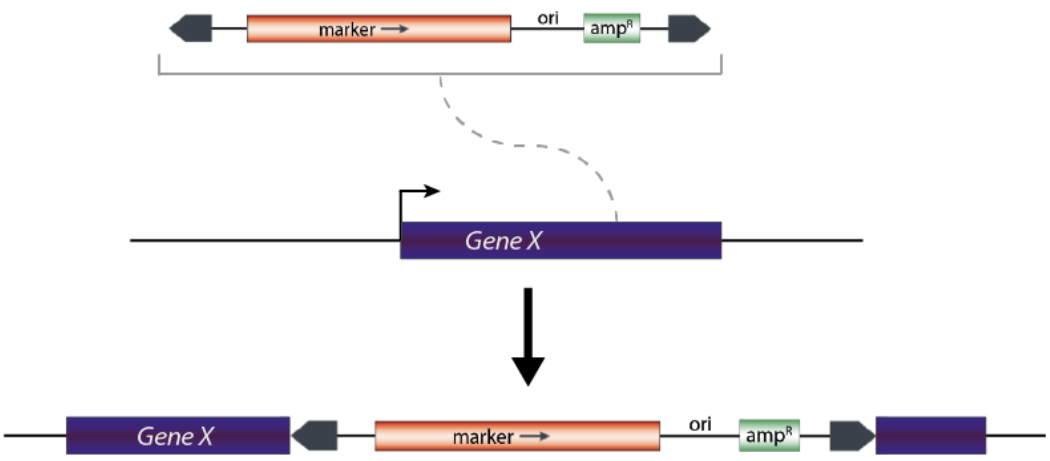

Figure 3-5 Gene disruption by P elements. A P element is inserted into a gene (*Gene X*, purple box), thereby disrupting its sequence.

P element excision: the fly knockout

P elements can also be used to create gene-specific deletions (gene knockouts) in the fly. If a P element is inserted in close vicinity of a gene of interest, a null allele can be created fairly easy using the "imprecise-excision" method in which the excision of the P element is induced by temporarily providing transposase activity (crossing into a transposase expressing line). Because transposase is only present temporarily in these organisms, its activity is mainly restricted to mediating the excision of existing P elements. Thus, excised P elements usually do not integrate at another locus. The method actually relies on the fact that excision of the P element results in double-strand breaks (DSB) at the insertion site, which have to be repaired by the cellular repair machinery. DNA repair mechanisms can process the DSB in different ways, generating different products (see figure 3-6).

In one scenario, the ends of the DSB are used to induce new DNA synthesis, using a homologous sequence as a template (this is called homology-directed repair). This can lead to either a restoration of the P element at the insertion site (if the P element on the sister chromatid served as a template for synthesis) or the restoration of the original chromosome without the P element (if the homologous chromosome without the P element served as a template for synthesis). The latter is also called a "precise excision", because the P element is excised without damaging the chromosomal region where it was inserted (see figure 3-6).

Alternatively, before repair, the DSB can be enlarged to a gap by the cellular repair machinery, and repairing the gap can lead to deletions that can extend in one or both directions of the original insertion site. Those events are relatively frequent such that imprecise-excision experiments will usually give a few different deletions for the gene of interest, which can be analyzed for their effect on gene function.

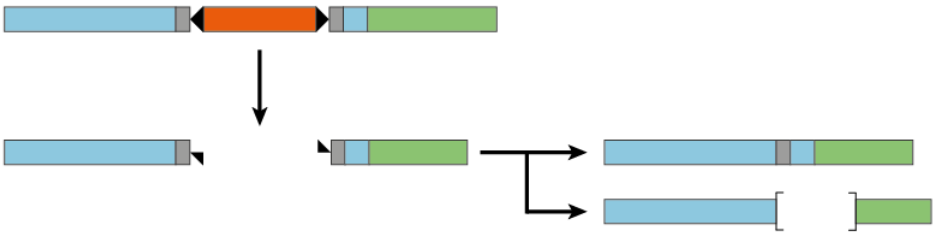

Figure 3-6 P element excision mutagenesis. A P element (orange) is inserted into a genomic location near a target gene (green). Transposase catalyzes the excision of the P element, which leaves a double strand break (DSB). Repair of these DSB can lead to either a precise excision of the P element (above) or to a deletion of sequences adjacent to the DSB, leading to a small deletions of the gene.

