CHAPTER 15

Advanced transgenic technology

Introduction

Gene transfer experiments in animals and plants have evolved beyond the simple process of introducing additional DNA sequences into target genomes. Now that DNA transfer is becoming routine in many species, attention is shifting to more sophisticated approaches which involve the precise control of transgenes and endogenous genes, both in terms of their structure and their expression. Parallel advances in animal and plant biotechnology have come about through the development of inducible expression systems which facilitate the external regulation of transgenes, and through the exploitation of site-specific recombination systems to make precise insertions and modifications in target sequences. Transgenic technology is particularly advanced in mice, where combinations of gene targeting, site-specific recombination, and inducible transgene expression make it possible to activate and inactivate both transgenes and endogenous genes in a conditional manner.

Other routes to conditional gene silencing have also been explored. These do not involve the direct modification of the target gene but rather the expression of inhibitory genes whose products interfere with the expression of the target. Although many such strategies have been developed, e.g. antisense RNA, ribozymes, interfering antibodies, and dominant negative mutants, one approach in particular has risen to dominate the field. From rather obscure beginnings, the phenomenon of RNA interference (RNAi) has emerged as one of the most exciting recent developments in molecular biology, and may soon make the transition from laboratory to clinic.

Inducible expression systems allow transgene expression to be controlled by physical stimuli or the application of small chemical modulators

In many gene transfer experiments, it is desirable for the introduced transgene to be expressed in a specific manner. In both animals and plants, cell- or tissue-specific promoters are used to restrict transgene expression to certain areas of the organism. In a commercial setting, for example, it is useful to restrict transgene expression in animals to the mammary glands, so that recombinant proteins can be recovered from milk (Wall *et al.* 1999). Similarly, it is useful to restrict transgene expression in plants to the seeds, since this is a stable environment for protein accumulation (Stoger *et al.* 2000). In other cases, it may be necessary to control transgene expression more precisely. 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 – under these circumstances the experimenter might want to choose the best time to switch the transgene off. Such issues can be addressed through the use of *inducible expression systems*, in which transgene expression is controlled by an external stimulus.

Some naturally occurring inducible promoters can be used to control transgene expression

A number of inducible expression systems have been developed for animals and plants 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, at least in the context of inducible expression constructs, in that the stimulus that activates it is physical. Most inducible promoters used to control transgene expression 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 analogs 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 tumor 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 by including a source of heavy metals in their drinking water. An example of zinc-induced activation of a rat growth-hormone gene in transgenic mice is discussed on p. 259. 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). An ethanol-inducible system has also been described (Roslan *et al.* 2001).

Endogenous chemically inducible promoters have also been used to control transgene expression in plants. Two systems have been widely employed: the *PR-1a* promoter, which is induced by pathogen infection and by chemical elicitors such as benzothiadiazole (Gorlach *et al.* 1996), and the maize *In2–2* (Inducible gene 2–2) promoter, which is induced by benzenesulfonamide safeners (agrochemicals that are used to increase the tolerance of plants to herbicides; De Veylder *et al.* 1997). The advantage of these systems is that neither inducer is toxic to plants so they can be applied safely in the field as well as under laboratory conditions. Physically inducible systems for plants have also been described: one example is a peroxidase gene promoter isolated from sweet potato (*Ipomoea batatas*), which is inducible by hydrogen peroxide, wounding, or ultraviolet light (Kim *et al.* 2003). A wound-inducible promoter from the tomato hydroxy-3-methylglutaryl CoA reductase 2 (*HMGR2*) gene has been used as the basis of a commercial inducible promoter system called MeGA-PharM (Mechanical Gene Activation Post-harvest Manufacturing), which allows transgenes encoding pharmaceutical proteins in plants to be switched on after the leaves have been harvested and shredded (Ma *et al.* 2003, Fischer *et al.* 2004).

Unfortunately, there are several limitations associated with the use of endogenous inducible promoters which limit their usefulness. First, they tend to be somewhat leaky, i.e. there is a low to moderate level of background transcription in the absence of induction. Second, the level of stimulation achieved by induction (the *induction ratio*) is often quite low, typically less than 10-fold. Third, there are often unwanted side effects caused by the activation of other endogenous genes that respond to the same inductive stimulus. The inducers used with many of these endogenous promoters are also toxic or damaging if contact is prolonged. Finally, the kinetics of induction are generally 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. For these reasons, there has been great interest in the development of alternative inducible expression systems.

Recombinant inducible systems are built from components that are not found in the host animal or plant

Many of the disadvantages of endogenous inducible promoters can be addressed using recombinant systems, since these can be designed to work more efficiently. All recombinant inducible expression systems are based on a two-component principle in which the transgenic organism is transformed with a transgene encoding a transcription factor whose activity is controlled by the external stimulus or chemical, as well as a transgene regulated by that transcription factor. Because the components are heterologous, there should be no coincidental activation of other endogenous genes in the transgenic organism. For the same reason, the inducer itself should be non-toxic, because it does not interfere with any endogenous processes. Current systems aspire to the ideal in which the inducing 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.

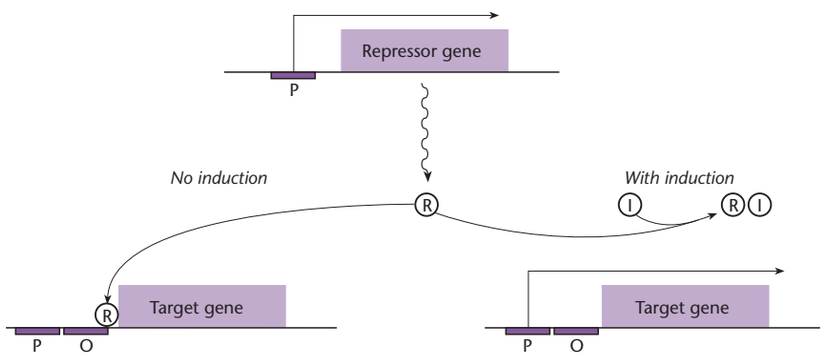
The *lac* and *tet* repressor systems are based on bacterial operons

The first heterologous systems were based on bacterial control circuits (Fig. 15.1). Hu & 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 *lac*I. 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 analog, 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 & Davison (1987) modified the *lac*I 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.

