CHAPTER 13

Advances in transgenic technology

Introduction

In the previous three chapters, we discussed the early development of transgenic technology, which focused on the evolution of reliable techniques for gene transfer and transgene expression in animals and plants. The standard use of this technology was to add genetic information to the genome, generating a *gain of function* in the resulting cells and transgenic organisms.

Now that such gene-transfer processes are routine, the focus of transgenic technology has shifted to the provision of more control over the behaviour of transgenes. New techniques have evolved in parallel in animals and plants. Advances have come about through the development of inducible expression systems that facilitate external regulation of transgenes and the exploitation of site-specific recombination systems to make precise modifications in target genomes. In mice, the combination of gene targeting, site-specific recombination and inducible transgene expression makes it possible to selectively switch on and off both transgenes and endogenous genes in a conditional manner. Other routes to gene silencing have also been explored, such as the expression of antisense RNA and the recently described phenomenon of RNA interference.

Transgenic animals and plants are also being used increasingly as tools for the analysis of genomes. We discuss the development of tagging systems for the rapid cloning of genes disrupted by insertional mutagenesis and the use of entrapment vectors, which allow the high-throughput analysis of gene expression and function.

Inducible expression systems

In many gene-transfer experiments, the production of a recombinant protein is the ultimate goal, in which case it is generally appropriate to maximize transgene expression using a strong constitutive promoter. However, there are also situations in which maximized transgene expression is undesirable. For example, if a recombinant protein is toxic, constitutive high-level expression would be lethal and would prevent the recovery of stably transformed cell lines. In other experiments, the timing of transgene expression is critical. This would be the case if we wished to study the effect of protein overexpression at a specific point in the cell cycle in cultured cells or at a particular developmental stage in transgenic animals and plants. These issues can be addressed through the use of *inducible expression systems*, in which transgene expression is controlled by an external stimulus.

Endogenous inducible promoters

A number of inducible expression systems have been developed based on promoters from endogenous cellular or viral genes. An early example is the *Drosophila* heat-shock promoter. Most cells respond to elevated temperature by synthesizing *heat-shock proteins*, which include molecular chaperons and other proteins with protective functions (Parsell & Lindquist 1983). The response is controlled at the level of transcription by a heat-labile transcription factor, which binds to heat-responsive promoters in the corresponding genes. The promoter of the *Drosophila hsp70* gene has been widely used to drive transgene expression, both in *Drosophila* itself (Lis *et al.* 1983) and in heterologous systems (e.g. Wurm *et al.* 1986). In transgenic flies, any gene linked to the *hsp70* promoter is more or less inactive at room temperature, but high-level expression in all cells can be induced by heating to 37°C for about 30 min.

The heat-shock promoter is unusual in that the stimulus that activates it is physical. Most inducible promoters respond to chemicals, which must be supplied to the transformed cells or transgenic

organisms in order to activate the expression of a linked transgene. In mammals, several promoters are known to be activated by glucocorticoid hormones or synthetic analogues, such as dexamethasone. Two of these have been extensively used for inducible transgene expression – the mouse metallothionein promoter (Hager & Palmiter 1981) and the long-terminal-repeat (LTR) promoter of mouse mammary tumour virus (MMTV) (Lee *et al.* 1981). The metallothionein promoter is also induced by interferons and heavy metals, such as cadmium and zinc, allowing the transgene to be activated in transgenic animals, e.g. by including a source of heavy metals in the drinking-water. An example of zinc-induced activation of a human growth-hormone gene in transgenic mice is discussed on p. 208. A metal-inducible expression system has also been developed for plants, although the components of the system are derived from yeast (Mett *et al.* 1993). Endogenous inducible promoters used in plants include the *PR-1a* promoter, which is responsive to benzothiadiazole (Gorlach *et al.* 1996).

There are several limitations associated with the use of endogenous inducible promoters, and most systems based on such promoters suffer from one or more of the following problems. First, they tend to be somewhat leaky, i.e. there is a low to moderate level of background transcription in the absence of induction. Secondly, the level of stimulation achieved by induction (the *induction ratio*) is often quite low, typically less than 10-fold. Thirdly, there are often cytotoxic side-effects, caused by the activation of other endogenous genes that respond to the same inductive stimulus. The inducers of many of the commonly used endogenous promoters are also toxic if contact is prolonged. Fourthly, the kinetics of induction are often not ideal. For example, in transgenic animals and plants, there may be differential uptake of the inducer into different cell types and it may be eliminated slowly. There has therefore been great interest in the development of alternative inducible expression systems.

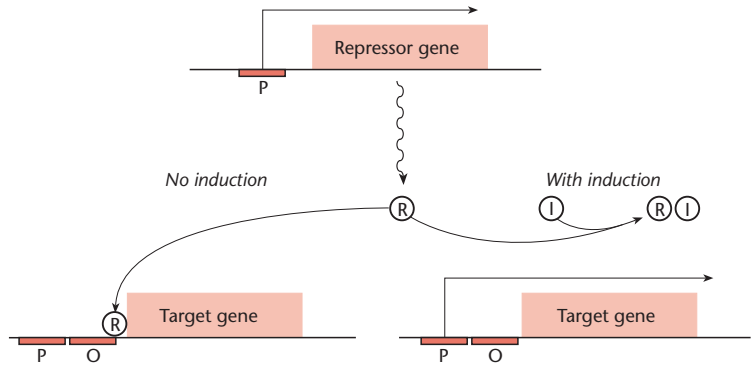
Recombinant inducible systems

Many of the disadvantages of endogenous inducible promoters can be addressed using recombinant systems. In the ideal system, induction is dependent on the supply of a non-toxic agent that specifically activates the transgene without interacting with any endogenous promoters. The agent is taken up rapidly and evenly, but has a short half-life, allowing rapid switching between induced and non-induced states. A range of inducers might be available with different properties. Steps towards this ideal system have been achieved using promoters and transcription factors that are either heterologous in the expression host or completely artificial.

The lac *and* tet *repressor systems*

The first heterologous systems were based on bacterial control circuits (Fig. 13.1). Hu and Davidson (1987) developed an inducible expression system for mammalian cells, incorporating the essential elements of the *lac* repressor control circuit. In *Escherichia coli*, transcription of the *lac* operon is switched off in the absence of lactose by a repressor protein encoded by the gene *lacI*. This protein binds to *operator* sites, the most important of which lies just downstream from the promoter, and thus inhibits transcriptional initiation. In the presence of lactose or a suitable analogue, such as isopropyl-β-D-thiogalactoside (IPTG), the *lac* repressor undergoes a conformational change that causes it to be released from the operator sites, allowing transcription to commence.

In order to use the *lac* circuit in eukaryotic cells, Hu and Davison (1987) modified the *lacI* gene by adding a eukaryotic initiation codon, and then made a hybrid construct in which this gene was driven by the Rous sarcoma virus (RSV) LTR promoter. The construct was introduced into mouse fibroblasts and a cell line was selected that constitutively expressed the *lac* repressor protein. This cell line was transiently transfected with a number of plasmids containing the *cat* reporter gene, driven by a modified simian virus 40 (SV40) promoter. Each of these plasmids also contained a *lac* operator site somewhere within the expression construct. The investigators found that, when the operator sites were placed in the promoter region, transcription from the reporter construct was blocked. However, transcription could be derepressed by supplying the cells with IPTG, resulting in strong chloramphenicol transacetylase (CAT) activity.



Component	Lac system	Tet system
Repressor gene	*lac* I	*tet* R
R	*Lac* repressor	*Tet* repressor
I	IPTG	tetracycline/doxycycline
O	*lac* operator	*tet* operator

Fig. 13.1 Summary of repression-based inducible expression-control circuits based in the *lac* and *tet* operons of *E. coli*.

A similar system, based on the *E. coli* tetracycline operon, was developed for tobacco plants (Gatz & Quail 1988). The *tet* operon is carried on a bacterial transposon that confers resistance to the antibiotic tetracycline. Similarly to the *lac* system, the *tet* operon is switched off by a repressor protein, encoded by the *tetR* gene, which binds to operator sites around the promoter and blocks transcriptional initiation. Tetracycline itself binds to this repressor protein and causes the conformational change that releases the *tet* operon from repression. Since tetracycline inhibits bacteria at very low concentrations, the *tet* repressor has a very high binding constant for the antibiotic, allowing derepression in the presence of just a few molecules. The *tet* repressor also has very high affinity for its operator sites. Therefore, cell cultures and transgenic tobacco plants expressing TetR were able to inhibit reporter gene expression from a cauliflower mosaic virus (CaMV) 35S promoter surrounded by three *tet* operator sites. This repressed state could be lifted rapidly by the application of tetracycline (Gatz *et al.* 1991).

The *lac* and *tet* repressor systems both show minimal background transcription in the presence of the appropriate repressor protein, and a high induction ratio is therefore possible. In the *lac* system the maximum induction ratio is approximately 50, whereas in the *tet* system up to 500-fold induction has been achieved (Figge *et al.* 1988, Gatz *et al.* 1992). Remarkably, the bacterial repressor proteins appear quite capable of interacting with the eukaryotic transcriptional apparatus and functioning as they do in bacteria, despite the many mechanistic differences in transcriptional control between prokaryotes and eukaryotes.

The tet *activator and reverse activator systems*

A disadvantage of repressor-based systems is that, in order to function effectively, high-level constitutive expression of repressor molecules is required to suppress background transgene activity. However, both the LacI and TetR proteins are cytotoxic at high levels.

To address these problems, TetR and LacI have been converted into activators by generating fusion proteins, in which the repressor is joined to the herpes simplex virus (HSV) VP16 transactivator (Labow *et al.* 1990, Gossen & Bujard 1992). In these systems, only the DNA-binding specificity of the repressor proteins is exploited. The binding of the modified bacterial proteins to operator sites within the transgene leads to transcriptional activation, because the VP16 protein acts positively on the transcriptional apparatus. The operator sites have effectively become enhancers and the inducers (IPTG and tetracycline) have effectively become repressors (Fig. 13.2). The *tet* transactivator (tTA) system has been more

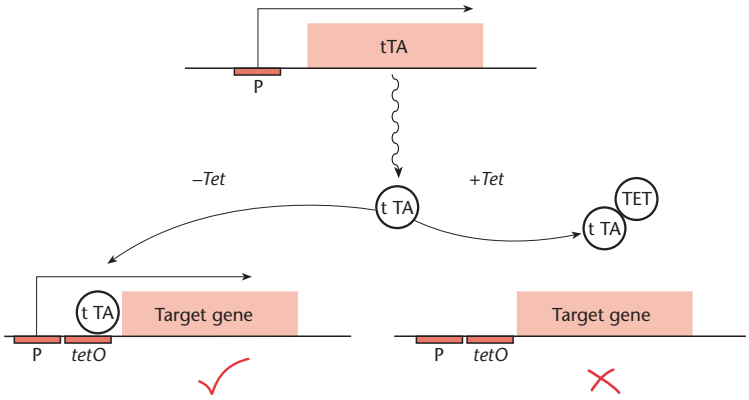


Fig. 13.2 The *tet* transactivator system (tTA).

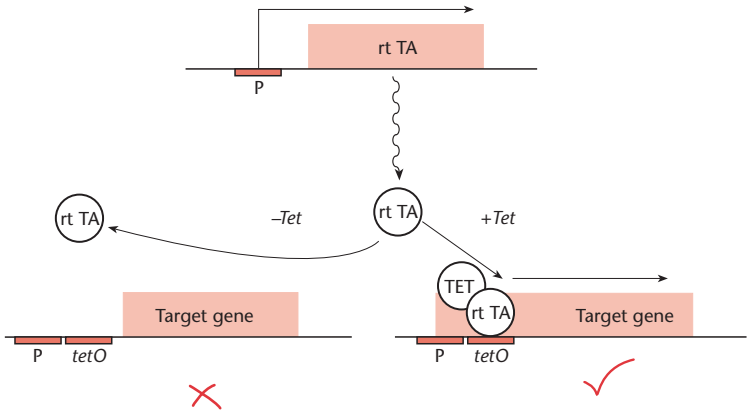


Fig. 13.3 The reverse *tet* transactivator system (rtTA).

widely used than the equivalent *lac* system, because very high levels of IPTG are required to inhibit LacI binding in mammalian cells and this is toxic (Figge *et al.* 1988). Many different proteins have been produced in mammalian cell lines using the tTA system, particularly cytotoxic proteins, whose constitutive expression would rapidly lead to cell death (Wu & Chiang 1996). In cells, a low background activity has been reported and an induction ratio of approximately 10^5 can be achieved. However, toxic effects of prolonged tetracycline exposure have been reported in transgenic animals, as well as unequal uptake of tetracycline into different organs, resulting in fluctuating basal transcription levels and cell-specific effects (reviewed by Saez *et al.* 1997).