Post-transcriptional regulation by RNA interference

So far, we have discussed different methods to alter the genome of an organism to create transgenic animals. Now, we want to introduce a different method of gene-expression regulation that does not alter the DNA sequence, but regulates on a post-transcriptional level. RNA interference (RNAi) is a naturally occurring cellular process in eukaryotes, in which short regulatory RNA molecules interfere with mRNA translation and thereby affect – usually suppress - protein production. How crucial this mechanism is for gene-expression regulation has only become apparent during the last 20 years. By now, we know that RNA interference is involved in practically all biological processes and forms a complex networks of regulatory pathways tightly linked with other cellular regulation mechanisms.

The general mechanism can be divided into three steps (see figure 3-7): i) double-stranded RNA fragments are cleaved into smaller pieces of 20 to 30 base pairs in length by ribonucleases; ii) these shorter RNA fragments are separated into single strands, of which one strand is incorporated into an enzyme complex known as the RISC complex; and iii) the RNA-enzyme complex recognizes mRNA molecules with a sequence that is complementary to the RNA fragment in the complex and blocks or cleaves it, preventing translation.

Now, the question becomes, where does an RNA fragment come from that provides sequence specificity to the inhibitory enzyme complex? In principle, there are two answers to this question: First, the RNA fragment was transcribed from non-protein-coding regions of the cell's own genome. The phenomenon that the genome of all animals and plants encodes RNA molecules that do not code for proteins, but have a regulatory role was discovered in the 1990th, and earned Andrew Fire and Craig Mello the Nobel Prize in 2006 for their discovery of the mechanism of RNA interference in *C. elegans*. We will discuss gene regulation by small non-coding RNAs in great detail in the next part of this course. Second, RNA can also be introduced into the cell, which provides interesting opportunities for experimental manipulation. What makes this mechanism so attractive for biological research is that is allows sequence-specific inhibition of a gene without altering the genome permanently. RNAi can be used to infer the function of a protein-coding gene by knocking it down and observing the effect, or, using libraries of RNA molecules, a gene encoding a known function can be identified. Looking at this last application, you may realize that some technologies cannot be easily categorized as forward- or reverse-genetics tools, but can be applied to a multitude of biological questions.