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).



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. 15.1 Summary of repression-based inducible expression-control circuits based on the *lac* and *tet* operons of *E. coli*. P = promoter.

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 were developed to circumvent some of the limitations of the original *tet* system

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. 15.2). The *tet* transactivator (tTA) system has been more 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. 15.3). 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 erythropoietin 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 significant implications for the use of inducible expression systems for gene therapy. Unfortunately, the mutations that reverse the effect of doxycycline also reduce its binding activity, so up to 10 times more antibiotic is required to activate rtTA than inhibit tTA. More recently, new versions of rtTA

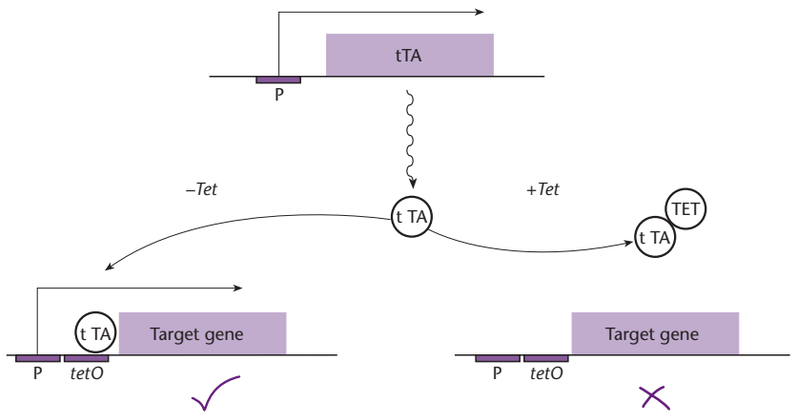


Fig. 15.2 The *tet* transactivator (tTA) system.

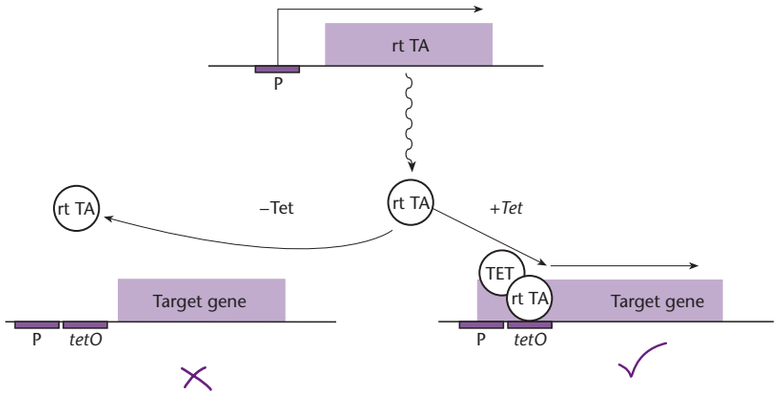


Fig. 15.3 The reverse *tet* transactivator (rtTA) system.

have been developed with mutations at a different site that does not affect doxycycline binding (Urlinger *et al.* 2000, Lamartina *et al.* 2002). Another limitation of the original rtTA system is that the regulator displays a significant residual binding to the *tet* operator in the absence of doxycycline. This has been addressed by the development of Tet-repressible transcriptional repressors or suppressors tTR and tTS, which out-compete rtTA for *tet* operator sites in the absence of doxycycline, but are unable to bind when the antibiotic is present. They do not, however, physically interact with rtTA (Zhu *et al.* 2001). A reverse-transactivator *lac* system has also been developed, with an induction ratio of approximately 10^4 (Baim *et al.* 1991). Another system, based on the pristinamycin (Pip) operon, is available in standard and reverse formats induced by antibiotics of the streptogramin family (Fussenegger *et al.* 2000). The availability of different systems should allow multiple transgenes to be controlled independently by different inducers (Corbel & Ross 2002).

Steroid hormones also make suitable heterologous inducers

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 advantageous because, in addition to their favorable kinetics, such molecules should not activate endogenous signaling 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 molting. As with other steroid-like signaling 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 signaling 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 analog, 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 homolog, 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).

The ecdysone system has also been employed successfully in plants. This is due to the identification and development of non-steroidal agonists of the ecdysone receptor which are safe for field use. For example, Martinez *et al.* (1999) developed a hybrid system consisting of the tobacco budworm ecdysone receptor ligand-binding domain fused to the glucocorticoid receptor DNA-binding domain and the VP16 transactivation domain. The receptor responds to tebufenozide (an insecticide better known by its trade name CONFIRM). Similarly, Padidam *et al.* (2003) have developed a system that is based on the spruce budworm ecdysone receptor ligand-binding domain, and responds to another

common insecticide, methoxyfenozide (INTREPID). Another system based on the European corn borer ecdysone receptor also responds to this insecticide (Unger *et al.* 2002).

The glucocorticoid receptor has also been developed as a heterologous system in plants (Schena *et al.* 1991, Aoyama & Chua 1997, Moore *et al.* 1997). The system described by Aoyama & Chua comprises the glucocorticoid-receptor steroid-binding domain 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. In a recent application, this system was used to express a viral replicase in a plant also expressing human γ-interferon. The intrinsic amplification stimulated gene expression more than 300-fold over the unamplified system (Mori *et al.* 2001). Another interesting development is a dual system which is activated by dexamethasone and repressed by tetracycline (Bohner *et al.* 1999).