A further modification to the *tet* system has led to marked improvements. A mutated form of the tTA protein has been generated whose DNA-binding activity is *dependent on* rather than abolished by tetracycline (Gossen *et al.* 1995). This protein is called reverse tTA (rtTA) and becomes an activator in the *presence* of tetracycline. In this system, the antibiotic is once again an inducer, but there is no requirement for prolonged exposure (Fig. 13.3). An early example of the use of this system is described by Bohl *et al.* (1997). Myoblasts were transformed with the rtTA system using a retroviral vector in which the erythropoietin complementary DNA (cDNA) was placed under the control of a tetracycline-inducible promoter. These cells were implanted into mice, and erythropoietin secretion could be controlled by feeding the mice doxycycline, a derivative of tetracycline with a shorter half-life. An important finding was that long-term control of the secretion of this hormone was possible, with important implications for the use of inducible expression systems for gene

therapy. A number of convenient plasmid- and virus-based vectors have been developed, allowing the components of the rtTA system to be transferred to the expression host in one experiment (for example, see Hofmann *et al.* 1996, Paulus *et al.* 1996, Schultze *et al.* 1996). A reverse-transactivator *lac* system has also been developed, with an induction ratio of approximately 10^4 (Baim *et al.* 1991).

Heterologous use of steroids: Drosophila *ecdysone in mammals and mammalian glucocorticoids in plants*

Steroid hormones are lipophilic molecules that penetrate cells rapidly and are eliminated within a few hours. The use of heterologous steroids for inducible transgene expression is therefore advantageous, because, in addition to their favourable kinetics, such molecules should not activate endogenous signalling pathways in the expression host and should therefore have limited toxicity.

Ecdysone is a steroid hormone found only in insects and is responsible for the extensive morphological changes that occur during moulting. As with other steroid-like signalling molecules, the hormone acts through a heterodimeric transcription factor of the nuclear receptor family. In *Drosophila*, this receptor comprises the products of the genes *ecdysone receptor* (*ecr*) and *ultraspiracle* (*usp*). The hormone and its signalling pathway are not found in mammalian cells. Therefore, transgenes including an ecdysone response element in the promoter can be induced by exogenously supplied ecdysone or its analogue, muristerone A, in cells or transgenic animals expressing the components of the *Drosophila* receptor. The unmodified *Drosophila* system has a poor induction ratio, but this can be improved using chimeric receptors and mammalian components (Yao *et al.* 1992, 1993). In a significant improvement, No and colleagues were able to achieve an induction ratio of 10^4 by generating a hybrid system in which the ecdysone receptor gene was expressed as a fusion with the HSV VP16 transactivator, and the ultraspiracle protein was replaced with a mammalian homologue, the retinoid X receptor. Background activity was reduced to near zero by altering the DNA sequence recognized by the hybrid receptor (No *et al.* 1996).

Although glucocorticoid induction has been used as an endogenous inducible expression system in mammalian cells, mammalian glucocorticoid receptors do not function properly in transgenic plants. However, the ligand-binding domain can be expressed as a fusion with other transcription factors and used to bring the activity of such transcription factors under inducible control. A potentially very powerful system has been described in which the glucocorticoid-receptor steroid-binding domain has been fused to the DNA-binding domain of the yeast transcription factor GAL4 and the VP16 transactivation domain. In this system, a CaMV 35S promoter modified to contain six GAL4-recognition sites is used to drive transgene expression. Genes placed under the control of this promoter can be induced 100-fold in the presence of dexamethasone (Aoyama & Chua 1997). An ecdysone-based system for plants has also been described, in which an agrochemical acts as the inducer, such that chemicals used in the field could function as inducers (Martinez *et al.* 1999).

Chemically induced dimerization

A novel strategy for inducible transgene regulation has been developed, exploiting essentially the same principles as the yeast two-hybrid system (p. 169). This technique, termed *chemically induced dimerization* (CID), involves the use of a synthetic divalent ligand to simultaneously bind and hence bring together separate DNA-binding and transactivation domains to generate a functional transcription factor. The initial system utilized the immunosuppressant drug FK-506. This binds with high specificity to an immunophilin protein called FKBP12, forming a complex that suppresses the immune system by inhibiting the maturation of T cells (reviewed by Schreiber 1991). For transgene induction, an artificial homodimer of FK-506 was created, which could simultaneously bind to two immunophilin domains. Therefore, by expressing fusion proteins in which the GAL4 DNA-binding domain and the VP16 transactivator were each joined to an immunophilin domain, the synthetic homodimer could recruit a functional hybrid transcription factor capable of activating any transgene carrying GAL4 recognition elements (Belshaw *et al.* 1996).

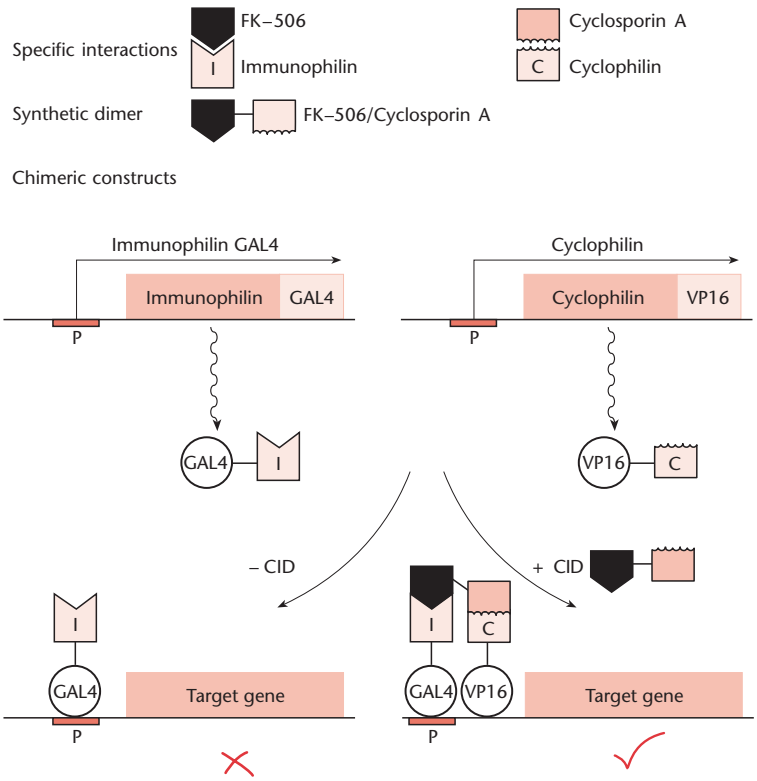


Fig. 13.4 Overview of chemically induced dimerization between the synthetic FK-506/cyclosporin A conjugate and fusion proteins containing immunophilin and cyclosporin domains. Dimerization assembles a functional transcription factor that can activate a promoter with GAL4 response elements.

Since this homodimer can also recruit non-productive combinations (e.g. two GAL4 fusions), an improved system has been developed using an artificial heterodimer specific for two different immunophilins (Belshaw *et al.* 1996). In this case, FK-506 was linked to cyclosporin A, a drug that binds specifically to a distinct target, cyclophilin. This heterodimer was shown to effectively assemble a transcription factor comprising an FKBP12–GAL4 fusion and a cyclophilin–VP16 fusion, resulting in strong and specific activation of a reporter gene in mammalian cells (Fig. 13.4). A more versatile system has been developed that exploits the ability of another immunosupressant drug, rapamycin, to mediate the heterodimerization of FKBP12 and a kinase known as FRAP (Rivera *et al.* 1996). In this system, FKBP12 and FRAP are each expressed as fusions with the components of a functional transcription factor. In the absence of rapamycin there is no interaction between these fusions, but when the drug is supplied they assemble into a hybrid transcription factor that can activate transgene expression. Transgenic mice containing a growth-hormone gene controlled by a CID-regulated promoter showed no expression in the absence of the inducer, but high levels of human growth hormone 24 h after induction with rapamycin (Magari *et al.* 1997). The advantage of this system is that rapamycin is rapidly taken up by cells *in vivo*, and it decays rapidly. The major disadvantage of immunosuppressant drugs as chemical inducers of dimerization is their pharmacological side-effects. An active area of current research is the design of modified CID systems that do not interact with endogenous targets in the immune system, and thus provide rapid and effective transgene induction with no side-effects (Liberles *et al.* 1997).

Inducible protein activity

The inducible expression systems discussed above are all regulated at the level of transcription, such

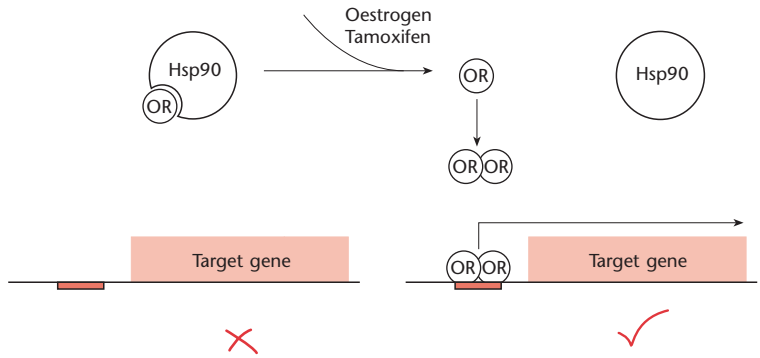


Fig. 13.5 Principle of the oestrogen-inducible expression system. The oestrogen receptor (OR) is normally sequestered by heat-shock protein 90 (Hsp90). However, in the presence of oestrogen or its analogue Tamoxifen, the receptor is released and can form dimers and activate target genes.

that there is often a significant delay between induction and response and between removal of induction and return to the basal state. Where a rapid response to induction is critical, inducible systems that operate at the post-translational level can be utilized. For example, the mammalian oestrogen receptor exists in an inert state in the absence of oestrogen, because the hormone-binding domain interacts with heat-shock protein 90 (Hsp90) to form an inactive complex. When oestrogen is present, it binds to its receptor, causing a conformational change that releases the receptor from Hsp90. The receptor is then free to dimerize and interact with DNA (Fig. 13.5). In principle, any protein expressed as a fusion with the oestrogen-binding domain will similarly interact with Hsp90 and form an inactive complex (Picard 1994). A recombinant protein can thus be expressed at high levels in an inactive state, but can be activated by feeding cells or transgenic animals with oestrogen or an analogue, such as Tamoxifen, which does not induce endogenous oestrogen-responsive genes (Littlewood *et al.* 1995). A similar system has been devised using a mutant-form progesterone receptor, which can no longer bind progesterone but can be induced with the antiprogestin RU486 (Garcia *et al.* 1992, Vegeto *et al.* 1992). An induction ratio of up to 3500 has been demonstrated in transgenic mice and, importantly, the inductive response occurs when the drug is supplied at a dose more than 100-fold below that required for it to function as an antiprogestin (Wang *et al.* 1997a,b).

Applications of site-specific recombination

Until recently, there was no generally applicable method for the precise *in vivo* manipulation of DNA sequences in animal and plant genomes. In mice, gene targeting by homologous recombination allows specific changes to be introduced into preselected genes, but it had proved impossible to extend the technique to other animals or to plants. Furthermore, even gene targeting is limited by the fact that the targeted gene is modified in the germ line; thus all cells in the mouse are similarly affected from the beginning of development and throughout its entire lifetime.

Over the last 10 years, general methods have become available that allow *in vivo* transgene manipulation in any animal or plant species. Importantly, by using such methods in concert with inducible or cell-type-specific expression systems, it is possible to generate transgenic organisms in which transgenes can be conditionally modified. In mice, the use of these methods in combination with gene targeting allows the production of conditional mutants (*conditional knockouts*), in which an endogenous gene is inactivated specifically in certain cell types or at a particular stage of development. These methods are based on a specialized genetic process, termed *site-specific recombination*.

Site-specific recombination

Site-specific recombination differs from homologous

Box 13.1 Visible marker genes

Reporter genes are widely used for *in vitro* assays of promoter activity (Box 10.1). However, reporters that can be used as cytological or histological markers are more versatile, because they allow gene expression profiles to be determined in intact cells and whole organisms.

