4

Vectors for plant transformation

INTRODUCTION

Vectors for plant transformation have to be created in the same way as any other plasmid-derived cloning vector, by using molecular biology techniques. It is therefore useful to look at some of the basic features that are necessary or desirable in any vector. Such features are described here because the usefulness of vectors often equates with the ease with which they can be manipulated, both in standard laboratory hosts such as *Escherichia coli* and *in vitro*, before being used to transform plants. Later in this chapter we will see that many of the recent advances in the design of plant transformation vectors are aimed at improving the ease with which they can be manipulated in the laboratory, rather than their ease of use for plant transformation per se. In addition, this chapter will describe the basic features of vectors used for plant transformation, their development into systems allowing the routine transformation of a range of plant species, and some of the problems associated with vector design and transgene expression.

Desirable features of any plasmid vector

● Plasmids need certain features to function efficiently as vectors.

Ideally, as well as the ability to replicate independently from the chromosome (which depends on the presence of an origin of replication), cloning vectors should have the following properties.

1 Be of a small size (low molecular mass). Small size has several important advantages. Small plasmids are easier to manipulate *in vitro* as they are less liable to damage by shearing. Small plasmids are also usually present in higher copy number and therefore plasmid yields are higher. There is also less chance of the vector having other sites for restriction endonucleases, making the design and integration of a multiple cloning site (MCS) simpler.

2 Confer a selectable phenotype on the host cells so that transformed cells can be selected for. Most plasmid cloning vectors therefore carry genes that confer resistance to an antibiotic, very often ampicillin (a penicillin derivative).

3 Contain single sites for a large number of restriction enzymes to enable the efficient production of recombinant vectors.

4 Enable the identification of bacterial colonies containing recombinant plasmids. This is usually achieved by clustering the unique restriction sites in one small area, the MCS, which is situated in the *lacZ*α sequence that encodes the N-terminal part of *E. coli* β-galactosidase. In suitable *E. coli* hosts, complementation results in the production of active β-galactosidase, which is detected by the bacterial colony turning blue due to hydrolysis of a chromogenic substrate. Cloning DNA into the MCS effectively results in the insertional mutagenesis of *lacZ*α. After transformation with these recombinant plasmids, *E. coli* will produce white colonies as no complementation can occur. Although not 100% reliable (both false-positives and false-negatives can be generated), such a system drastically reduces the number of colonies that need to be screened in order to identify those harbouring recombinant plasmids.

These features are illustrated in Figure 4.1, which shows a typical cloning vector widely used in molecular biology.

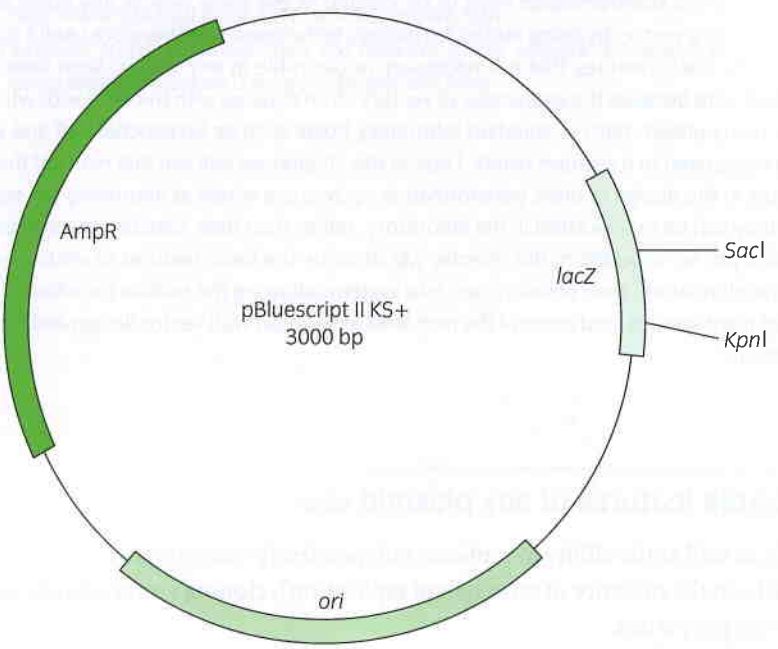


Figure 4.1 Diagram of a typical 'standard' cloning vector commonly used in molecular biology, pBluescript II KS+ in this example, although many fundamentally similar products exist. The following features are worth noting.