Chemically induced dimerization exploits the ability of a divalent ligand to bind two proteins simultaneously

A further strategy for inducible transgene regulation has been developed, exploiting essentially the same principles as the yeast two-hybrid system (p. 459). 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).

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 assemble effectively 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. 15.4). 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. Various analogs of rapamycin (rapalogs) have therefore been developed that do not have significant pharmacological effects. Indeed, the best systems use rapalogs that no longer bind to endogenous FKBP and FRAP, but only interact with modified derivatives developed especially for the CID system. Other advantages of rapalogs include their prolonged bioavailability following oral administration and their ability to cross the blood–brain barrier. In the most popular version of the system (Fig. 15.5) three copies of FKBP are joined to a synthetic DNA-binding domain comprising a pair of zinc fingers and a homeodomain, which binds a novel response element, whereas the FKBP-rapamycin binding domain of FRAP is joined to the p65 activation domain subunit of NF-κB. Generally, both hybrid components are expressed from a single transcription unit using an internal ribosome entry site. The target promoter contains 12 copies of the target site followed by a TATA box (Auicchio *et al.* 2002 a,b, Chong *et al.* 2002).

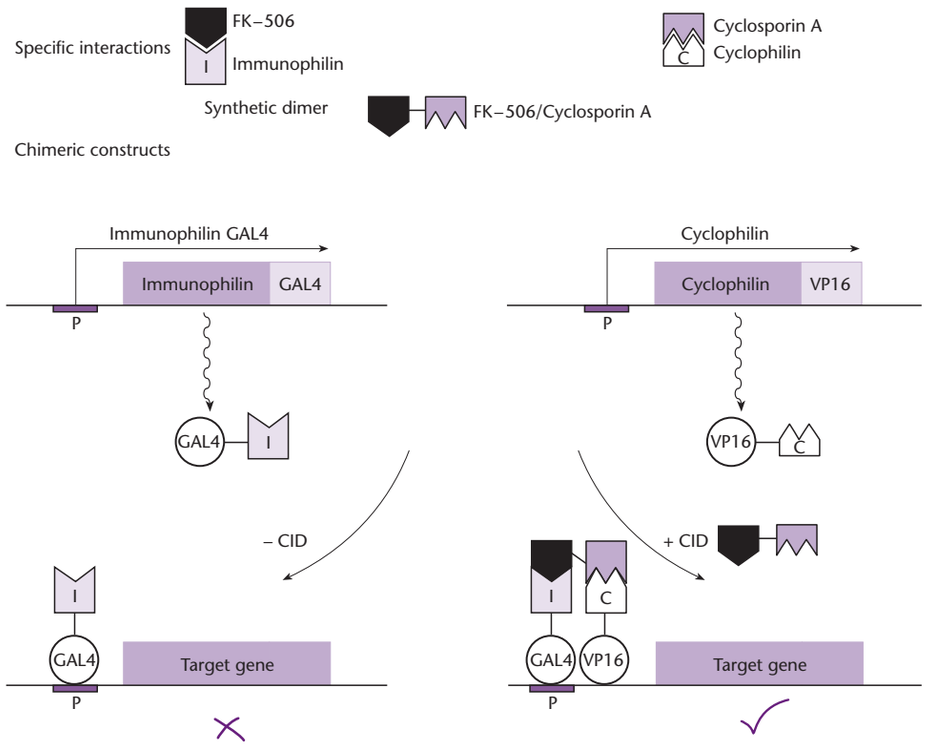


Fig. 15.4 Overview of chemically induced dimerization between the synthetic FK-506/cyclosporin A conjugate and fusion proteins containing immunophilin and cyclosporin domains. Dimerization assembles a functional transcription factor that can activate a promoter with GAL4 response elements.