β-galactosidase and β-glucuronidase

The *E. coli lacZ* gene encodes β-galactosidase, an enzyme that hydrolyses β-D-galactopyranosides, such as lactose, as well as various synthetic analogues. Like CAT, β-galactosidase activity can be assayed *in vitro*, although with the advantage that the assays are non-radioactive. For example, cell lysates can be assayed spectrophotometrically using the chromogenic substrate ONPG*, which yields a soluble yellow compound (Norton & Coffin 1985). Alternatively, a more sensitive fluorometric assay may be preferred, using the substrate MUG*. For histological staining, the substrate X-gal* yields an insoluble blue precipitate that marks cells brightly. The *lacZ* gene was first expressed in mammalian cells by Hall *et al.* (1983) to confirm transfection. For these experiments, the gene was linked to the SV40 early promoter and the mouse mammary tumour virus (MMTV) LTR promoter. Fusions between the *hsp70* promoter and *lacZ* were also constructed and shown to drive heat-shock-inducible β-galactosidase expression in *Drosophila* (Lis *et al.* 1983). One disadvantage of *lacZ* as a marker is that certain mammalian cells, and many plants, show a high level of endogenous β-galactosidase activity, which can obscure the analysis of chimeric genes (Helmer *et al.* 1984). The *E. coli gusA* gene, which encodes the enzyme β-glucuronidase (GUS), is an acceptable alternative (Jefferson *et al.* 1986). This marker is preferred in plants, due to the minimal background activity of the endogenous enzyme (Jefferson *et al.* 1987a), but has also been used successfully in animals (e.g. Jefferson *et al.* 1987b). Similar *in vitro* and histological assay formats to those described for β-galactosidase are also available for GUS, e.g. a histochemical substrate, X-gluc*, which yields an insoluble blue precipitate.

Luciferase

CAT, GUS and β-galactosidase are stable proteins, which persist in the cells that express them. One problem with stable reporter proteins is that, while they provide useful markers for gene activation, they are less useful for assaying transcriptional repression or rapid changes in gene activity. Luciferase was introduced as a novel reporter gene in 1986, for use in both plants (Ow *et al.* 1986) and animals (De Wet *et al.* 1987). The original marker gene, *luc*, was isolated from the North American firefly *Photinus pyralis* and encoded a single polypeptide of 550 amino acids. The enzyme catalyses the oxidation of luciferin, in a reaction requiring oxygen, ATP and the presence of magnesium ions. When excess substrate is supplied, a flash of light is emitted that is proportional to the amount of enzyme present. This can be detected using a luminometer, a scintillation counter as a luminometer or even photographic film (Wood & DeLuca 1987). Important advantages of the luciferase system include its very high sensitivity (more than 100-fold more sensitive than *lacZ*) and the rapid decay of light emission. Luciferase has therefore been used to analyse the activity of genes with oscillating expression profiles, such as the *Drosophila period* gene (Brandes *et al.* 1996). The amenability of the luciferase system has been expanded by the isolation of alternative luciferases from other organisms, which bioluminesce in different colours (e.g. see Thompson *et al.* 1990). A bacterial luciferase gene, *luxA*, has also been used as a marker in transgenic plants (Koncz *et al.* 1987).

Green fluorescent protein

The most recent addition to the growing family of reporters is green fluorescent protein (GFP), from the jellyfish *Aequoria victoria*. Over the last 5 years, this remarkable molecule has emerged as one of the most versatile tools in molecular and cellular biology and is being used to investigate an increasing variety of biological processes in bacteria, yeast, animals and plants (reviewed by Tsien 1998, Haseloff *et al.*

continued

Box 13.1 *continued*

1999, Ikawa *et al.* 1999, Naylor 1999). GFP is a bioluminescent marker that causes cells to emit bright green fluorescence when exposed to blue or ultraviolet light. However, unlike luciferase, GFP has no substrate requirements and can therefore be used as a vital marker to assay cellular processes in real time. Other advantages of the molecule include the fact that it is non-toxic, it does not interfere with normal cellular activity and it is stable even under harsh conditions (Ward & Bokman 1982).

GFP was first used as a heterologous marker in *Caenorhabditis elegans* (Chalfie *et al.* 1994). However, early experiments with GFP expression in a variety of other organisms, including *Drosophila* (Wang & Hazelrigg 1994), mammalian cell lines (Marshall *et al.* 1995) and plants (Haseloff & Amos 1995, Hu & Chen 1995, Sheen *et al.* 1995), identified a number of difficulties in the heterologous expression of the *gfp* gene. Modifications have been necessary for robust GFP expression in some plants (Chiu *et al.* 1996). In *Arabidopsis*, for example, the original *gfp* gene is expressed very poorly due to aberrant splicing. This problem was addressed by removing a cryptic splice site recognized in this plant (Haseloff *et al.* 1997). The original *gfp* gene has been extensively modified to alter various properties of the protein, such as the excitation and emission wavelengths, to increase the signal strength and to reduce photobleaching (e.g. Heim & Tsein 1996, Zolotukhin *et al.* 1996, Cormack *et al.* 1997). As a result, a range of variant GFPs are available which can be used for dual labelling (e.g. Tsien & Miyawaki 1998; reviewed by Ellenberg *et al.* 1999). Fluorescent proteins of other colours are also available, including a red fluorescent protein from *Anthozoa* (Matz *et al.* 1999). A mutant form of this protein changes from green to red fluorescence over time, allowing it to be used to characterize temporal gene expression patterns (Terskikh *et al.* 2000).

GFP is particularly useful for generating fusion proteins, providing a tag to localize recombinant proteins in the cell. This facilitates the investigation of intracellular protein trafficking, and even the transport of proteins between cells. An early example of this application was the use of GFP to monitor the movement of ribonucleprotein particles during oogenesis in *Drosophila* (Wang & Hazelrigg 1994). Kohler *et al.* (1997) have used GFP to study the exchange of molecules between plant organelles, while Wacker *et al.* (1997) have investigated the transport of a GFP-tagged protein along the secretory pathway. The use of GFP to study the real-time dynamics of a systemic viral infection in plants was described by Padgett *et al.* (1996).

* Abbreviations: ONPG O-nitrophenyl-β-D-galactopyranoside; MUG 4-methylumbelliferyl-β-D-galactoside; X-gal: 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; X-gluc: 5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid.

recombination in several important respects. In terms of gene manipulation, the most important differences between these processes concern the availability of the recombinase and the size and specificity of its target sequence. Homologous recombination is a ubiquitous process that relies on endogenous recombinase enzymes present in every cell, whereas site-specific recombination systems are very specialized and different systems are found in different organisms. Homologous recombination occurs between DNA sequences with long regions of homology but no particular sequence specificity, whereas site-specific recombination occurs at short, specific recognition sites. This means that target sites for site-specific recombination can be introduced easily and unobtrusively into transgenes, but recombination will only occur in a heterologous cell if a source of recombinase is also supplied. The power of site-specific recombination as a tool for genome manipulation thus relies on the ability of the experimenter to supply the recombinase enzyme on a conditional basis.

A number of different site-specific recombination systems have been identified and several have been studied in detail (reviewed by Craig 1988, Sadowski 1993). Some recombinases, such as bacteriophage λ integrase, require various accessory proteins for efficient recombination. However, the simplest systems

require only the recombinase and its target sequence. Of these, the most extensively used are Cre recombinase from bacteriophage P1 (Lewanodski & Martin 1997) and FLP recombinase (flippase) from the 2 µm plasmid of the yeast *Saccharomyces cerevisiae* (Buchholz *et al.* 1998). These have been shown to function in many heterologous eukaryotic systems including mammalian cells and transgenic animals and plants (reviewed by Sauer 1994, Ow 1996, Metzger & Feil 1999). Both recombinases recognize 34 bp sites (termed *loxP* and *FRP*, respectively) comprising a pair of 13 bp inverted repeats surrounding an 8 bp central element. *FRP* possesses an additional copy of the 13 bp repeat sequence, although this has been shown to be non-essential for recombination. Cre recombinase has been used most extensively in mammals, because it works optimally at 37°C. The optimal temperature for FLP recombinase is 30°C (Buchholz *et al.* 1996).

Site-specific deletion of transgene sequences

Site-specific deletion of unwanted transgenes

The reaction catalysed by Cre recombinase is shown in Fig. 13.6. If two *loxP* sites are arranged as direct repeats, the recombinase will delete any intervening DNA, leaving a single *loxP* site remaining in the genome. If the *loxP* sites are arranged as inverted repeats, the intervening DNA segment is inverted. Both reactions are reversible. However, when the *loxP* sites are arranged in the same orientation, excision is favoured over reintegration, because the excised DNA fragment is rapidly degraded.

The ability of flanking *loxP* sites to delineate any sequence of interest for site-specific deletion has numerous applications. The most obvious of these is the deletion of unwanted sequences, such as marker genes. This approach has been used, for example, as a simplified strategy to generate targeted mutant mice containing point mutations. Recall from Chapter 11 that traditional strategies for generating subtle mutants in mice involve two rounds of homologous recombination in embryonic stem cells (ES cells) (p. 207). Such strategies are very inefficient, because homologous recombination is a rare event. However, in the Cre recombinase-based approach, a second round of homologous recombination is unnecessary (Kilby *et al.* 1993). As shown in Fig. 13.7, gene targeting is used to replace the wild-type allele of a given endogenous gene with an allele containing a point mutation, and simultaneously to introduce markers, such as *neo* and *Tk*, for positive and negative selection. The positive–negative selection markers within the homology region are flanked by *loxP* sites.

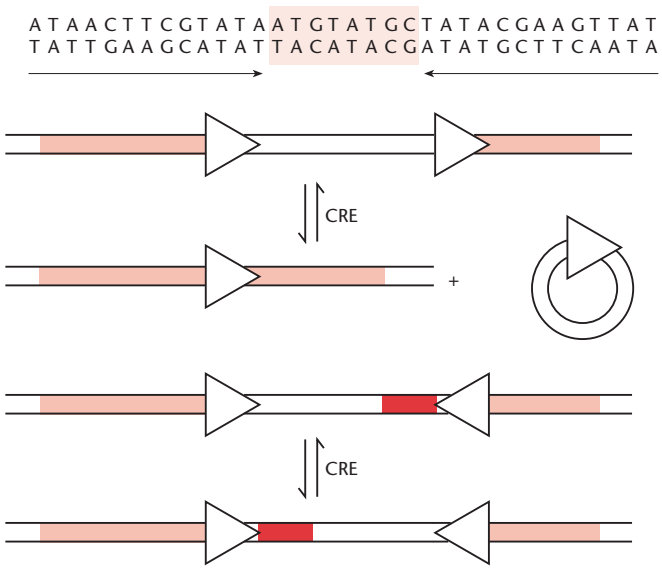


Fig. 13.6 Structure of the *loxP* site and reactions catalysed by Cre recombinase when paired *loxP* sites, shown as arrows, are arranged in different orientations.

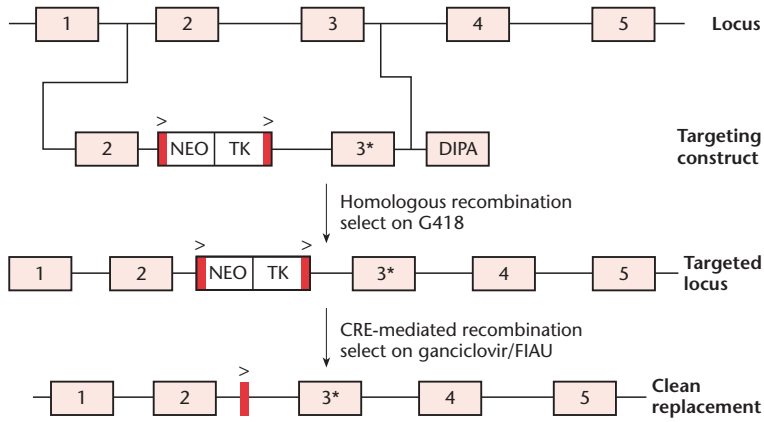


Fig. 13.7 Gene targeting followed by marker excision, catalysed by Cre recombinase. Initially, positive and negative markers (*neo* and *Tk*), flanked by *loxP* sites, are introduced by homologous recombination (a second negative marker, in this case encoding diphtheria toxin, is included outside the homology region to eliminate random integration events). Following selection for *neo* on G418, Cre recombinase is used to excise both markers, leaving a single *loxP* site remaining in the genome. The excision event can be identified by selection for the absence of *Tk*, using ganciclovir or FIAU. Asterisk indicates mutation.