1 It is of a small size (3 kb) to facilitate ease of handling and high transformation efficiencies.
2 It has an origin of replication (a pUC *ori*) to enable replication of the plasmid to high copy number in *E. coli*.
3 It has an antibiotic-resistance marker to facilitate the recovery of transformed bacteria.
4 It has a multiple cloning site (MCS) in the *lacZ*α fragment to ease cloning and facilitate the identification of recombinant plasmid-containing colonies by blue/white screening. The MCS is between the *Sac*I and *Kpn*I sites (in this particular plasmid there are restriction sites for over 20 other restriction enzymes in the MCS). (Redrawn with permission from the pBluescript II KS+ figure on the Stratagene website, www.stratagene.com.)

Development of plant transformation vectors

As the process of *Agrobacterium*-mediated plant transformation was elucidated (see Chapter 3), important information came to light that made the development of efficient plant transformation vectors possible. Chief among these were the following discoveries: that the only features of the T-DNA necessary for integration into the host plant genome were the short border sequences; that the removal of the oncogene sequences enabled plants to be regenerated from transformed plant tissue by manipulating the plant hormone composition of the medium; and that the *vir* genes function in *trans*.

Basic features of vectors for plant transformation

Some of the basic factors that contribute to the design of a successful vector have already been considered. However, when considering the design of plant transformation vectors, several additional degrees of complication become apparent, as follows.

- Like all cloning vectors, plant transformation vectors need a minimum set of features to function efficiently.

1. The plasmid (if it is to be used in *Agrobacterium*-mediated transformation) needs to be able to replicate not only in *E. coli* (so that routine manipulations can be carried out) but also in *Agrobacterium*.
2. Additional selectable markers need to be included so that the successfully transformed plants can be identified.
3. Border sequences need to be incorporated into the design of plasmid vectors for *Agrobacterium*-mediated transformation to ensure integration of the genes of interest into the host plant's genome.
4. The genes (particularly if they are from prokaryotes or non-plant eukaryotes) that are to be integrated into the genome of the host plant may need to be made 'plant-like'. This includes the use of appropriate promoters and terminators to ensure that expression of the genes occurs. These features are often incorporated into the basic vector.

Promoters and terminators

- A variety of promoters can be used to direct transgene expression.

An obvious requirement for any genes that are to be expressed as transgenes in plants is that they are expressed correctly (or at least in a predictable fashion). It is known that the major determinant of gene expression (level, location, and timing) is the region upstream of the coding region. This region, termed the promoter, is therefore of vital importance (see Chapter 1). Any genes that are to be expressed in the transformed plant have to possess a promoter that will function in plants. This is an important consideration as many of the genes that are to be expressed in plants, particularly reporter genes and selectable marker genes, are bacterial in origin. They therefore have to be supplied with a promoter that will drive their expression in plants.

Transgenes also need to have suitable terminator sequences at their 3′ terminus to ensure that transcription ceases at the correct position. Failure to stop transcription can lead to the production of aberrant transcripts and can result in a range of deleterious effects, including inactivation of gene products and increased gene silencing.

In addition to the basic need for the promoter to be capable of driving expression of the gene in plants, there are other considerations that need to be taken into account, such as promoter strength, tissue specificity, and developmental regulation.

Agrobacterium-derived promoter and terminator sequences

The genes from the Ti plasmid of *Agrobacterium* that code for opine synthesis, and in particular the nopaline synthase (*nos*) gene, are widely used as a source of both promoters and terminators in plant transformation vectors. Although derived from bacterial genes, their presence on the T-DNA means they are adapted to function in plants. The *nos* promoter is usually considered to be constitutive.