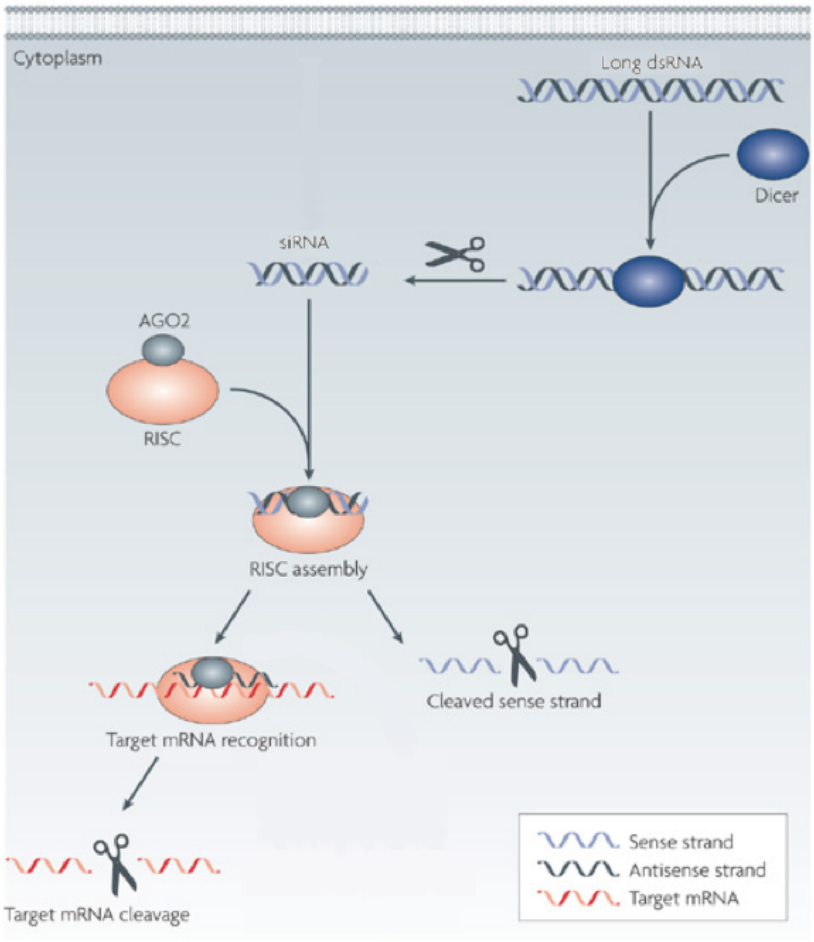

Figure 3-7 Long double-stranded RNA (dsRNA) is cleaved into small-interfering RNA (siRNA) by the enzyme Dicer. The siRNA is then incorporated into the RNA-induced silencing complex (RISC). The sense strand of the siRNA is cleaved by the protein argonaute 2 (AGO2) and only the antisense strand of the siRNA is used for target mRNA recognition. The activated RISC–siRNA complex seeks out, binds to, and degrades complementary mRNA, which leads to the silencing of the target gene. (adapted from Anderson *et al.*, *Nat. Rev. Drug Discov.*, 2009)

RNAi in model organisms

RNAi has become a widely used method for studying gene function. The generation of loss-of-function phenotypes by depletion of the corresponding transcript has facilitated functional studies of genes in various organisms, from *C. elegans* and *Drosophila*, to human cells and organisms such as planaria and house flies for which genetic techniques were not available so far. RNA interference can be applied in different model organisms in various ways. In *C. elegans*, double-stranded RNAs are introduced simply by feeding the worms with bacteria that express these RNAs (see figure 3-8A). In *Drosophila* cultured cells, dsRNA molecules that can be easily generated by *in vitro* transcription of plasmids are added to the culture medium (see figure 3-8B). The double-stranded RNA has to be processed inside C. elegans or *Drosophila* cells by dicer to produce siRNAs that can be incorporated into the RNA-induced silencing complex (RISC). In human and mouse cell lines, siRNA can be generated *in vitro* (there are even companies that synthesize siRNAs) and is introduced into cultured cells by transfection. These siRNAs can be directly incorporated into the RISC to mediate gene silencing (see figure 3-8 C).