A second negative marker (e.g. the gene for diphtheria toxin) is included outside the homology region to select against random integration events. Cells that have lost the diphtheria-toxin gene and survive selection for *neo* are likely to represent authentic targeting events. Such cells are then transfected with a plasmid expressing Cre recombinase, which catalyses the excision of the remaining markers, leaving a clean point mutation and no evidence of tinkering except for a single *loxP* site remaining in one intron. Negative selection using ganciclovir or 1-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)-5-iodouracil (FIAU) identifies cells that have lost the markers by selection against *Tk*.

Similar strategies can be used to remove marker genes from plants, as first demonstrated by Dale and Ow (1991). These investigators used *Agrobacterium* to transform tobacco-leaf explants with a CaMV 35S-luciferase reporter construct. The transfer DNA (T-DNA) also contained a selectable marker for hygromycin resistance, flanked by *loxP* sites. Transgenic plants were regenerated under hygromycin selection and leaf explants from these plants were then transformed with a second construct, in which Cre recombinase was driven by the CaMV 35S promoter. This construct also contained the *nptII* gene and the second-round transgenic plants were selected on kanamycin. Ten of the 11 plants tested were found to be hygromycin-sensitive, even though they continued to express luciferase, showing that the original marker had been excised. Since the *cre*/*nptII* construct was introduced separately, it was not linked to the original T-DNA and segregated in future generations, leaving 'clean' transgenic plants containing the luciferase transgene alone.

Site-specific transgene activation and switching

While commonly used as a method to inactivate transgenes by deletion, site-specific recombination can also activate transgenes or switch between the expression of two transgenes (Fig. 13.8). In one method, termed *recombinase-activated gene expression* (RAGE), a blocking sequence, such as a polyadenylation site, is placed between the transgene and its promoter, such that the transgene cannot be expressed. If this blocking sequence is flanked by *loxP* sites, Cre recombinase can be used to excise the sequence and activate the transgene.

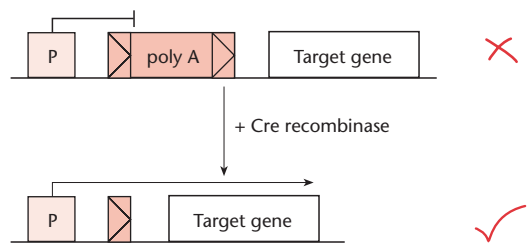


Fig. 13.8 Overview of the recombinase activation of gene expression (RAGE) strategy. A polyadenylation signal is inserted between the promoter and target gene, blocking its expression. However, if this signal is flanked by *loxP* sites, Cre recombinase can be used to excise the block, bringing the promoter and gene into juxtaposition and thus activating gene expression.

This strategy was first used in transgenic mice to study the effect of SV40 T-antigen expression in development (Pichel *et al.* 1993). In this case, Cre recombinase was expressed under the control of a developmentally regulated promoter. Essentially the same strategy was used in transgenic tobacco plants to activate a reporter gene in seeds (Odell *et al.* 1994). In this case, Cre recombinase was expressed under the control of a seed-specific promoter. An important feature of both these experiments was the use of two separate transgenic lines, one expressing Cre recombinase in a regulated manner and one containing the target gene. Crosses between these lines brought both transgenes together in the hybrid progeny, resulting in the conditional activation of the transgene based on the expression profile of Cre. This is an extremely versatile and widely used strategy, because it allows 'mix and match' between different Cre transgenic and 'responder' lines. We return to this subject below.

Site-specific transgene integration

Site-specific integration of transgenes can occur if the genome contains a recombinase recognition site. This may be introduced by random integration or (in mice) by gene targeting. Using an unmodified Cre–*loxP* system, transgene integration occurs at a low efficiency, because, as discussed above, the equilibrium of the reaction is shifted in favour of excision. Initial attempts to overcome this problem by providing transient Cre activity had limited success (see Sauer & Henderson 1990, Baubonis & Saur 1993). However, high-efficiency Cre-mediated integration has been achieved in plants (Albert *et al.* 1995) and mammalian cells (Feng *et al.* 1999) using mutated or inverted *loxP* sites. Site-specific transgene integration into mammalian cells has also been achieved using FLP recombinase (O'Gorman *et al.* 1991).

Transgene integration by site-specific recombination has many advantages over the random integration that is normally achieved by illegitimate recombination. For example, if a region of the genome can be identified that is not subject to negative position effects (Box 11.1), transgenic lines with a *loxP* site at this position can be used for the stable and high-level expression of any transgene (e.g. Fukushige & Sauer 1992). Also, due to the precise nature of site-specific recombination, transgenic loci generated by this method are likely to be less complex than loci generated by random integration.

Chromosome engineering

Site-specific recombination between widely separated target sites or target sites on different chromosomes can be used to generate large deletions, translocations and other types of chromosome mutation. Chromosome engineering by site-specific recombination was first reported by Golic (1991), using FLP recombinase in *Drosophila*, but similar experiments have now been carried out in plants and mice. Precise intrachromosomal deletions can be generated in mice by two rounds of gene targeting, introducing *loxP* sites at distant sites, followed by Cre-mediated recombination (Ramirez-Solis *et al.* 1995, Li *et al.* 1996). In plants, where gene targeting is very inefficient, an ingenious scheme has been developed where *loxP* sites are arranged in tandem on a transformation construct, one inside a *Ds* transposon and one outside (Fig. 13.9). The transposon is placed between a marker gene and its promoter. When this construct is introduced into tobacco plants containing the autonomous transposon *Ac* to provide a source of transposase, the *Ds* element can excise from the transgene, as revealed by marker-gene expression. In most heterologous plants, *Ac–Ds* elements reintegrate at a position that is linked to the original site. Although the site of reintegration cannot be controlled, this nevertheless defines a large chromosomal segment that can be excised by Cre recombinase (Medberry *et al.* 1995, Osbourne *et al.* 1995). Translocations are more difficult to engineer, because interchromosomal site-specific recombination is inefficient, and inventive selection strategies are required to identify the desired products (e.g. see Qin *et al.* 1994, Smith *et al.* 1995, Van Deursen *et al.* 1995).

Cre-mediated conditional mutants in mice

In mice, gene targeting and site-specific recombination can be used in a powerful combined approach to generate conditional knockout mutants. Essentially, targeting vectors are designed so that part of a selected endogenous gene becomes flanked by *loxP*

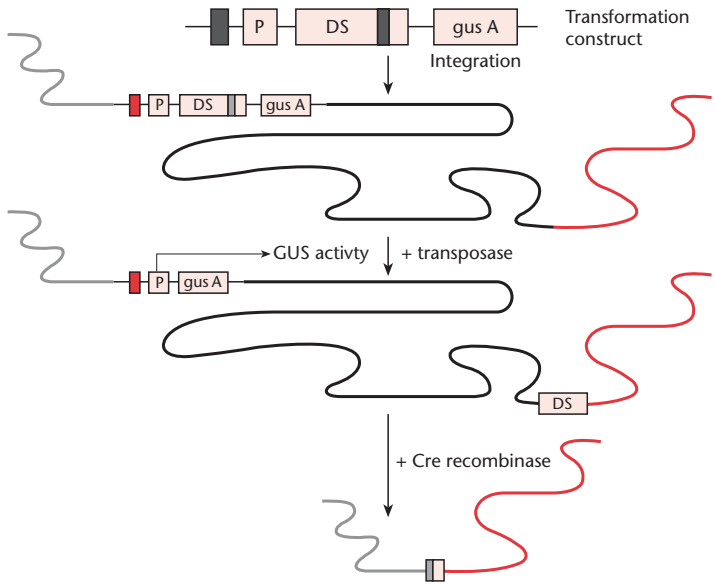


Fig. 13.9 Chromosome engineering in plants with Cre recombinase. A construct integrates into the plant genome. The construct contains a promoter separated from the reporter gene *gusA* by a *Ds* transposon containing a *loxP* site. Another *loxP* site is located just upstream of the promoter. The addition of transposase (e.g. by an *Ac* element elsewhere in the genome) causes the *Ds* element to excise, thereby bringing the *gusA* gene adjacent to the promoter. The excision event can thus be identified by the onset of GUS activity in the plant. The *Ds* element tends to insert a few tens of kilobases away on the same chromosome. Cre recombinase can then delete the large genomic region between the two *loxP* sites.

sites, or *floxed*. The usual strategy is to insert the *loxP* sites into introns flanking an essential exon, since this generally does not interfere with the normal expression of the gene. Cre recombinase is then supplied under the control of a cell-type-specific, developmentally regulated or inducible promoter, causing the gene segment defined by the *loxP* sites to be deleted in cells or at the developmental stage specified by the experimenter. This addresses a major limitation of traditional gene-knockout techniques, i.e. that, if the mutation has an embryonic lethal phenotype, only its earliest effects can be investigated.

The general methodology for such experiments, as we discussed earlier, is to cross two lines of transgenic mice, one carrying the floxed target gene and the other carrying the conditional *cre* transgene. As the number of reports of such experiments has increased, more and more transgenic mouse lines are becoming available, with Cre expressed under the control of different conditional promoters. For example, a mouse line with Cre expressed specifically in the lens of the eye was generated by Lasko *et al.* 1992. Lines are also available with Cre expressed specifically in the mammary gland (Wagner *et al.* 1997) and developing sperm (O'Gorman *et al.* 1997). Lines in which Cre is expressed in germ cells or in early development are known as 'deleter' lines and are used to remove marker genes and generate Cre-mediated constitutive gene knockouts.

In the first examples of the conditional knockout approach, Gu *et al.* (1994) generated a Cre transgenic line expressing the recombinase under the control of the *lck* promoter, such that it was expressed only in T cells. This strain was crossed to targeted mice in which part of the DNA polymerase β gene was floxed, leading to T-cell-specific excision of an essential exon. Kuhn *et al.* (1995) mutated the same gene, but they used the metallothionein promoter to express Cre recombinase, allowing induction of site-specific recombination with interferon. Although successful, this experiment highlighted many of the inadequacies of inducible promoters. There was pronounced variation in the efficiency of excision in different tissues, probably reflecting differential uptake of the inducer. Furthermore, high-level background activity of Cre was observed in the spleen, resulting in excision of the gene segment in the absence of induction, probably caused by the presence of endogenous interferons. The tTA system has been used to bring Cre expression under the control of tetracycline administration, although a high level of background activity was also observed in this experiment, resulting in excision of the target gene prior to induction (St Ogne *et al.* 1996). Tighter control has been possible using post-translational

induction. For example, Cre has been expressed as a fusion with the ligand-binding domain of the oestrogen receptor (Fiel *et al.* 1996). When this transgene was crossed into an appropriate responder strain, the background excision was minimal and Cre was strongly induced by Tamoxifen. Several strains of Cre mice are now available, in which Tamoxifen- or RU486-induced site-specific recombination has been shown to be highly efficient (e.g. Brocard *et al.* 1997, Wang *et al.* 1997a, Schwenk *et al.* 1998).

Further transgenic strategies for gene inhibition

Traditional gene-transfer strategies add new genetic information to the genome, resulting in a gain of function phenotype conferred by the transgene. Gene targeting and site-specific recombination now provide us with the ability to disrupt or delete specific parts of the mouse genome, allowing loss-of-function phenotypes to be studied, but this approach cannot be used in other animals or in plants. A range of alternative, more widely applicable transgenic strategies have therefore been developed for gene inhibition. These strategies involve the introduction of new genetic information into the genome, but, instead of conferring a gain of function, the transgene interferes with the expression of an endogenous gene, at either the RNA or the protein level. The actual target gene is not affected. The resulting loss of function effects are termed *functional knockouts* or *phenocopies*.

Gene inhibition at the RNA level

Antisense RNA transgenes

Antisense RNA has the opposite sense to mRNA. The presence of complementary sense and antisense RNA molecules in the same cell can lead to the formation of a stable duplex, which may interfere with gene expression at the level of transcription, RNA processing or possibly translation (Green *et al.* 1986). Antisense RNA is used as a natural mechanism to regulate gene expression in a number of prokaryote systems (Simons & Kleckner 1988) and, to a lesser extent, in eukaryotes (e.g. Kimelman & Kirchner 1989, Lee *et al.* 1993, Savage & Fallon 1995).