The 35S promoter

The most widely used promoter used to drive expression of genes in plant transformation vectors is the promoter of the cauliflower mosaic virus 35S RNA gene (35S promoter). This promoter is considered to be expressed in all tissues of transgenic plants (although not necessarily in all cell types). In dicotyledons it drives expression at high levels, although in monocotyledons the level of expression is not as high. This makes the 35S promoter ideal for driving the expression of selectable marker genes, and in some cases of reporter genes, as expression is more or less guaranteed. The activity of the 35S promoter can be further increased by the inclusion of one or more copies of the enhancer region (Figure 4.2 shows the structure of the 35S promoter). High-level constitutive expression is not always a desirable feature though, particularly if the gene product is toxic in large amounts.

In monocotyledons, alternatives, such as the maize ubiquitin I promoter or the rice actin promoter/first intron sequence, are often used to drive the high-level expression of transgenes.

Tissue-specific promoters

Considerable effort has been put into isolating promoters that can be used to drive expression in a tissue-specific manner, although the expression is, in reality, tissue-enhanced, as expression is not completely restricted to a given tissue. Such promoter systems are potentially useful in a number of situations. The expression of any potentially harmful substances can be limited to tissues that are not consumed by animals or humans, and genes involved in specific processes can be limited to tissues in which that process occurs. It is beyond the scope of this chapter to look at all the different tissue-specific promoters that have been isolated. However, it should be noted that several have been used to successfully drive transgene expression in the predicted pattern (some examples will be seen in later chapters). It should also be noted, however, that there are examples where promoters have been found not to function, or not to drive expression in the predicted pattern, in heterologous systems. Considerable care has therefore to be taken with the use of promoters. Unless the promoter is well

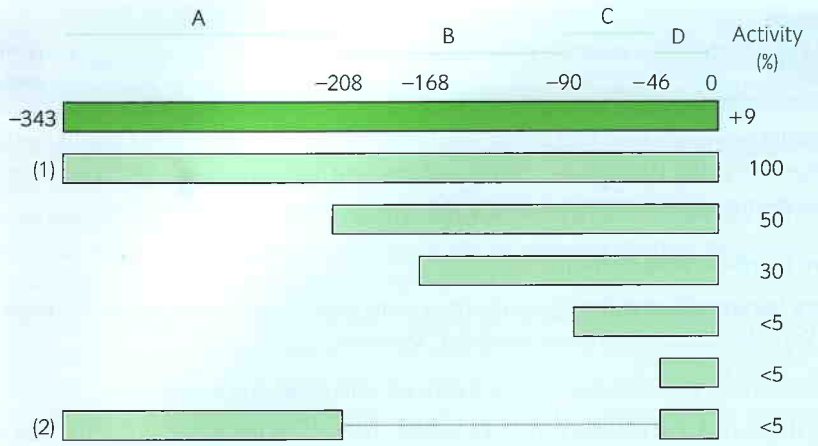


Figure 4.2 Structure/activity relationships in the 35S promoter as elucidated by deletion analysis. The uppermost box represents the 35S promoter region as defined in Fang *et al.* (1989), including 343 bp of upstream 'promoter'. The letters above designate four functional regions that can be identified in the 35S promoter. The numbers above the uppermost box give the position of the end of the promoter fragments (5′ deletions) used in the deletion analysis. (1) The boxes below represent the promoter fragments used to assess the chloramphenicol acetyltransferase (CAT) activity of various promoter fragments. (2) The broken line represents a piece of the promoter removed from a fragment, the two flanking pieces (−343 to −208 and −46 to 0) being ligated together to form a single promoter fragment. Functional regions: (A) (−343 to −208) This has no activity in itself (activity depends on the presence of the −90 to −46 region). It is responsible for about 50% of the promoter activity. (B) (−208 to −90) This functions as an enhancer and is responsible for most of the rest of the activity of the promoter. (C) (−90 to −46) This region has an accessory role in increasing the transcriptional activity of regions A and B. (D) (−46 to 0) The core promoter region, containing the TATA box. The −208 to −46 region of the promoter is often used in transformation vectors as an enhancer to increase the expression of the 35S promoter. Increasing the number of copies of the enhancer increases the transcriptional activity linearly, until a maximum stimulation of activity is reached with four copies. The enhancer also functions with heterologous promoters to increase transcriptional activity. The −343 to −90 region is responsible primarily for promoter activity in leaves, while the −90 to −46 region is required for expression in roots. (Data taken with permission from Fang *et al.* (1989).)