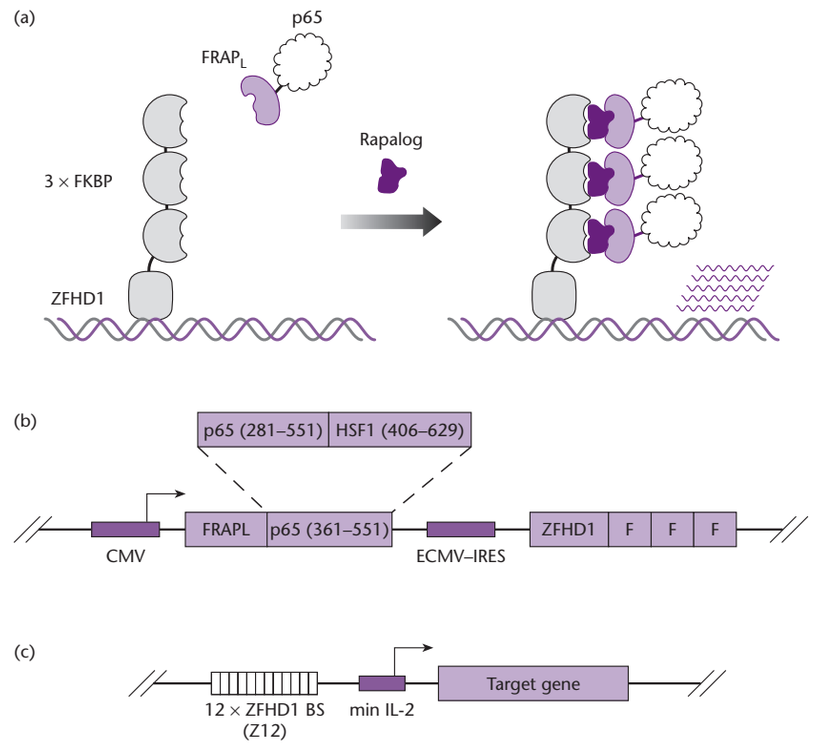


Fig. 15.5 (a) A frequently used configuration of the rapamycin system, in which three tandem copies of the human FKBP are fused to a synthetic DNA-binding domain and FRAP is fused to the DNA activation domain of p65. Rapamycin facilitates interaction between FKBP and FRAP, leading to the assembly of functional transcription factor. (b) Detailed structure of the construct encoding the three copies of FKBP and the synthetic DNA-binding domain ZFHD1 (zinc finger homeodomain 1). ECMV-IRES is the internal ribosome entry site from encephalomyocarditis virus. (c) Structure of a typical target gene showing 12 tandem copies of the ZFHD1-binding site.

Not all inducible expression systems are transcriptional switches

The inducible expression systems discussed above are all regulated at the level of transcription, such 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 estrogen receptor exists in an inert state in the absence of estrogen, because the hormone-binding domain interacts with heat-shock protein 90 (Hsp90) to form an inactive complex. When estrogen 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. 15.6). In principle, any protein expressed as a fusion with the estrogen-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 estrogen or an analog, such as Tamoxifen, which does not induce endogenous estrogen-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).

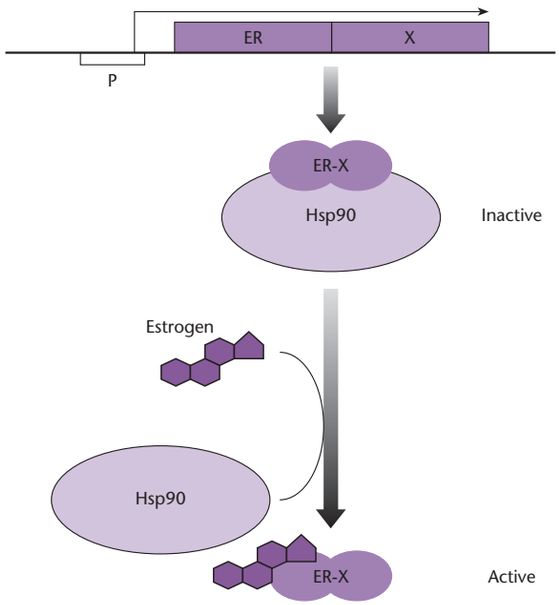


Fig. 15.6 The estrogen-inducible expression system, which works at the post-translational level – see text for details.

Site-specific recombination allows precise manipulation of the genome in organisms where gene targeting is inefficient

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 pre-selected 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 germline; 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 differs from homologous 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). However, the greatest advantages are seen if multiple systems can be used simultaneously, as has been proposed as a strategy to develop marker-free transgenic plants (Srivastava & Ow 2004).

Site-specific recombination can be used to delete unwanted transgenes

The reaction catalyzed by Cre recombinase is shown in Fig. 15.7. 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

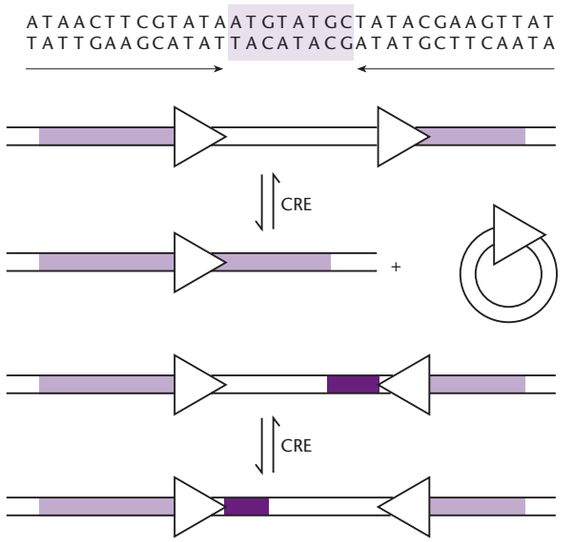


Fig. 15.7 Structure of the *loxP* site and reactions catalyzed by Cre recombinase when paired *loxP* sites, shown as arrows, are arranged in different orientations.

repeats, the intervening DNA segment is inverted. Both reactions are reversible. However, when the *loxP* sites are arranged in the same orientation, excision is favored 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 13 that traditional strategies for generating subtle mutants in mice involve two rounds of homologous recombination in embryonic stem cells (ES cells) (p. 258). 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. 15.8, 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. A second negative marker (e.g. the gene for diphtheria toxin) is included outside the homology region to select against random integration

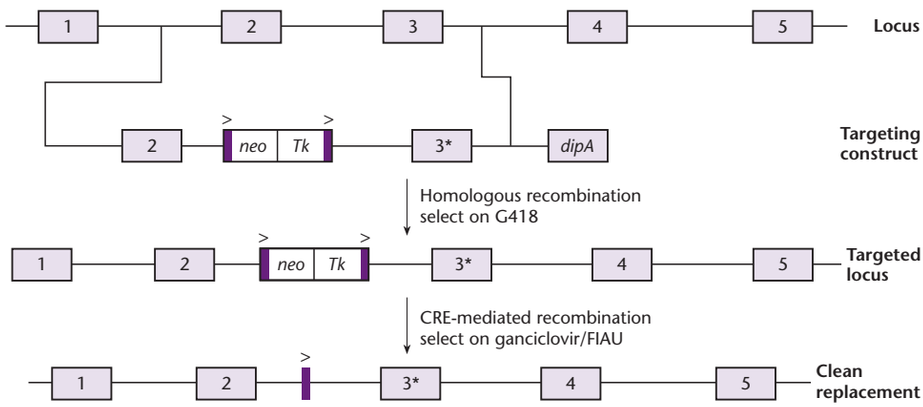


Fig. 15.8 Gene targeting followed by marker excision, catalyzed by Cre recombinase. Initially, positive and negative markers (*neo* and *Tk*), flanked by *loxP* sites, are introduced by homologous recombination (a second negative marker, in this case encoding diphtheria toxin (DIPA), is included outside 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.