Transient inhibition of particular genes can be achieved by directly introducing antisense RNA or antisense oligonucleotides into cells. However, the transformation of cells with antisense transgenes (in which the transgene is inverted with respect to the promoter) allows the stable production of antisense RNA and thus the long-term inhibition of gene expression. This principle was established in transgenic animals and plants at about the same time. Katsuki *et al.* (1988) constructed an expression cassette in which the mouse myelin basic protein (MBP) cDNA was inverted with respect to the promoter, thus producing antisense RNA directed against the endogenous gene. In some of the transgenic mice, there was up to an 80% reduction in the levels of MBP, resulting in the absence of myelin from many axons and generating a phenocopy of the myelin-depleted 'shiverer' mutation. Smith *et al.* (1988) generated transgenic tomato plants carrying an antisense construct targeting the endogenous polygalacturonase (*pg*) gene. The product of this gene is an enzyme that causes softening and leads to overripening. The levels of *pg* mRNA in transgenic plants were reduced to 6% of the normal levels and the fruit had a longer shelf-life and showed resistance to bruising.

Antisense constructs have been widely used in transgenic animals and plants for gene inhibition. However, the efficiency of the technique varies widely and the effects can, in some cases, be non-specific. In some experiments, it has been possible to shut down endogenous gene activity almost completely, as demonstrated by Erickson *et al.* (1993), who used an inverted cDNA to generate antisense RNA against the mouse *wnt-1* gene and reduced endogenous mRNA levels to 2% of normal. Conversely, Munir *et al.* (1990) designed a construct to generate antisense RNA corresponding to the first exon and intron of the mouse *Hprt* gene, and observed no reduction in endogenous mRNA levels at all, even though the presence of antisense RNA was confirmed. The level of inhibition apparently does not depend on the size of the antisense RNA or the part of the endogenous gene to which it is complementary. For example, Moxham *et al.* (1993) achieved a 95% reduction in the level of $G_{\alpha i2}$ protein through the expression of antisense RNA corresponding to only 39 bp of the gene's 5′ untranslated region.

Conditional gene silencing can be achieved by placing antisense constructs under the control of an inducible promoter. The expression of antisense c-*myc* under the control of the MMTV LTR promoter resulted in the normal growth of transformed cells in the absence of induction, but almost complete growth inhibition in the presence of dexamethasone (Sklar *et al.* 1991). Experiments in which the tTA system was used to control antisense expression in plants have also been reported (e.g. Kumar *et al.* 1995).

Ribozyme constructs

Ribozymes are catalytic RNA molecules that carry out site-specific cleavage and (in some cases) ligation reactions on RNA substrates. The incorporation of ribozyme catalytic centres into antisense RNA allows the ribozyme to be targeted to particular mRNA molecules, which are then cleaved and degraded (reviewed by Rossi 1995, James & Gibson 1998). An important potential advantage of ribozymes over antisense RNA is that ribozymes are recycled after the cleavage reaction and can therefore inactivate many mRNA molecules. Conversely, antisense inhibition relies on stoichiometric binding between sense and antisense RNA molecules.

The use of ribozyme constructs for specific gene inhibition in higher eukaryotes was established in *Drosophila*. In the first such report, Heinrich *et al.* (1983) injected *Drosophila* eggs with a P-element vector containing a ribozyme construct targeted against the *white* gene. They recovered transgenic flies with reduced eye pigmentation, indicating that expression of the endogenous gene had been inhibited. A ribozyme construct has also been expressed under the control of a heat-shock promoter in *Drosophila* (Zhao & Pick 1983). In this case, the target was the developmental regulatory gene *fushi tarazu* (*ftz*). It was possible to generate a series of conditional mutants with *ftz* expression abolished at particular stages of development, simply by increasing the temperature to 37°C.

Ribozymes have also been used in mammalian cell lines, predominantly for the study of oncogenes and in attempts to confer resistance to viruses (reviewed by Welch *et al.* 1998). There has been intensive research into ribozyme-mediated inhibition of HIV, and remarkable success has been achieved using retroviral vectors, particularly vectors carrying multiple ribozymes (reviewed by Welch *et al.* 1998, Muotri *et al.* 1999). So far, there have been relatively few reports of ribozyme expression in transgenic mice. Larsson *et al.* (1994) produced mice expressing three different ribozymes targeted against β_2-macroglobulin mRNA, and succeeded in reducing endogenous RNA levels by 90%. Tissue-specific expression of ribozymes has also been reported. A ribozyme targeted against glucokinase mRNA was expressed in transgenic mice under the control of the insulin promoter, resulting in specific inhibition of the endogenous gene in the pancreas (Efrat *et al.* 1994). Recently, retroviral delivery of anti-neuregulin ribozyme constructs into chicken embryos has been reported (Zhao & Lemke 1998). Inhibition of neuregulin expression resulted in embryonic lethality, generating a very close phenocopy of the equivalent homozygous null mutation in mice.

Cosuppression

Cosuppression refers to the ability of a sense transgene to suppress the expression of a homologous endogenous gene. This surprising phenomenon was first demonstrated in plants, in a series of experiments designed to increase the levels of an endogenous protein by introducing extra copies of the corresponding gene. In an attempt to increase the amount of pigment synthesized by petunia flowers, Napoli *et al.* (1990) produced transgenic petunia plants carrying multiple copies of the chalcone synthase (*chs*) gene. This encodes an enzyme that converts coumaroyl-CoA and 3-malonyl-CoA into chalcone, a precursor of anthocyanin pigments. The presence of multiple transgene copies was expected to increase the level of enzyme and result in deeper pigmentation. However, in about 50% of the plants recovered from the experiment, exactly the opposite effect was observed, i.e. the flowers were either pure white or variegated with purple and white sectors. Similar findings were reported by Van der Krol *et al.* (1988) using a transgene encoding another pigment-biosynthesis enzyme, dihydroflavonol-4-reductase. In both cases, it appeared that integration of multiple copies of the transgene led to the suppression of some or all of the transgenes and the cosuppression of homologous endogenous genes.

While troublesome in terms of generating plant lines with high transgene expression levels, cosuppression can also be exploited as a tool for specific gene inactivation. There have been many reports of this nature. For example, transgenic tomatoes have been produced containing a partial copy of the *pg* gene in the sense orientation (Smith *et al.* 1990). As with the antisense *pg* transgenic plants generated previously by the same group (see above), strong inhibition of the endogenous gene was achieved, resulting in fruit with a prolonged shelf-life. Cosuppression has also been demonstrated in animals (Pal-Bhadra *et al.* 1997, Bahramian & Zabl 1999, Dernberg *et al.* 2000) and is probably related to a similar phenomenon called *quelling*, which has been described in fungi (reviewed by Selker 1997, 1999).

The mechanism of cosuppression in plants is complex and can involve silencing at either the transcriptional or post-transcriptional levels (for details, see Box 13.2). One of the most remarkable

Box 13.2 Transgene silencing, cosuppression and RNA interference

Transgene silencing is a complex phenomenon, occurring in all eukaryotes, caused by the introduction of foreign nucleic acid into the cell. Typically, expression of the affected transgene is reduced or abolished, associated with increased methylation at the transgenic locus. An understanding of transgene silencing is important, first, because it is a serious impediment to the use of animal and plant systems for the expression of foreign genes and, second, because if a transgene is homologous to an endogenous gene, the endogenous gene can also be silenced by *cosuppression*. Several different forms of silencing can be distinguished.

Position-dependent silencing and sequence-dependent silencing

These forms of silencing can affect single-copy transgenes and are not, therefore, homology-dependent. Position-dependent silencing occurs where a transgene integrates into a genomic region containing heterochomatin. The repressive chromatin structure and DNA methylation can spread into the transgenic locus from the flanking genomic DNA (Matzke & Matzke 1998); therefore silencing results from a negative position effect (position effects are discussed in more detail in Box 11.1). Single-copy transgenes may also be silenced, even if they integrate into a genomic region that lacks negative position effects. For example, integrated retrovirus vectors often undergo *de novo* silencing associated with increased levels of DNA methylation (Jahner *et al.* 1982) and, indeed, this methylation may spread outwards into flanking host DNA and inactivate linked genes (Jahner & Jaenisch 1985). Many unrelated transgenes in animals and plants have been subject to this type of silencing, suggesting that a specific sequence is not responsible. It is possible that eukaryotic genomes possess mechanisms for scanning and identifying foreign DNA sequences, perhaps based on their unusual sequence context, and then inactivating them by methylation (Kumpatla *et al.* 1998). Prokaryotic DNA may be recognized in this manner, since prokaryotic sequences act as a strong trigger for *de novo* methylation, e.g. in transgenic mice (Clark *et al.* 1997).

Homology-dependent gene silencing (HDGS)

HDGS is caused by the presence of multiple transgene copies (reviewed by Gallie 1998, Grant 1999, Plasterk & Ketting 2000, Hammond *et al.* 2001). The suppression of transgene expression can occur at the transcriptional or post-transcriptional levels, and homologous endogenous genes are often cosuppressed. Single-copy transgenes homologous to an endogenous gene may, in some cases, also be sufficient to induce cosuppression. In *transcriptional gene silencing* (TGS), no mRNA is produced from the silenced genes and affected loci act as nucleation points for heterochromatin formation and DNA methylation. In *post-transcriptional gene silencing* (PTGS), transcription is actually required for silencing to take place, and it induces the degradation of mRNA, so that very little accumulates in the

continued

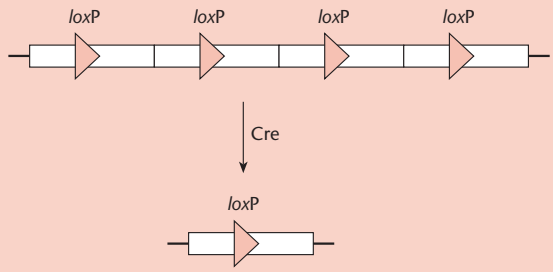


Fig. B13.1 An experiment to demonstrate homology-dependent gene silencing in mammals (also called repeat-induced gene silencing (RIGS)). A transgene construct containing the human β-globin cDNA was modified to contain a single *loxP* site, which is recognized by Cre recombinase. Transgenic mice were generated carrying multiple copies of the transgene, and in these animals the locus was highly methylated and β-globin expression was low. However, when Cre recombinase was expressed, recombination between the *loxP* sites resulted in the excision of all copies of the transgene except one. Reduction in the transgene copy number resulted in increased expression, accompanied by reduced methylation at the transgenic locus. (After Garrick *et al.* 1998.)

cytoplasm. The severity of silencing often correlates with transgene copy number, and this can be demonstrated directly through the use of site-specific recombination, as shown in the figure above.

Different processes may trigger homology-dependent silencing, including the ability of homologous DNA sequences to form secondary structures and the synthesis of aberrant RNA molecules, leading to the production of dsRNA. Both processes are stimulated if the transgenic locus is complex, particularly if transgenes or partial transgenes are arranged as inverted repeats. There may be some cross-talk between the TGS and PTGS mechanisms, but this is currently unclear. In plants, transgenes carried by RNA viruses can induce PTGS, a phenomenon that has been termed virus-induced gene silencing (VIGS). Only replication-competent viruses have this effect, providing further evidence that dsRNA is the trigger for this process. The high-level expression of sense transgenes can also induce PTGS, and it has been suggested that the plant cell may recognize such aberrant transcripts and convert them into dsRNA.

PTGS and RNAi as a common mechanism – the evidence

The involvement of dsRNA in many cases of PTGS in plants suggests that the underlying basis of this phenomenon may be similar to RNA interference (RNAi), which is also a post-transcriptional process. As discussed in the text, both PTGS and RNAi are systemic, i.e. silencing can spread from the site at which it was induced throughout the entire organism. Recently, small RNA molecules, called *guide RNAs*, have been identified in plants affected by PTGS and animals subjected to RNAi, which can also spread systemically (Hamilton & Baulcombe 1999, Hammond *et al.* 2000, Parrish *et al.* 2000). It is proposed that these RNAs may assemble with certain proteins to generate a catalytic endonuclease complex that cleaves target mRNA molecules efficiently and in a homology-dependent fashion. In cases of PTGS, guide RNAs may form from aberrant transcripts generated by integrated transgenes or directly from dsRNA viral genomes.

What is silencing for?

It is possible that both homology-dependent and sequence-dependent silencing have evolved as a defence against 'invasive' nucleic acids (Yoder *et al.* 1997, Jones *et al.* 1998, Jensen *et al.* 1999, Li *et al.* 1999). This has been supported by the recent isolation of mutants in several organisms that show deficiencies in PTGS or RNAi. Animals impaired for RNAi show increased rates of transposon mobilization, whereas plants impaired for PTGS are more susceptible to viral infection. Interestingly, similar gene products have been identified in diverse organisms, providing further evidence for a link between PTGS and RNAi. A comprehensive discussion of this exciting new area of research is outside the scope of this book, but the interested reader can consult several excellent reviews on the subject (Plasterk & Ketting 2000, Hammond *et al.* 2001).

aspects of post-transcriptional gene silencing (PTGS) is that it is a systemic phenomenon, suggesting that a diffusible signal is involved. This can be demonstrated by grafting a non-silenced transgenic scion on to a silenced transgenic host. The silencing effect is able to spread into the graft, and the systemic effect works even if the two transgenic tissues are separated by up to 30 cm of wild-type stem (Palauqui *et al.* 1997, Voinnet *et al.* 1998).