characterized and widely used in the plant species you intend to transform, characterization of the expression pattern conferred by the promoter (promoter::reporter gene fusions could be used) should be carried out. This is particularly so as it becomes apparent that many *cis*-acting regulatory sequences are, in fact, found in regions other than the 'promoter'.

Inducible promoters

Considerable effort has also been put into isolating and developing inducible promoter systems for use in plants. Inducible expression systems can be divided into three main categories: (1) non-plant-derived systems; (2) plant-derived systems that respond to environmental signals; and (3) plant-derived systems based on developmental control of gene expression (examples are given in Box 4.1). Theoretically, such systems allow the timing of transgene expression to be controlled carefully.

● Promoters that confer inducible expression can be used to direct expression of transgenes.

BOX 4.1 *Systems for inducible gene expression in plants*

Inducible expression gives finer control over transgene expression than does a constitutive promoter such as the 35S promoter. Several systems have been developed with the aim of giving inducible transgene expression in plants. Some examples will be considered here.

Non-plant-derived systems

Some features are considered to be desirable in any inducible expression system that depends on the application of an exogenous chemical. These are:

1 there should be no expression of the transgene in the absence of the inducer;
2 the system should respond to only one inducer, or one class of inducer, so that the system is specific to the inducer;
3 induction of gene expression should be rapid following the application of the inducer;
4 gene expression should cease rapidly following withdrawal of the inducer;
5 the inducer should be non-toxic;
6 the inducer should not cause non-specific changes in gene expression.

Some examples of inducible expression systems are given below.

Tetracycline

The antibiotic tetracycline can be used to either de-repress or inactivate gene expression. In bacteria, the tetracycline repressor (TetR) binds to the *tet* operator and negatively regulates its expression. Upon binding to tetracycline, the TetR is modified and released from the operator. This system has been adapted to function in plants. The TetR is constitutively overexpressed from a 35S promoter. The transgene is under the control of a chimaeric promoter consisting of a core 35S promoter and several copies of the *tet* operator (both upstream and downstream of the TATA box). The TetR normally represses expression of the transgene, but with the application of tetracycline it is released from the operators in the promoter of the transgene. This then results in transgene expression. This system was one of the first to be developed, but it does have some drawbacks. It does not appear to work in all plant species (such as *Arabidopsis*) as the plant cells must be able to tolerate the required levels of TetR, and tetracycline has to be applied constantly to ensure transgene expression because it has a short half-life.

Tetracycline can also be used to inactivate transgene expression. A similar promoter is used to control the expression of the transgene, but the TetR is modified to convert it to an activator (tetracycline transactivator or tTA). The tTA binds to the operators and induces gene expression in the absence of tetracycline. When tetracycline is added to the system, the tTA is released and transgene expression ceases.

Systems based on promoter repression are proposed to be less efficient at controlling gene expression than systems based on promoter activation.

Alcohol (ethanol)-inducible

In this system a chimaeric promoter is again used to drive transgene expression. In addition to a core 35S promoter region, the promoter contains binding sites for the AlcR transcription factor

BOX 4.1 *Continued*

(from *Aspergillus nidulans*), which is expressed constitutively. Upon application of ethanol to the system, AlcR binds to the AlcR-binding sites and activates transcription. The ethanol (it is actually ethanal that causes binding of the AlcR to its cognate binding sites) can be applied as a drench or vapour and there is no endogenous activation of the transgene in plants in the absence of AlcR. Activation of the promoter in tissue culture due to the induction of AlcR by endogenous metabolites (proposed to be produced due to anoxia and tissue culture conditions) precludes widespread use of this system in tissue and cell culture. There are also some doubts as to its general applicability, with some researchers showing that ethanol, at levels that would be used for inducing gene expression, can have profound effects in non-transformed plants. The system has therefore to be used with great care after carefully controlled experimentation to determine specific and non-specific ethanol effects.