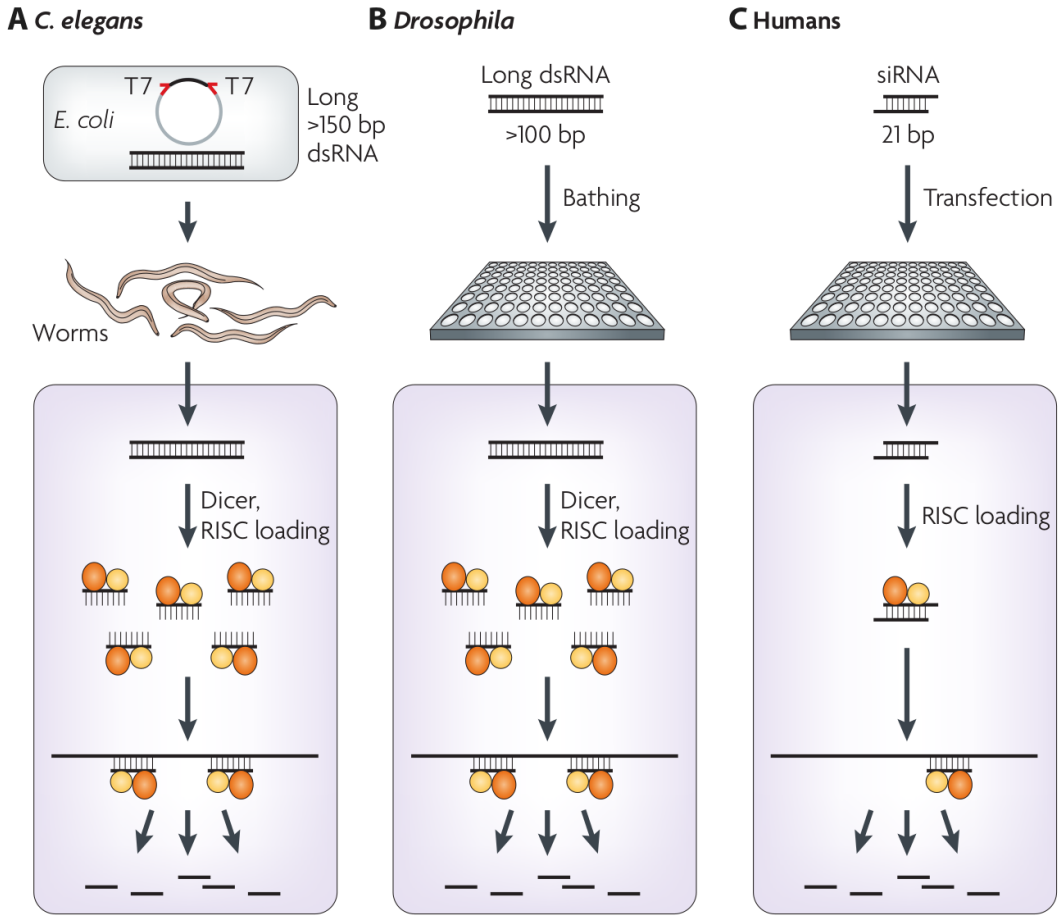

Figure 3-8 Overview of RNAi screening approaches used in different organisms. Long double-stranded (ds) RNAs are introduced into *C. elegans* simply by ingestion of *E. coli* expressing dsRNAs **(A)** or into *Drosophila* cells (by bathing) and are intracellularly diced into small-interfering RNAs (siRNAs, **B**). This leads to highly efficient knockdown, because many different siRNAs are generated from each dsRNA. In worms and flies, long double-stranded RNA (dsRNA) is introduced into the cytoplasm, where it is cleaved into small interfering RNA (siRNA) by the enzyme Dicer. In human (or vertebrate) cells, siRNA can be introduced directly into the cell by transfection **(C)**. The siRNA is incorporated into the RNA-induced silencing complex (RISC) and the activated RISC–siRNA complex seeks out, binds to and degrades complementary mRNA, which leads to the silencing of the target gene. (M. Boutros and J. Ahringer, *Nat. Rev. Genet.*, 2008)

Adding dsRNA to *Drosophila* cultured cells is easy and rapid (silencing occurs with 72 h after adding the dsRNA) and therefore well suited for large-scale screening. However, the goal of most reverse genetics studies is to determine phenotypes that are associated with the loss of gene function in the organism. Therefore, transgenic flies have been generated that express an RNA with a long inverted repeat that can fold back on itself to become double stranded (a so-called hairpin RNA). This hairpin RNA is then processed into siRNAs by dicer to mediate RNA degradation. By placing the RNA hairpin construct under the control of *UAS* sequences, it is possible to express the dsRNA in specific tissues or at specific times during development using the Gal4-UAS system we already described in the forward genetics lessons. In brief, the binding of the transcription factor Gal4 to *UAS* sequences activates the expression of the gene downstream of the *UAS* sequences. Since many different Gal4 lines exist that drive the expression of Gal4 in different tissues, or those where the Gal4 is under the control of an inducible promoter, both tissue-specific as well a temporal expression of the dsRNA can be achieved. A library of transgenic flies expressing different RNA hairpin constructs was constructed, targeting over 88% of all protein-coding genes in the fly. This library is publically available and can be used to systematically silence genes and analyze their functions in any tissue and any stage of *Drosophila*