PTGS in plants can be induced not only by integrated transgenes but also by RNA viruses, as long as there is a region of homology between the virus genome and an integrated gene. For example, the virus may carry a sequence that is homologous to an endogenous gene or to a transgene integrated into the host genome. The effect also works if the plant is transformed with a cDNA construct corresponding to part of the virus genome, as demonstrated by Angell and Baulcombe (1997). The rationale behind this experiment was to transform plants with a cDNA construct corresponding to a chimeric potato virus X (PVX) genome containing the *gusA* reporter gene. Expression of the transgene was expected to generate very high levels of β-glucuronidase (GUS) activity, because, after transcription of the transgene, the resulting viral RNA would be amplified by the virus's own replication system. However, disappointingly, all of the transgenic plants produced extremely low levels of viral RNA and GUS activity. The plants also showed an absence of PVX symptoms and were resistant to PVX infection. The virus-induced silencing effect only worked using replication-competent vectors, suggesting that the double-stranded RNA (dsRNA) intermediate involved in viral replication was the trigger for silencing (see Box 13.2).

Such is the efficiency with which PVX RNA can silence homologous genes in the plant genome that PVX vectors have been used very successfully to generate functional knockouts in plants (reviewed by Baulcombe 1999). For example, Burton *et al.* (2000) described the infection of tobacco plants with PVX vectors containing a cDNA sequence putatively encoding a cellulose synthase. The inoculated plants showed a dwarf phenotype, and levels of cellulose in affected leaves were reduced by 25%. On the basis of this evidence, the investigators concluded that the cDNA did indeed encode such an enzyme and was capable of cosuppressing the endogenous cellulose synthase gene.

RNA interference

RNA interference (RNAi) is a novel phenomenon that has the potential to become an extremely powerful tool for gene silencing in any organism. The process was discovered by Fire *et al.* (1998) while investigating the use of antisense and sense RNA for gene inhibition in the nematode worm *Caenorhabditis elegans*. In one experiment, they introduced both sense and antisense RNA into worms simultaneously and observed a striking and specific inhibitory effect, which was approximately 10-fold more efficient than either single RNA strand alone.

Further investigation showed that only a few molecules of dsRNA were necessary for RNAi in *C. elegans*, indicating that, like the action of ribozymes, the effect was catalytic rather than stoichiometric. Interference could be achieved only if the dsRNA was homologous to the exons of a target gene, indicating that it was a post-transcriptional process. The phenomenon of RNAi appears to be quite general, and this strategy has been used more recently for gene silencing in many other organisms, including *Drosophila* (Kennerdell & Carthew 2000), mice (Wianny & Zernicka-Goetz 2000) and plants (Waterhouse *et al.* 1998, Chuang & Meyerowitz 2000). In *C. elegans*, RNAi is now the standard procedure for gene inactivation, and it is becoming increasingly favoured in *Drosophila* and plants, due to its potency and specificity. Direct injection of the RNA is unnecessary. The use of a construct containing adjacent sense and antisense transgenes producing hairpin RNA (e.g. Chuang & Meyerowitz 2000, Tavernarakis *et al.* 2000) or a single transgene with dual opposing promoters (Wang *et al.* 2000) provides a stable source of dsRNA and hence the potential for permanent gene inactivation. Like PTGS in plants, RNAi is also systemic. RNAi-mediated silencing can be achieved in *C. elegans* by injecting dsRNA into any part of the body (Fire *et al.* 1998), but, more remarkably, simply placing worms in a solution containing dsRNA or feeding them on bacteria that synthesize dsRNA is sufficient to trigger the effect (Tabara *et al.* 1998, Timmons & Fire 1998). The similarities between RNAi and PTGS in plants

suggest a common molecular basis, which we discuss briefly in Box 13.2.

Gene inhibition at the protein level

Intracellular antibodies

Antibodies bind with great specificity to particular target antigens and have therefore been exploited in many different ways as selective biochemical agents. Examples discussed in this book include the immunological screening of cDNA expression libraries (p. 109) and the isolation of recombinant proteins by immunoaffinity chromatography (p. 76). The microinjection of antibodies into cells has been widely used to block the activity of specific proteins, but the limitation of this approach is that the inhibitory effect is transient (reviewed by Morgan & Roth 1988). Specific inhibitory effects can also be achieved by microinjecting cells with RNA from hybridoma cell lines (Valle *et al.* 1982, Burke & Warren 1984). Such experiments provided the first evidence that non-lymphoid cells can synthesize and assemble functional antibodies.

To achieve long-term inhibition of specific proteins, cells can be transformed with cDNA constructs that allow the expression of intracellular antibodies (sometimes termed *intrabodies*) (Richardson & Marasco 1995). An important consideration here is that antibodies are large multimeric proteins with, in addition to antigen binding, various effector functions that are non-essential for intracellular protein inhibition. The strategy for expressing intracellular antibodies has been radically simplified using modified antibody forms, such as single-chain Fv (scFv) fragments (Fig. 13.10). These comprise the antigen-binding variable domains of the immunoglobulin heavy and light chains, linked by a flexible peptide arm. Such fragments retain the specificity of the parent monoclonal antibody, but are encoded by a single, relatively small transgene. Further modifications to the expression construct allow the antibody to be targeted to particular intracellular compartments, such as the nucleus, mitochondria or cytosol. It should be noted, however, that antibodies are normally folded and assembled in the endoplasmic reticulum (ER) and Golgi apparatus and are generally less stable in cell compartments outside the secretory pathway.

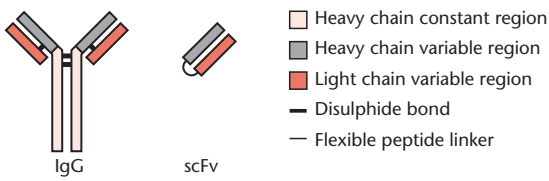


Fig. 13.10 Comparison of a normal immunoglobulin molecule with a single-chain Fv fragment.

Due to their long half-life in the ER, intracellular antibodies have been particularly useful for the inhibition of cell-surface receptors, which pass through this compartment *en route* to the plasma membrane. For example, the cell-surface presentation of functional interleukin-2 (IL2) receptors was completely abolished in Jurkat cells stably expressing an anti-IL2Rα scFv fragment in the ER, rendering these cells insensitive to exogenously applied IL2 (Richardson *et al.* 1995). More recently, the same result has been achieved using lentivirus vectors expressing the scFv fragment, demonstrating how intracellular antibodies can be valuable for gene therapy (Richardson *et al.* 1998). Intracellular antibodies have also been used to abolish the activity of oncogenes (Beerli *et al.* 1994, Cochet *et al.* 1998, Caron de Fromentel *et al.* 1999) and to confer virus resistance by inhibiting replication (reviewed by Rondon & Marasco 1997). Functional antibodies, both full-sized immunoglobulins and fragments, can also be expressed in plants. Hiatt *et al.* (1989) were the first to demonstrate the expression of plant recombinant antibodies, dubbed '*plantibodies*', and subsequent experiments have shown that this strategy can be used, as in animal cells, to combat viral diseases by targeting specific viral proteins (Conrad & Fielder 1998). Antibodies expressed in plants have also been used to interfere with physiological processes in the plant, e.g. antibodies against abscisic acid have been used to disrupt signalling by this hormone in tobacco (Artsaenko *et al.* 1995). Note that plants can also be used as bioreactors to produce recombinant antibodies for therapeutic use (Chapter 14).

Dominant-negative mutants

In diploid organisms, most loss-of-function mutations generate recessive or semidominant (dosage-related)

phenotypes, because the remaining wild-type copy of the gene provides enough gene product for normal or near-normal activity. However, some loss-of-function mutations are fully dominant over the wild-type allele, because the mutant gene product interferes with the activity of the wild-type protein. Such mutants are known as *dominant negatives*, and principally affect proteins that form dimers or larger multimeric complexes.

The deliberate overexpression of dominant-negative transgenes can be used to swamp a cell with mutant forms of a particular protein, causing all functional molecules to be mopped up into inactive complexes. The microinjection of DNA constructs or *in vitro*-synthesized dominant-negative RNA into *Xenopus* embryos has been widely used to examine the functions of cell-surface receptors in development, since many of these are dimeric (e.g. see Amaya *et al.* 1991, Hemmati-Brivanlou & Melton 1992). Dominant-negative proteins stably expressed in mammalian cells have been used predominantly to study the control of cell growth and proliferation. A dominant-negative ethylene receptor from *Arabidopsis* has been shown to confer ethylene insensitivity in transgenic tomato and petunia. The effects of transgene expression included delayed fruit ripening and flower senescence (Wilkinson *et al.* 1997).

Transgenic technology for functional genomics

In the last 5 years, the complete sequences of the genomes of several important model eukaryotic species have been published: the yeast *S. cerevisiae* (Mewes *et al.* 1997), the nematode worm *C. elegans* (*C. elegans* Sequencing Consortium 1998), the fruit fly *Drosophila melanogaster* (Adams *et al.* 2000), the model plant *Arabidopsis thaliana* (*Arabidopsis* Genome Initiative 2000) and, most recently, the human genome itself (International Human Genome Sequencing Consortium 2001, Venter *et al.* 2001). With the wealth of information that has been generated by these genome projects, the next important step is to find out what all the newly discovered genes actually do. This is the burgeoning field of *functional genomics*, which aims to determine the function of all the transcribed sequences in the genome (the *transcriptome*) and all the proteins that are made (the *proteome*).

In other parts of the book, we have already discussed a number of functional genomics techniques, including microarrays (p. 116) and variants of the yeast two-hybrid system (p. 169). However, perhaps the most important way to establish a gene's function is to see what happens when that gene is either mutated or inappropriately expressed in the context of the whole organism. This chapter has focused on the development of novel transgenic strategies for the analysis of gene function in animals and plants, but such approaches are only applicable to individual genes. For functional genomics, technologies must be available that allow the high-throughput analysis of gene function, which is the only way we can begin to understand what the 10 000–40 000 genes in the genomes of higher eukaryotic cells are for. Such technologies are discussed below.

Insertional mutagenesis

Traditional techniques for generating mutations involve the use of radiation or chemical mutagens. These tend to generate point mutations, and isolating the genes corresponding to a particular mutant phenotype can be a laborious task, particularly in the large genomes of vertebrates and plants.

An alternative is to use DNA as a mutagen. The genomes of most animals and plants contain *transposable elements* – DNA elements that have the ability to jump from site to site in the genome, occasionally interrupting genes and causing mutations. Some of the gene-silencing methods discussed in the previous section appear to be based on defence strategies, many involving DNA methylation, that exist to resist the movement of such elements in the genome. However, if these endogenous transposable elements *can* be mobilized at a sufficient frequency, they can be used to deliberately interrupt functional genes and generate insertional mutants. Importantly, populations of animals or plants carrying transposable elements can be set up for saturation mutagenesis, such that a suitable number of transposition events is induced to theoretically interrupt each gene in the genome at least once somewhere in the population (e.g. see Walbot 2000).

Endogenous transposable elements are not the only sequences that cause insertional mutagenesis. Foreign DNA introduced into a cell may occasionally integrate into an existing gene and disrupt its expression. Insertional mutants have therefore been identified as by-products of other gene-transfer experiments. As an example, we consider the study of Yokoyama *et al.* (1993). They introduced a transgene encoding the enzyme tyrosinase into the male pronuclei of fertilized mouse eggs and recovered a number of founders that went on to produce transgenic lines. In one line, they found a unique and unexpected phenotype – situs inversus (the reversal of left–right asymmetry among the developing organs) and inversion of embryonic turning. This phenotype had nothing to do with tyrosinase activity, since the other lines were not similarly affected, but represented integration of the transgene into a locus required for correct specification of the left–right axis. Generally, aberrant results such as these are not reported unless they warrant investigation in their own right, so insertional mutants are likely to be produced in many unrelated transgenic experiments. A number of serendipitous cases of insertional mutagenesis in transgenic mice, caused by either pronuclear microinjection or retroviral transfer of foreign DNA, have been reviewed (Rijkers *et al.* 1994).