The original alcohol-inducible expression system utilized a modified 35S promoter, thereby conferring expression wherever the inducing ethanol reached. Although not a problem in some applications, this lack of specificity in spatial expression pattern limited the usefulness of the system. It was known from experiments using physical confinement of the ethanol (using plastic bags for instance) that transgene expression could be limited only to areas exposed to ethanol. This was not a particularly elegant solution to the problem of achieving some spatial specificity to transgene expression and recently modifications to the system have produced ethanol-inducible promoters that do confer some spatial specificity in expression. The simplest approach has been to replace the constitutive 35S promoter elements with tissue-specific promoters (which were discussed previously in this chapter).

Steroid-inducible

A variety of systems conferring steroid inducibility on transgene expression exist. Most consist of a modified transcriptional activator capable of binding a steroid hormone or its analogue. Binding sites for the modified transcription factor are included in a chimaeric promoter, which controls transgene expression. In the absence of an inducer the transcription factor is inactive, but on application of the inducer it binds to the promoter and activates expression. Systems that respond to different steroidal inducers have been developed. Among these are ones that respond to glucocorticoids (the synthetic analogue dexamethasone is used), oestrogen, and ecdysone (an insect hormone). Although the dexamethasone system is the one most widely adopted for inducible expression in *Arabidopsis*, it appears to have some serious drawbacks that may limit its usefulness. The development of the ecdysone-based system is, however, particularly interesting as it also responds to the non-steroidal ecdysone agonists such as RH5992 (tebufenozide) and methoxyfenozide. Tebufenozide and methoxyfenozide are widely used as agrochemicals, which opens up the possibility of using this system in greenhouse and field situations.

Copper-inducible

This system is based on the yeast metallothionein regulatory system. A transcription factor is expressed constitutively. Upon binding copper, the transcription factor binds to elements in a chimaeric promoter and activates expression.

(Continued overleaf)

BOX 4.1 *Continued*

Plant-derived systems based on response to environmental signals

Although not inducible in the same sense as the non-plant-derived systems, some systems have been developed that respond to a variety of environmental signals. These systems depend on either the construction of chimaeric promoters or on the use of complete promoters from wound-inducible genes to control the expression of transgenes. The chimaeric promoters usually consist of a core promoter element (usually from the 35S promoter) and sequence elements that bind transactivating factors that respond to particular environmental signals. Several systems are described briefly here.

Wound-inducible

Many wound-inducible genes have been identified, and the promoters of some of these, particularly some proteinase inhibitors, characterized in some detail. The promoters of these genes have been found to confer wound inducibility on transgene expression. The wound inducibility may be mediated by other factors such as methyl jasmonate (see also Chapter 1), which can be used to mimic the wound response in some cases. Wound-inducible promoters can be used to drive the expression of pest-resistance genes after insect damage.

Pathogen-induced

The promoters from a range of genes induced by pathogen attack have been cloned and characterized. Many pathogen-induced promoters also show wound inducibility, and may not therefore confer the optimum expression pattern on transgenes. However, several pathogen-induced promoters have been identified that at least have a greater degree of specificity in response, being insensitive to wounding, but responding to signals such as salicylic acid.

Heat-shock-inducible

Sequence elements, termed heat-shock elements or HSEs, mediate the heat-shock inducibility of genes (see Chapter 9). If included in chimaeric promoters, they confer heat-shock inducibility on transgene expression.

Plant-derived systems based on developmental control of gene expression

Again, although not inducible in the sense of the non-plant-derived systems, many genes have been identified that are expressed at particular stages in plant development. The promoters from some of these genes have been studied in detail. As in the previous cases, either chimaeric promoters or complete promoters are used to drive transgene expression.

Senescence-specific gene expression

The promoters from two *Arabidopsis* genes, termed *SAG12* and *SAG13* (for senescence-activated gene), which are induced during senescence, have been shown to confer senescence-inducible expression on transgenes. For example, expressing the *ipt* gene from *Agrobacterium* (see Chapter 3), which is involved in cytokinin production, under the control of the *SAG12* promoter delays senescence in tobacco.