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 catalyzes 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 & 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 recombination can be used to activate transgene expression or switch between alternative transgenes

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. 15.9). 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.

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

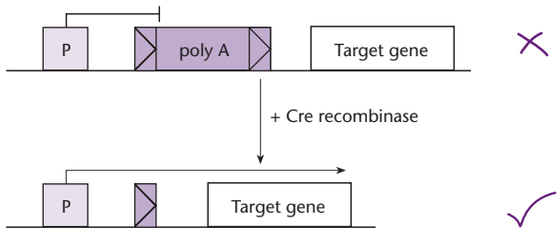


Fig. 15.9 Overview of the recombinase-activated gene expression (RAGE) strategy. A polyadenylation signal is inserted between the promoter and target gene, blocking its expression. However, if this signal is flanked by *loxP* sites, Cre recombinase can be used to excise the block, bringing the promoter and gene into juxtaposition and thus activating gene expression.

strategy, because it allows "mix and match" between different Cre transgenic and "responder" lines. We return to this subject below.

Site-specific recombination can facilitate precise 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 favor 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).

With recent pressure to develop marker-free transgenic animals and plants in the biotechnology industry, this area of research has benefited from a large amount of funding. Particular progress has been made with commercially important crops such as rice, where targeted integration is now possible at efficiencies that yield hundreds of clones per experiment (Srivastava & Ow 2002, Srivastava *et al.* 2004). 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 13.2), 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.

Site-specific recombination can facilitate 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 intra-chromosomal 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. 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).

Inducible site-specific recombination allows the production of conditional mutants and externally regulated transgene excision

In mice, gene targeting and site-specific recombination can be used in a powerful combined approach

Box 15.1 Visible marker genes

Reporter genes are widely used for *in vitro* assays of promoter activity (Box 12.1). 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 hydrolyzes β-D-galactopyranosides, such as lactose, as well as various synthetic analogs. 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 Xgal* 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 tumor 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 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 catalyzes 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 analyze 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 colors (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

continued

Box 15.1 *continued*

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.* 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, many variations of the protein are now available such as cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP), which can be used for dual labeling (e.g. Tsien & Miyawaki 1998; reviewed by Ellenberg *et al.* 1999). Fluorescent proteins of other colors are also available, many from coral reef organisms. For example, the proteins DsRed, AmCyn, and ZsYellow are all coral-derived fluorescent proteins available from Clontech. A mutant form of red fluorescent protein from *Anthoza* (Matz *et al.* 1999) 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; Xgal: 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; X-gluc: 5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid.

to generate conditional knockout mutants. Essentially, targeting vectors are designed so that part of a selected endogenous gene becomes flanked by *loxP* 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 estrogen 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).

As well as creating conditional mutants, inducible site-specific recombination can be used to control transgene excision externally. A number of reports have been published recently in which Cre or FLP recombinase has been expressed under inducible control in plants using one of the systems discussed at the beginning of the chapter. This allows the recombinase gene to be maintained in an inactive state until the stage of development at which the marker needs to be removed. Furthermore, this strategy is suitable for plants that are propagated vegetatively. The typical strategy, which is to cross plants containing a floxed transgene with plants containing a conditionally expressed *cre* gene, is not possible in plants that do not reproduce by sexual crossing. However, by placing the *cre* gene under inducible control, both the floxed transgene and the *cre* gene can be introduced at the same time (Lyznik *et al.* 1995, Sugita *et al.* 2000, Zuo *et al.* 2001, Hoff *et al.* 2001, Zhang *et al.* 2003).

Many strategies for gene inactivation do not require the direct modification of the target gene

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 routinely 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*.

Antisense RNA blocks the activity of mRNA in a stoichiometric manner

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 over-ripening. 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).

Ribozymes are catalytic molecules that destroy targeted mRNAs

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 centers 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 their catalytic activity: 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 is the inhibition of an endogenous gene by the presence of a homologous sense transgene

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 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 15.2). One of the most remarkable 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 onto 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 & 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 15.2).

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.

Box 15.2 Gene silencing

As discussed in the main text, there are several forms of gene silencing in eukaryotes, which act in a sequence-specific manner to inhibit the activity of particular genes or transgenes. These forms of silencing can occur at either the transcriptional or post-transcriptional levels. In the former case, no mRNA is produced from the affected gene, while in the latter case transcription is not only permitted, but is actually necessary for silencing to occur. Transcriptional silencing reflects the structure of chromatin, which can form an open configuration (euchromatin) that is permissive for gene expression or a closed configuration (heterochromatin) that represses gene expression. Post-transcriptional silencing (now more commonly referred to as RNA silencing) reflects the activity of particular protein complexes that target mRNAs with a specific sequence for destruction. Both forms of silencing have arisen as mechanisms of defense against invasive nucleic acids (viruses, transposons etc.), and both have been subjugated as mechanisms for gene regulation. There is extensive cross-talk between these silencing pathways.

Position-dependent silencing and context-dependent silencing

These forms of transcriptional 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 heterochromatin. 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 12.1). 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).