Gene tagging

The use of transposable elements or transgenes for insertional mutagenesis is advantageous over traditional mutagenesis methods, because the interrupted locus becomes 'tagged' with a unique sequence. Cloning genes tagged with endogenous transposons can sometimes be complicated if there are many copies of the element in the genome, but the use of heterologous transposons or transgenes can provide a truly unique sequence tag that identifies the interrupted gene. A genomic library can therefore be generated from the mutant, and the interrupted gene can be identified using hybridization or the polymerase chain reaction (PCR) to identify clones containing the tag. Such clones will also contain flanking DNA from the interrupted genomic locus.

Compared with chromosome walking and similar techniques (p. 107), tagging is a simple procedure that homes in directly on the mutant gene. Several

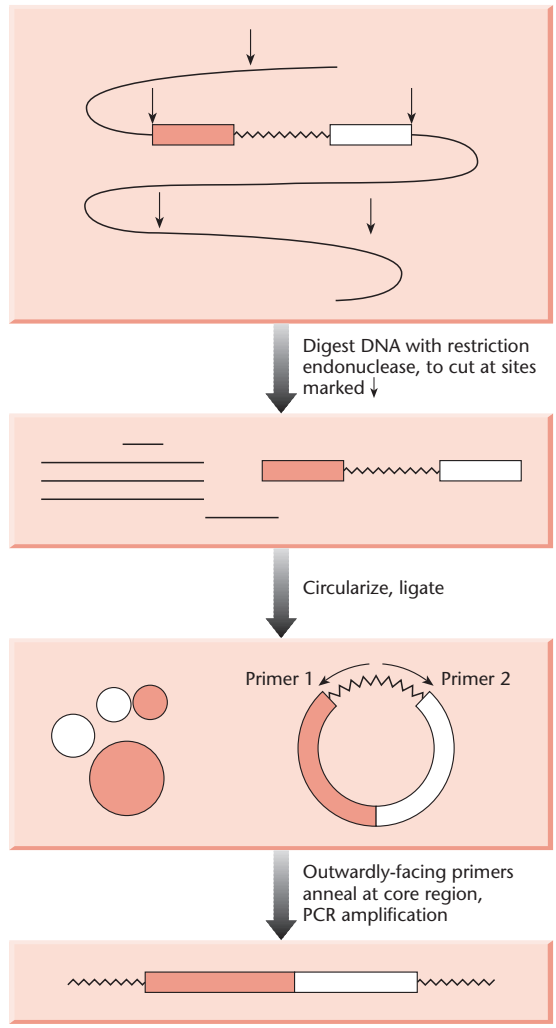


Fig. 13.11 Inverse PCR. The core region is indicated by the wavy line. Restriction sites are marked with arrows, and the left and right regions which flank the core region are represented by closed and open boxes. Primers are designed to hybridize with core sequences and are extended in the directions shown. PCR amplification generates a linear fragment containing left and right flanking sequences.

PCR-based techniques can be used to directly amplify genomic DNA flanking a transposon or transgene tag, thus avoiding the necessity for library construction in phage-λ or cosmid vectors (reviewed by Maes *et al.* 1999). The inverse PCR strategy is shown in Fig. 13.11 (Ochman *et al.* 1988). Genomic DNA from the mutant is digested with a restriction enzyme for which there is no site in the insertional tag. The DNA

is thus divided into fragments, one of which will contain the entire insertion sequence flanked by genomic DNA tails. The tails, however, have compatible sticky ends and in a suitably dilute ligation reaction can be self-ligated to generate a circular genomic fragment. PCR primers annealing at the edges of the transposon tag but facing *outwards* can then be used to amplify the flanking genomic sequences. Such PCR strategies have been used to generate fragment libraries of flanking sequences for the rapid identification of genes associated with particular phenotypes (Parinov *et al.* 1999, Tissier *et al.* 1999). If appropriate, these products can be used to screen traditional genomic or cDNA libraries to isolate the uninterrupted clone.

An even greater simplification of this procedure is possible if the origin of replication from a bacterial plasmid and an antibiotic-resistance gene are included in the tag (Perucho *et al.* 1980a). In this case, self-ligation of the construct as above generates an autonomously replicating plasmid containing genomic DNA. If whole preparations of genomic DNA are used to transform competent bacterial cells, only the tag-derived plasmid will replicate, while other genomic circles will be rapidly diluted from the culture. Furthermore, only bacteria carrying the antibiotic-resistance marker will survive selection, allowing pure plasmid DNA corresponding to the flanking regions of the genomic insert to be automatically prepared. This *plasmid-rescue* technique is illustrated in Fig. 13.12.

Both endogenous transposable elements and transgene insertions have therefore been used to isolate genes by tagging in a variety of species. In *D. melanogaster*, P elements have been widely used because of the high transposition frequency and the availability of strains that lack P elements, allowing lines containing individual insertions of interest to be derived by back-crossing (Bingham *et al.* 1981). *Tc*1 elements have been used for transposon tagging in *C. elegans* (e.g. Moerman *et al.* 1986), *Ac–Ds* elements have been widely used in maize (reviewed by Gierl & Saedler 1992) and *Tam* elements have been used in the snapdragon (*Antirrhinum majus*). *Tam* elements are particularly useful because the transposition frequency can be increased more than 800-fold if the temperature at which the plants are raised is reduced to 15°C (Coen *et al.* 1989). Also, it

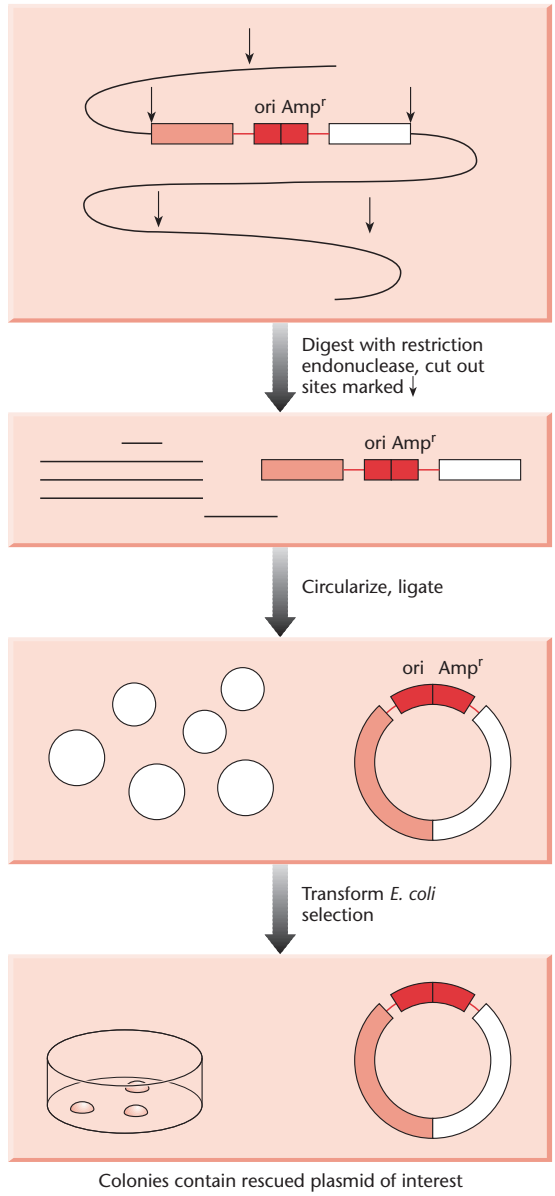


Fig. 13.12 The principle of plasmid rescue, a technique for isolating genomic DNA sequences flanking a transposon or transgene integration site.

is notable that approximately 5% of all naturally occurring recessive mutations in the mouse are caused by a particular family of retroviruses. The cloning of several mouse genes has been facilitated by virtue of their linkage to a proviral sequence (e.g. Bowes *et al.* 1993).

The transposition of *Drosophila* P elements appears to be restricted to drosophiloid insects (Handler *et al.* 1993), but other transposable elements are more promiscuous. The *Ac–Ds* elements of maize have been shown to function in a wide range of plants – dicots as well as monocots – and have been extensively used as insertional mutagens and gene tags in heterologous species (reviewed by Sundaresan 1996). T-DNA can also be used as a tag following transfer by *Agrobacterium tumefaciens*. T-DNA mutagenesis and tagging has been used in *Arabidopsis* (Feldmann 1991, Krysan *et al.* 1999) and has recently been extended to other plant systems, such as rice (Jeon *et al.* 2000).

Vectors for insertional mutagenesis and tagging

The development of vectors specifically for insertional mutagenesis and gene tagging arose directly from the use of transposons and the recovery of serendipitous insertional mutants in gene-transfer experiments. The transposable elements of several species have been developed as vectors. Modifications have been carried out in order to control the number of insertion events and to facilitate the cloning and analysis of tagged genes. The use of recombinant P-element derivatives for controlled transposon mutagenesis in *Drosophila* is a good example (reviewed by Cooley *et al.* 1988). Two fly strains are involved. One strain contains the *mutator*, a defective P element carrying useful marker genes, which can be mobilized when provided with a source of transposase. A second strain contains the *jumpstarter*, a *wings-clipped* element, which provides transposase *in trans* but lacks the *cis*-acting elements required for its own mobilization. During a controlled mutagenesis screen, a single jumpstarter element is crossed into a mutator-containing strain, whereupon transposition occurs. In subsequent generations the mutator is stabilized when the chromosome carrying the mutator element segregates from the chromosome bearing the jumpstarter element. The inclusion of a marker gene in the mutator P element allows screening or selection for transformed flies and facilitates subsequent manipulation of the modified locus. The inclusion of a bacterial selectable-marker gene and an *E. coli* origin of replication allows recovery of the tagged gene by plasmid rescue, as discussed above.

Entrapment constructs

Entrapment constructs are genetic tools that combine three important principles of gene-transfer technology: (i) the random integration of transgenes causes insertional mutagenesis and tags the mutated gene; (ii) randomly integrated DNA sequences are subject to variable position effects (Box 11.1); and (iii) reporter genes can be used to assay the activity of regulatory elements to which they become joined (Box 10.1). Entrapment constructs are insertional mutagenesis vectors adapted to provide information about the genomic region into which they integrate. Such vectors contain a reporter gene, whose expression is activated by regulatory elements in the surrounding DNA. When they integrate into the genome, the pattern of reporter-gene expression reveals the activity of nearby genes, allowing investigators to screen for appropriate or interesting insertion events. The corresponding gene can be cloned, because it is tagged by the insertion.

Enhancer traps in Drosophila

The prototype entrapment construct was the *Drosophila* enhancer trap, which originated as an important application of P-element vectors (p. 219). The method employs a *lacZ* reporter construct, in which the reporter gene is transcribed from a minimal promoter, such as a TATA box. Expression from the promoter is weak, because the promoter lacks an enhancer to stimulate its transcriptional activity. P-element-mediated transposition is used to transpose the construct into many different genomic positions in separate fly lines. In some flies, by chance, the construct is transposed to a position where it comes under the influence of an enhancer that activates transcription from the weak promoter (Fig. 13.13). It is often found in practice that, when using a histochemical stain for β-galactosidase activity, the pattern of expression shows cell specificity. Sometimes the pattern of expression is remarkably refined and detailed (O'Kane & Gehring 1987). The pattern of *lacZ* expression is assumed to reflect the cell-type specificity of the enhancer. Presumably, an endogenous gene located within range of the enhancer's effect has the same pattern of expression as the reporter. This assumption is known to be valid in

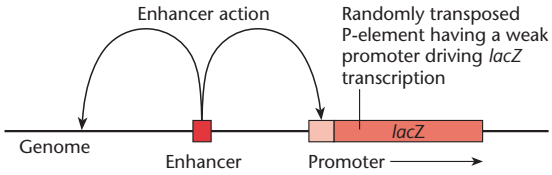


Fig. 13.13 Enhancer trapping. A reporter construct consists of *lacZ* linked to a weak promoter, which requires an enhancer for significant transcriptional activity. P-element-mediated transposition is used to insert the construct at random sites in the fly genome. When the promoter is inserted within the active range of an enhancer, expression of *lacZ* can be detected.

some cases, but the reporter activity does not always exactly match that of an endogenous gene.

Since enhancers are often found a considerable distance from the gene they activate, the enhancer trap cannot be used to directly clone genes by tagging, although it can be used as the basis for a chromosome walk to identify the gene of interest. Enhancer-trap lines have been used to identify and clone novel *Drosophila* genes, but the cell-type-specific expression that is revealed can also be harnessed in other ways. For example, instead of driving expression of a *lacZ* reporter, the principle could be applied to the expression of a toxic gene (such as ricin or diphtheria toxin), leading to cell death and thus ablation of specific cell lineages in the fly. This has enormous potential, e.g. in studies of the development of the nervous system (O'Kane & Moffat 1992).