BOX 4.1 *Continued*

Abscisic acid-inducible gene expression

Sequence elements responsible for abscisic acid (ABA) (see Chapter 2) induction of gene expression have been identified (see Chapter 1) and if combined with the core 35S promoter in a chimaeric promoter confer ABA inducibility on transgene expression.

Auxin-inducible gene expression

Similarly, several sequence elements responsible for auxin (see Chapter 2)-induced gene expression have been identified. These auxin-response elements (AuxREs) when combined with a core 35S promoter confer auxin (and inactive auxin analogue) inducibility on transgene expression.

Synthetic promoters for inducible gene expression

In some of the previous examples, we have seen how sequence elements from genes that, for example, confer heat and ABA inducibility, can be combined with core promoters to effectively build synthetic promoters that confer the desired expression characteristics on the transgene. As more and more *cis*-elements are identified and characterized, the possibility of building synthetic promoters combining one or more of these elements with a minimal promoter to confer the desired expression characteristics becomes more realistic. Multiple sequence elements responsible for pathogen and wound induction, tissue-enhanced expression, cell-specific expression, and transcriptional up- and down-regulation have all been identified. Combinations of several sequence elements involved in pathogen-induced expression have been combined with a minimal promoter to create synthetic promoters that have been tested and shown to function effectively as pathogen-induced promoters, with the added advantage that multiple elements involved in the same response may act synergistically. It may also be possible to construct synthetic promoters that confer the characteristics of several of these elements in some cases. Construction of these synthetic promoters is not without its difficulties, as the spatial arrangement of the various *cis*-elements has to be controlled carefully and can be difficult to predict. However, the potential benefits of synthetic promoters suggests an important role in driving transgene expression, at least in demanding situations requiring precise regulation of expression.

The advantage of non-plant-derived systems is that they are independent of normal plant processes, requiring the application of a specific substance to induce expression. This is, however, also their weakness, as the application of (in some cases expensive) inducers on an agricultural scale is, in most but not all cases, not feasible or economically viable (see Box 4.1 for an example where it may prove to be both feasible and economically viable). Experimentally, however, some of these have proved to be quite useful.

Systems based on plant-derived components do not have the advantage of independence from normal plant processes. This does, however, make their use in agriculture potentially simpler as the application of an inducer is not required. It should be noted that no one system is suited to all situations, and potentially more damaging is the fact that some of the systems developed have proved to be difficult to

use or have proved to be 'leaky' (i.e. they direct expression even in the absence of the inducing agent).

Selectable markers

Selectable marker genes introduced during transformation allow transformed tissue to be selected, normally by conferring resistance to a toxic substance.

Plant transformation is, in many cases, a very-low-frequency event. It is therefore vital that some means for selecting the transformed plant tissue is provided by the plant transformation vector. In most cases this selection is based on the inclusion into the culture medium of a substance that is toxic to plants. The selectable marker on the vector confers resistance to the toxic substance when expressed in transformed plant tissue. In many cases, as in the situation when using *E. coli*, antibiotic-resistance genes can be used as selectable markers in plants. Although plants are eukaryotic, antibiotics efficiently inhibit protein synthesis in the organelles, particularly the chloroplasts. Perhaps the most widely used selectable marker gene is the *nptII* gene, which confers resistance to the antibiotic kanamycin. Other antibiotic-resistance genes have also been used with some considerable success in plant transformation vectors. In part, this use of alternative selectable marker genes was driven by the observations that some plant species exhibited a very high degree of natural resistance to kanamycin (such as cereals), and that some species (some soft fruits for example) were too sensitive to kanamycin for it to be used successfully. This made the selection of transformed tissue very difficult, leading to a high number of false-positive results or the inability to recover transformed plants. Thus, the use of any selectable marker has to be controlled closely, with appropriate concentrations being determined by kill-curves, on a case-by-case basis. The development of alternative selectable marker genes also allows for the re-transformation of plant tissue that already expresses one or more different selectable markers. Thus, genes conferring resistance to antibiotics such as bleomycin, spectinomycin, and hygromycin are used quite widely. These selective agents can be used at lower concentrations than kanamycin, and therefore usually result in a cleaner selection of transformed tissue. Other resistance genes can also be used as selectable markers in plants. Among those used widely are genes that confer resistance to herbicides such as chlorsulphuron and bialaphos (see Chapter 5).