RNA silencing

All forms of RNA silencing (i.e. PTGS and VIGS in plants, cosuppression, RNAi, quelling in fungi) appear to depend in some way or other on the presence of double-stranded RNA. The biogenesis of dsRNA can occur in many ways,

continued

Box 15.2 *continued*

e.g. through the deliberate introduction or expression of complementary RNA molecules in the cell in RNAi, the production of aberrant dsRNA from complex transgenes in PTGS, or the production of viral replication intermediates in VIGS. Once the dsRNA has formed, it becomes the substrate of a nuclease called Dicer, which reduces it to short duplexes, 21–25 bp in length with overhangs. These duplexes are known as small interfering RNAs (siRNAs). They assemble with several proteins, including one of the Argonaute family, into an endonucleolytic complex known as the RNA induced silencing complex (RISC), which uses one strand of the siRNA to target complementary mRNAs and cleave them. The RISC is efficient, resulting in potent silencing (Tijsterman & Plasterk 2004). In many organisms (but not mammals) there is amplification of the siRNA by RNA-dependent RNA polymerase to increase the potency of the silencing effect. The siRNA molecules can also move between cells, explaining why RNAi and other RNA silencing phenomena are systemic.

RNA silencing is a form of genomic defense

It is likely that RNA silencing 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).

RNA silencing and gene regulation

Small RNA molecules are also involved in gene regulation, and the structure of these endogenous regulators is very similar to that of the siRNAs produced during RNA interference experiments. These so-called microRNAs (miRNAs) are produced not through the cleavage of longer dsRNA precursors, but are the direct transcripts of small miRNA genes that generate hairpin RNA precursors. The processed miRNAs assemble into a complex called the miRNP (micro-ribonucleoprotein) complex, which binds to the 3′ untranslated regions of target mRNAs and regulates their translation. Once considered a peculiarity, the discovery of hundreds of miRNA genes in eukaryotic genomes (Lim 2003) now suggests miRNA may be a major form of gene regulation. Researchers are focusing on identifying the targets of miRNAs and the processes they help to regulate. While siRNAs and miRNAs were once distinguished by their mechanism (mRNA cleavage or translational repression), the extensive overlap between the two pathways now means they are classified by origin (one from dsRNA produced either inside or outside the cell, one from transcription of an endogenous gene). The chosen pathway appears to depend on the type of complex involved – siRNAs are processed by Dicer and assemble into RISCs which cause cleavage, while at least in animals miRNAs are also processed by another enzyme, Drosha, and assemble into miRNPs, which have a regulatory role. The choice between these complexes appears to depend on how perfect the match is between the complementary strands. Perfect duplexes generally behave as siRNAs whereas duplexes with mismatches or bubbles behave as miRNAs (Bartel 2004, Kim 2005).

Cross-talk between RNA silencing and transcriptional silencing

The RNA silencing apparatus also has significant interactions with chromatin-

continued

Box 15.2 *continued*

modifying proteins, including histone-modifying enzymes and DNA methyltransferases. Recent evidence has accumulated that the RNAi machinery can set epigenetic marks in plants and yeast (Wassenegger 2005) as well as in *Drosophila* and mammalian cells (Matzke & Birchler 2005). Although the mechanisms involved are complex and yet to be elucidated fully, it appears that repetitive regions of the genome such as tandem repeats of transposons (and perhaps integrated transgenes) can generate long dsRNA molecules which are processed into siRNA-like structures known as rasiRNAs (repeat-associated siRNAs). These are cleaved by Dicer and form complexes known as RITS (RNA-induced initiation of transcriptional silencing), which mediate histone modification and DNA methylation (Xie *et al.* 2004).