lifespan. By crossing flies containing a UAS-RNA hairpin construct targeting a gene of interest with flies expressing Gal4 under a tissue specific promoter, the expression of the gene of interest can be inhibited (see figure 3-9).

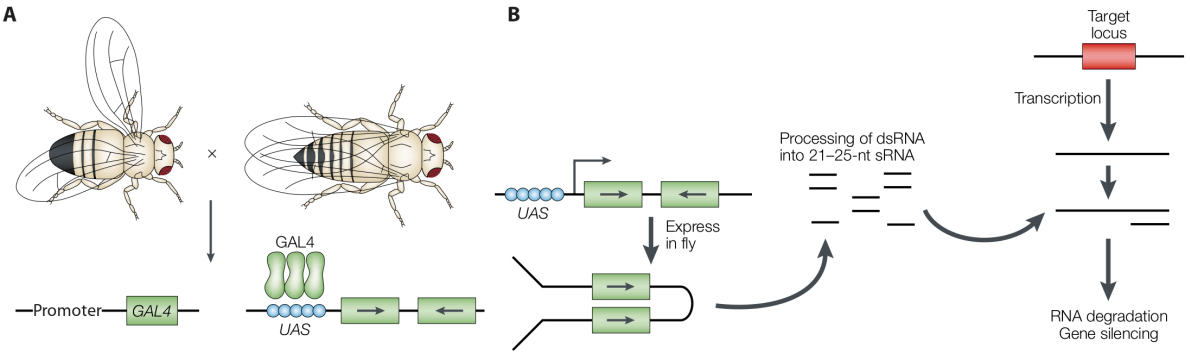

Figure 3-9 The Gal4-UAS system for directed gene expression is used to downregulate genes in *Drosophila* by RNA interference. (A) Flies containing the *GAL4* gene under the control of a tissue-specific promoter are crossed with flies carrying inverted repeat sequences to produce a hairpin RNA. **(B)** Double-stranded DNA (dsRNA) that is homologous to the target gene is delivered by expression of an inverted repeat RNA. The dsRNA is processed into 21-25-nucleotide small-interfering RNAs (siRNA) by the Dicer ribonuclease. The siRNA is used to guide the sequence-specific degradation of mRNA, leading to post-transcriptional silencing of the target locus. (A: adapted from D. St Johnston, *Nat. Rev. Genet.*, 2002; B: adapted from M.D. Adams and J.J. Sekelsky, *Nat. Rev. Genet.*, 2002)

Notably, the generation of the transgenic fly library carrying the hairpin RNA constructs was only possible due to the existence of P elements. The hairpin constructs were cloned into plasmids containing inverted repeats from a P element, which mediated the insertion of the hairpin construct into the genome of the flies. This represents one more example how P elements aided in developing gene technologies in the fly.

Summary

In this lesson, we have discussed several reverse-genetics approaches, mainly focusing on *Drosophila*. *Drosophila* genetics is dominated by the use of P elements. They are the basis of almost all transgenic technologies and were the key to develop most of the sophisticated tools that are now available to *Drosophila* geneticists. We have seen how P elements served to develop different transgenic methods (e.g., enhancer trap or misexpression libraries) or technologies (e.g., generation of Flp/FRT insertions and RNAi) and how they are used for gene disruption to create mutants (insertional mutagenesis and imprecise excisions). However, although these tools originate from *Drosophila*, a lot of them (for example, the Flp/FRT technique or transposon-mediated transgenesis) were adapted for other organisms, such as *C. elegans* and mice.