Public concern has, in recent years, questioned the general desirability of growing transformed crops, and, more specifically, the merits of using antibiotic- or herbicide-resistance genes as selectable markers during plant transformation. There is obvious concern about the creation of so-called super-weeds and the transfer of antibiotic-resistance genes. Although in the long term it is undoubtedly better to remove these selectable marker genes from the transformed plants once their job has been done (this issue will be considered in more detail later in this chapter), other, more acceptable, selectable marker genes are also being introduced.

Some of the most interesting are based on the principal of facilitating alternative carbon-source utilization. Thus, genes from bacteria that allow the use of mannose or xylose (and, in fact, other carbohydrates) as carbon sources have been successfully used as selectable markers in plant transformation. These genes have also generally proved to be superior to standard antibiotic-selection genes in some transformation

protocols. These are examples of positive-selection methods, so-called because untransformed tissue does not die (due to the presence of low concentrations of sucrose in the medium used for regeneration) but does not proliferate.

Other potential selectable markers are also being investigated; these include: the reporter gene *gfp*, which can be used for the visual screening of transformants; the *Agrobacterium* isopentenyl transferase (*ipt*) gene, which determines growth and morphology during regeneration; and the use of cytokinin glucuronides in combination with β-glucuronidase (GUS). Examples of various selectable markers are shown in Table 4.1.

Reporter genes

● Reporter genes confer an easily screened phenotype on transformed plant tissues.

Reporter genes (see Table 4.2) are widely used in plant transformation vectors, both as a means of assessing gene expression by promoter analysis (see Chapter 1) and as easily scored indicators of transformation (indeed, they have been used in place of selectable marker genes in some cases). Ideally, reporter genes should be easy to assay, preferably with a non-destructive assay system, and there should be little or no endogenous activity in the plant to be transformed. At present, only a small number of reporter genes are in widespread use in plant transformation vectors, these being β-glucuronidase (*uidA* or *gus*), green fluorescent protein (*gfp*), luciferase genes (*lux* and *luc*), and, to a lesser degree (although it is widely used in animal systems), the chloramphenicol acetyltransferase gene (*cat*). Each of these reporter genes will be looked at in more detail.

β-Glucuronidase

This is perhaps the most widely used reporter gene in plant transformation vectors. Its widespread acceptance is due to its many advantages over the previously existing marker genes. GUS can be assayed extremely sensitively using quick, easy, and non-radioactive methods. It can be used to obtain both quantitative (i.e. the level of gene expression) and qualitative (i.e. localization of gene expression) data. There is also little or no endogenous activity in most plant tissues (with the possible exception of reproductive tissues).

Quantitative data is obtained by assays utilizing fluorogenic substrates such as 4-methylumbelliferryl-β-D-glucuronide (4-MUG), which is hydrolysed to 4-MU (4-methylumbelliferone). Standard enzyme assay protocols can be used and results compared with a standard curve of 4-MU fluorescence. Qualitative data can be obtained from histochemical assays that allow tissue- and cell-specific localization of GUS activity. These histochemical assays are conducted *in situ* with the chromogenic substrate 5-bromo-4-chloro-3-indolyl β-D-glucuronide (X-gluc). GUS activity results in the deposition of an insoluble blue precipitate, effectively identifying the precise location of expression.

In order to avoid problems associated with GUS expression in *Agrobacterium*, a *gus* gene containing an intron (which can therefore only be expressed in eukaryotes) is often used in plant transformation vectors.