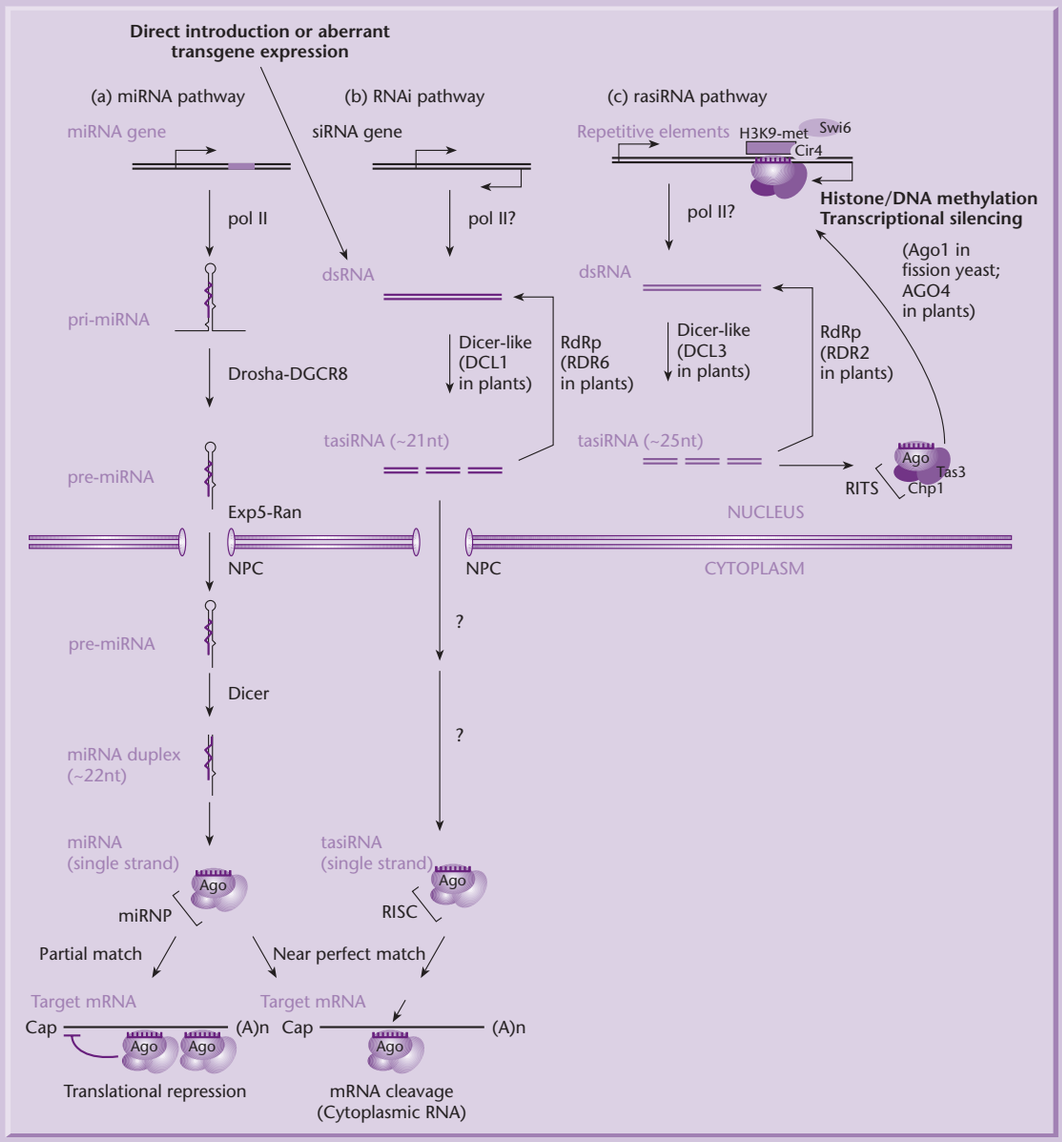


Fig. B15.1 Current model of small RNA pathways which lead to transcriptional and post-transcriptional silencing.

RNA interference is a potent form of silencing caused by the direct introduction of double-stranded RNA into the cell

RNA interference (RNAi) is a sequence-specific gene silencing phenomenon caused by the presence of double-stranded RNA. The process was discovered by researchers working on gene silencing in the nematode worm *C. elegans*, when they observed that either sense *or* antisense RNA could suppress the expression of a homologous gene. It was Fire *et al.* (1998) who first showed that the deliberate introduction of both sense and antisense RNA into worms *at the same time* caused a striking and specific inhibitory effect, which was approximately 10-fold more efficient than either single RNA strand alone. They correctly postulated that the silencing effects seen in the earlier experiments were in fact due to the presence of contaminating dsRNA. Although discovered in *C. elegans*, the principle of RNAi can be traced back even further, since both cosuppression and virus-induced gene silencing in plants are thought to involve the production of dsRNA (due to aberrant transgene expression and viral replication, respectively).

Initial investigation showed that only a few molecules of dsRNA were necessary to induce RNAi in *C. elegans*, suggesting that like ribozymes, RNAi is catalytic rather than stoichiometric. Interference can be achieved only if the dsRNA is homologous to the exons of a target gene, indicating that it is a post-transcriptional process. The phenomenon of RNAi appears to be quite general, and has been used for gene-silencing experiments in many other organisms, including *Drosophila* (Kennerdell & Carthew 2000), mice (Wianny & Zernicka-Goetz 2000), and plants (Waterhouse *et al.* 1998, Chuang & Meyerowitz 2000). Indeed, RNAi is fast becoming the method of choice for large-scale functional analysis in these organisms due to its simplicity, specificity, and potency (see for example Hammond *et al.* 2001, Hannon 2002). We return to the topic of large-scale RNAi screens in Chapter 19.

In *C. elegans* microinjection is the most consistently effective way to induce RNAi. Typically, *in vitro* synthesized dsRNA is injected into the germline of adult worms, and progeny are screened for RNAi-induced phenocopies. However, because RNAi is a systemic phenomenon, microinjection is not the only way it can be achieved. Technically simpler ways to induce RNAi include adding dsRNA to the worms' liquid medium, or even feeding the worms on bacteria that have been engineered to express dsRNA (Maeda *et al.* 2001, Fraser *et al.* 2000). More recently, transgenic strategies to achieve RNAi have become popular. 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.

Until 2001, RNAi was not possible in mammals due to an unrelated phenomenon called the interferon response, which shuts down protein synthesis in the presence of dsRNA molecules greater than 30 bp in length and masks any specific effects of gene silencing. However, as more was learned about the mechanism of RNAi, a way around this problem was envisaged. As shown in Box 15.2, RNAi is mediated by an enzyme called Dicer, which chops the dsRNA into short fragments, 21 or 22 bp in length with 2-nt overhangs; these are known as short interfering RNAs (siRNAs). Two groups independently showed that chemically synthesized siRNAs transfected into mammalian cells were capable of inducing specific RNAi effects without inducing the interferon response (Elbashir *et al.* 2001, Caplen *et al.* 2001), although as more experiments have been performed, some examples of the interferon response being induced by RNAi have been reported (Bridge *et al.* 2003, Sledz *et al.* 2003). Another problem with RNAi in mammalian cells is that part of the RNAi pathway appears to be missing. In other organisms, but not in mammals, there is some form of intrinsic amplification of the triggering RNA which prolongs the effect and makes it more potent. In *C. elegans*, for example, the effect of dsRNA injected into adult worms can persist in the offspring of the injected worm! Because of the absence of amplification in mammalian cells, there has been much interest in the development of transgenic systems for the expression of siRNA. Conventional strategies for transgene expression cannot be used because the siRNA genes are so short. Instead, expression cassettes have been designed which are transcribed by RNA polymerase III, since this enzyme transcribes the naturally occurring short RNA genes in mammalian genomes. In one approach, a plasmid is constructed which contains two pol III transcription units in tandem, each producing one of the siRNA strands. The separate strands are thought to assemble spontaneously into the siRNA duplex *in vivo*. A second approach is very similar, but the pol III transcription units are supplied on separate vectors. The individual RNA strands assemble in the same manner. In a