In order to facilitate the use of enhancer trapping as a general method for driving cell-specific expression, a modified strategy has been developed. This depends upon the ability of the yeast transcription factor GAL4 to activate transgenes containing its recognition site in the heterologous environment of the fly. The enhancer-trap principle shown in Fig. 13.13 is modified so that the *lacZ* reporter gene is replaced by the coding region for GAL4 (Fig. 13.14). In such flies, GAL4 is now expressed in the pattern dictated by a local enhancer, although this cannot be seen. The pattern of GAL4 expression can be revealed by crossing the enhancer-trap line to flies carrying a reporter transgene in which the *lacZ* gene is coupled to a promoter containing GAL4-binding sites. The beauty of this system is that a bank of fly stocks with different trapped enhancers can be built up, each with a defined pattern of GAL4 expression. Once the patterns of GAL4 expression are known, crosses can be performed to introduce chromosomes containing constructs in which any desired gene is coupled to a GAL4-dependent promoter, giving a particular cell-specific pattern of expression.

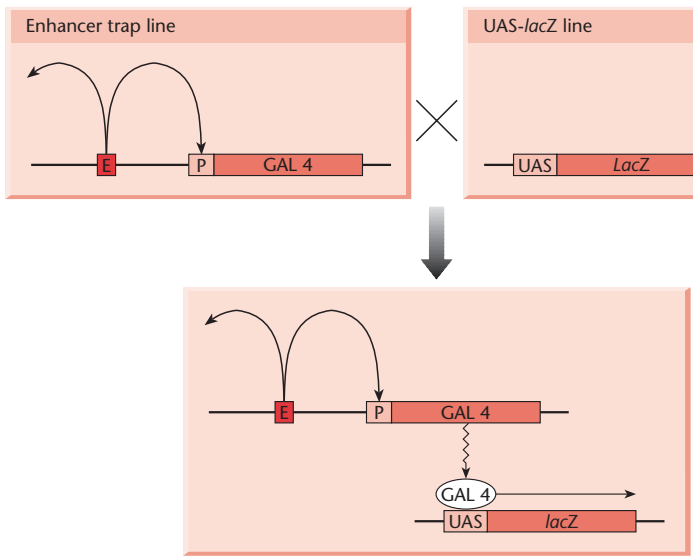


Fig. 13.14 Second-generation enhancer trap, in which the yeast gene for the GAL4 transcription factor is activated by the enhancer. This can be crossed to a responder line in which *lacZ* is driven by a GAL4-dependent promoter.

Enhancer traps in other species

Simple *lacZ* enhancer traps based on the original *Drosophila* model have also been used in mice (Allen *et al.* 1988, Kothary *et al.* 1988). The constructs were introduced into eggs by pronuclear microinjection and the resulting transgenic embryos could be screened for β-galactosidase activity. One disadvantage of this strategy is that, unlike P-element insertion, gene transfer by pronuclear microinjection tends to generate complex, multicopy, transgene arrays, often combined with major structural rearrangements of the genome. These factors complicate the interpretation of reporter expression patterns and hamper efforts to recover the corresponding genes. The use of ES cells and alternative transformation techniques have addressed these problems. ES cells can be transformed by electroporation, resulting in simpler transgenic loci with a lower transgene copy number (Gossler *et al.* 1989), or by retroviral vectors, which generate single-copy insertions (Friedrich & Soriano 1991, Von Melchner *et al.* 1992). Enhancer-trap vectors have also been developed for plants, using T-DNA insertions carrying a *gusA* gene driven by a weak promoter adjacent to the right border repeat (Goldsbrough & Bevan 1991, Topping *et al.* 1991). The activity of endogenous enhancers can be determined by screening the plants for GUS activity. The use of *lacZ* and *gusA* as histological reporters is discussed in Box 13.1.

Gene traps and functional genomics

Although enhancer traps have been widely used in animals and plants, there have been relatively few reports of successful gene-cloning efforts other than in *Drosophila*. Much more use has been made of an alternative type of entrapment construct, the *gene trap*, which is activated only upon direct integration into a functional gene. Depending on the site of integration, the insertion may or may not generate a mutant phenotype, but in either case the gene can be directly cloned by virtue of its unique insertional tag.

The first entrapment vectors in plants were gene traps that employed a promoterless selectable marker providing resistance to the antibiotic neomycin, in an attempt to directly select for T-DNA insertions into functional genes (Andre *et al.* 1986, Teeri *et al.* 1986). A problem with this approach was that selection favoured plants with multiple copies of abnormal T-DNA inserts (Koncz *et al.* 1989). A second selectable marker was therefore incorporated into the vector, allowing transformed plants to be selected for hygromycin resistance, while the trapped neomycin-resistance marker was used as a reporter gene to confirm integration into functional genes. Replacing the neomycin-resistance marker with *gusA* provided the increased versatility of direct visual screening of gene-trap lines, and vectors of this type, based on transposons or T-DNA, have been widely used (Federoff & Smith 1993, Lindsey *et al.* 1993). An alternative type of gene trap incorporates a splice acceptor site, in addition to the promoterless reporter gene, thus making reporter-gene expression dependent on the formation of a transcriptional fusion with the trapped gene (Sundaresan *et al.* 1995). Promoter-trap and splice-trap vectors are compared in Fig. 13.15.

In mice, gene-trap technology has become very sophisticated and can be used, in ES cells, as an excellent high-throughput strategy to identify and characterize new genes. The technique was developed in parallel in a number of laboratories, predominantly based on the splice-trap principle (Gossler *et al.* 1989, Friedrich & Soriano 1991, Skarnes *et al.* 1992, Wurst *et al.* 1995).

Several clever strategies have been employed to address potential pitfalls of the technique and improve the versatility of gene-trap vectors. For example, in order to select ES cells with productive gene-trap insertions, the selectable marker *neo* has been fused to the reporter gene *lacZ* to generate a hybrid marker called β-geo (Gossler *et al.* 1989). A potential disadvantage of the gene-trap strategy is that only in-frame insertions will generate a functional reporter protein. This has been addressed using an internal ribosome entry site (Box 13.3), so that translation of the reporter gene occurs independently of the transcript in which it is embedded (Skarnes *et al.* 1995, Chowdhury *et al.* 1997). The possibility that transiently expressed genes will be overlooked in large-scale screens has been addressed, using a binary system in which Cre recombinase is expressed in the gene-trap construct and activates the *lacZ* reporter

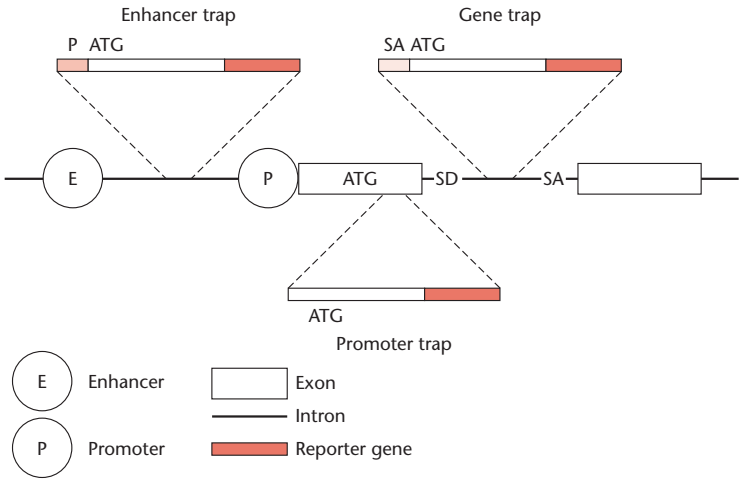


Fig. 13.15 Comparison of the structure and target-integration positions of three different types of entrapment vector – the enhancer trap, a typical gene trap and a promoter-dependent gene trap.

gene, using the RAGE strategy discussed earlier in the chapter (p. 257). Thus, β-galactosidase is expressed constitutively in all cells where Cre was transiently active (Thorey *et al.* 1998). In other vectors, the reporter gene incorporates a nuclear targeting sequence, which carries the enzyme into the nucleus and increases the concentration of the signal (Takeuchi *et al.* 1995). However, relying on endogenous targeting signals provided by the interrupted gene can allow visual screening for proteins that function in specific cell compartments (e.g. Tate *et al.* 1998).

The applicability of gene trapping to high-throughput functional genomics has been demonstrated by Wiles and colleagues, who were able to generate a library of over 12 000 gene-trapped ES cell lines, using retroviral gene-trap vectors (Wiles

Box 13.3 Internal ribosome entry sites

The translation of most eukaryotic mRNAs is dependent on the 5′ cap, with which the ribosome associates before scanning for the initiation codon. However, certain RNA viruses and a few endogenous mRNAs contain an *internal ribosome entry site* (IRES), which allows cap-independent translation (reviewed by Mountford & Smith 1995).

An IRES can be very useful in expression vectors because it allows multiple transgenes to be expressed on a single transcript. One way in which this can be exploited is to increase the efficiency of cotransformation. In traditional plasmid-based cotransformation techniques, the transgene and a selectable marker are transcribed independently; therefore many selected cells express the marker but not the transgene. However, if both genes are expressed as a single, dicistronic mRNA, then selection for the marker necessarily identifies cells in which the non-selected transgene is also expressed (Kaufman *et al.* 1991). Cotransformation with retroviral vectors is also much more efficient if a dicistronic expression system is used in preference to dual-promoter or alternative-splicing vectors, since, in the latter cases, the non-selected gene is often not expressed (e.g. Ghattas *et al.* 1991).

Linking an IRES to a reporter gene also has numerous applications. For example, most transgenes do not confer a convenient phenotype that can be used to track their expression. However, by placing an IRES-controlled reporter gene downstream of the transgene, the expression pattern of the transgene can be established very easily. In this context, IRES elements are particularly useful where transcriptional fusions are generated, because this removes any dependence on in-frame insertions to generate a functional protein. This is important in gene-trap vectors, where the position of integration cannot be controlled (Skarnes *et al.* 1995; see p. 271).

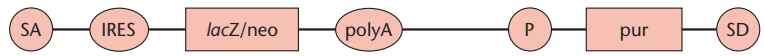


Fig. 13.16 A highly versatile gene-trap vector designed for high-throughput screening. See text for details.

et al. 2000). Such a library of clones can be grown, stored frozen in microtitre plates and systematically analysed to build up a molecular library of flanking sequences. In order to characterize the expression patterns of the trapped genes, it is possible to undertake large-scale screens of mouse embryos or particular tissues of adult mice produced from the ES cells, but prescreening is also possible by exploiting the ability of ES cells to differentiate along different pathways in specific culture media (e.g. Reddy *et al.* 1992, Baker *et al.* 1997, Yang *et al.* 1997). It is also possible to identify ES cells in which particular functional subsets of genes are trapped, e.g. secreted proteins (see below). An elegant gene-trap vector suitable for high-throughput screening has been designed by Zambrowicz *et al.* (1998) (Fig. 13.16). There are two expression cassettes in this vector: the β-geo hybrid marker downstream of a splice acceptor and internal ribosome entry site, and the selectable marker *pur*, conferring resistance to puromycin, under the control of a constitutive promoter but lacking a polyadenylation signal, upstream of a splice donor site. Using this vector, expression of the reporter is dependent on the formation of a transcriptional fusion with the trapped gene, but, since the selectable marker is driven by its own promoter but dependent on the structure of the gene for correct processing, selection for insertions is not biased in favour of genes that are actively expressed in ES cells.

Recently, a *Xenopus* gene trap has also been developed, based on the transgenesis procedure of Kroll and Amaya (1996) (p. 216). In this system, the reporter gene for green fluorescent protein is used (Box 13.1), allowing reporter expression patterns to be analysed in living tadpoles (Bronchain *et al.* 1999). This system could become a powerful tool for functional genomics in *Xenopus* because hundreds or thousands of embryos can be generated in a single day and screened for gene-trap events in real time.

Function-specific trapping

The design of gene-trap constructs can be modified to select for particular classes of genes. For example, Skarnes *et al.* (1995) describe a construct in which the β-geo marker is expressed as a fusion to the transmembrane domain of the CD4 type I protein. If this inserts into a gene encoding a secreted product, the resulting fusion protein contains a signal peptide and is inserted into the membrane of the ER in the correct orientation to maintain β-galactosidase activity. However, if the construct inserts into a different type of gene, the fusion product is inserted into the ER membrane in the opposite orientation and β-galactosidase activity is lost.