third strategy, the plasmid produces a hairpin RNA which assembles into siRNA by self-pairing. Either RNA polymerase II or III can be used to produce hairpin siRNAs because the transgene is longer. Because of these developments, many RNAi experiments have now been performed in mammalian cells (reviewed by Tuschl & Borkhardt 2002, Mittal 2004) and the first reports are beginning to appear concerning the germline transmission of siRNA transgenes in mice and rats, in some cases mirroring the equivalent mutant phenotype (Carmell *et al.* 2003, Hasuwa *et al.* 2002, Kunath *et al.* 2003). Further experiments have shown how siRNA transfected *in vivo* into mouse organs also results in a knockdown phenotype (McCaffrey *et al.* 2002). The medical applications of RNAi are potentially very exciting, and are discussed in Chapter 26.

RNAi is also being applied in plant biotechnology, and recent reports discuss several experiments in which crop plants have been improved using RNAi transgenes. Examples include using RNAi to overcome genetic redundancy in polyploids (Lawrence & Pikaard 2003), modifying plant height in rice by interfering with gibberellin metabolism (Sakamoto *et al.* 2003), changing the glutelin content of rice grains (Kusaba *et al.* 2003) and the oil content of cotton seeds (Liu *et al.* 2002), and controlling the development of leaves (Palatnik *et al.* 2003). The most interesting development is the production of coffee plants that make decaffeinated coffee by using RNAi to suppress caffeine biosynthesis (Ogita *et al.* 2003).

Gene inhibition is also possible at the protein level

Intracellular antibodies and aptamers bind to expressed proteins and inhibit their assembly or activity

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 (Chapter 6), the isolation of recombinant proteins by immunoaffinity chromatography (Chapter 23), and the development of antibody-based protein chips (Chapter 20). Similarly, oligonucleotides that bind to proteins – known as aptamers – can be used as specific capture agents. The microinjection of antibodies into cells has been widely used to block the activity of 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.

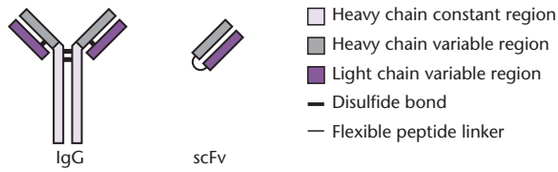


Fig. 15.10 Comparison of a normal immunoglobulin molecule with a single-chain Fv fragment.

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. 15.10). 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.

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-IL2Ra 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 & Fiedler 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 signaling by this hormone in tobacco (Artsaenko *et al.* 1995). There also some examples of expressed aptamers (known as *intramers*) being used to suppress protein activity (e.g. Good *et al.* 1997, Konopka *et al.* 2000, Thomas *et al.* 1997, Shi *et al.* 1999a). As is the case with siRNAs, the expression of intramers usually requires a pol III expression cassette (Famulok & Verma 2002).

Active proteins can be inhibited by dominant-negative mutants in multimeric assemblies

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).

Suggested reading

Albanese C., Hulit J., Sakamaki T. *et al.* (2002) Recent advances in inducible expression in transgenic mice. *Seminal Cell Development & Biology* **13**, 129–41.

Dykxhoorn D.M. & Lieberman J. (2005) The silent revolution: RNA interference as basic biology, research tool, and therapeutic. *Annual Review Medicine* **56**, 401–23.

Hannon G.J. (2002) RNA interference. *Nature* **418**, 244–51.

Gossen M. & Bujard H. (2002) Studying gene function in eukaryotes by conditional gene inactivation. *Annual Review of Genetics* **36**, 153–73.

Gunsalus K.C. & Piano F. (2005) RNAi as a tool to study cell biology: building the genome–phenome bridge. *Current Opinion in Cell Biology* **17**, 3–8.

Metzger D. & Feil R. (1999) Engineering the mouse genome by site-specific recombination. *Current Opinion in Biotechnology* **10**, 470–6.

Muller U. (1999) Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis. *Mechanical Development* **82**, 3–21.

Padidam M. (2003) Chemically regulated gene expression in plants. *Current Opinion in Plant Biology* **6**, 169–77.

Ristevski S. (2005) Making better transgenic models – Conditional, temporal, and spatial approaches. *Molecular Biotechnology* **29**, 153–63.

Saez E., No D. & West A. (1997) Inducible gene expression in mammalian cells and transgenic mice. *Current Opinion in Biotechnology* **8**, 608–16.

Srivastava V. & Ow D.W. (2004) Marker-free site-specific gene integration in plans. *Trends in Biotechnology* **22**, 627–9.

Tijsterman M. & Plasterk R.H.A. (2004) Dicers at RISC: the mechanism of RNAi. *Cell* **117**, 1–3.

Tuschl T. & Borkhardt A. (2002) Small interfering RNAs: a revolutionary tool for the analysis of gene function and gene therapy. *Molecular Interventions* **2**, 158–67.

Voinnet O. (2002) RNA silencing: small RNAs as ubiquitous regulators of gene expression. *Current Opinion in Plant Biology* **5**, 444–51.

Wang R.H., Zhou X.F. & Wang X.Z. (2003) Chemically regulated expression systems and their applications in transgenic plants. *Transgenic Research* **12**, 529–40.