Table 4.1 Selectable marker genes used in plant transformation, their source, and mode of action

Selectable marker gene	Abbreviation	Source of gene	Selection mechanism	Selective agent
Antibiotic resistance				
Aminoglycoside adenyltransferase	*aadA*	*Shigella flexneri*	Antibiotic resistance	Streptomycin, spectinomycin
Bleomycin resistance	*ble*	*E. coli*	Antibiotic resistance	Bleomycin
Dihydropteroate synthase	*sul/dhps*	*E. coli*	Antibiotic resistance	Sulphonamides
Dihydrofolate reductase	*dhfr*	Mouse	Antibiotic resistance	Methotrexate
Hygromycin phosphotransferase	*hpt/aphIV/hyg*	*E. coli*	Antibiotic resistance	Hygromycin
Neomycin phosphotransferase II	*nptII/neo*	*E. coli*	Antibiotic resistance	Kanamycin, geneticin (G418)
Neomycin phosphotransferase III	*nptIII*	*Streptococcus faecalis*	Antibiotic resistance	Kanamycin, geneticin (G418)
Herbicide resistance				
Acetolactate synthase	*als*	*Arabidopsis* spp./ maize/tobacco	Herbicide resistance	Sulphonylureas
Enolpyruvylshikimate phosphate synthase	*epsps/aroA*	*Petunia hybrida/ Agrobacterium* spp.	Herbicide resistance	Glyphosate
Glyphosate oxidoreductase	*gox*	*Achromobacter LBAA*	Herbicide resistance	Glyphosate
Phosphinothricin acetyltransferase	*bar/pat*	*Streptomyces hygroscopicus/S. viridochromogenes*	Herbicide resistance	Bialophos, glufosinate, L-phosphinothricin
Cyanamide hydratase	*cah*	*Myrothecium verrucaria*	Herbicide resistance	Cyanamide
Bromoxynil nitrilase	*bxn*	*Klebsiella pneumoniae*	Herbicide resistance	Bromoxynil
Others				
Lysine–threonine aspartokinase	*lysC*	*E. coli*	Resistance to high levels of lysine or threonine	Lysine and threonine
D-Serine ammonia lyase	*dsdA*	*E. coli*		

Bromoxynil nitrilase *bxn*

Others				
Lysine–threonine aspartokinase	*lysC*	*E. coli*	Resistance to high levels of lysine or threonine	Lysine and threonine
D-Serine ammonia lyase	*dsdA*	*E. coli*	Resistance to D-serine	D-Serine
Mannose-6-phosphate isomerase[a]	*pmi/manA*	*E. coli*	Alternative carbon source	Mannose
Xylose isomerase	*xylA*	*Thermoanaerobacterium thermosulfurogenes/ Streptomyces rubiginosus*	Alternative carbon source	Xylose
Arabitol dehydrogenase	*atl*D	*E. coli*	Alternative carbon source	Arabitol
Trehalose-6-phosphate synthase	*AtTPS1*	*Arabidopsis*	Resistance to high levels of glucose	Glucose
β-Glucuronidase	*gus/uidA*	*E. coli*	Complementation of missing medium component	Cytokinin glucuronide
Betaine aldehyde dehydrogenase[b]	*badh*	Spinach	Detoxification	Betaine aldehyde
Green fluorescent protein	*gfp*	*Aequorea victoria* (jellyfish)	Visual screening	None
Isopentenyl transferase[c]	*ipt*	*Agrobacterium* spp.	Growth in absence of exogenous cytokinin	None
Leafy cotyledon[d]	*lec*	Maize	Growth	None
Knotted1	*Kn*1	Maize	Growth	None

[a]Also known as phosphomannose isomerase.
[b]See Chapter 9. Used as a selectable marker in chloroplast transformation.
[c]If used with a constitutive promoter, plants have poor root formation and are often sterile, thus limiting the use of the *ipt* gene as a selectable marker. However, use with an inducible promoter has resulted in the development of an antibiotic-free selectable-marker system.
[d]Leafy cotyledon gives the transformed cells a competitive growth advantage, and when coupled with a visual marker to identify transformed cells can be used as an antibiotic-free selectable-marker system.

Table 4.2 Reporter genes used in plant transformation

Reporter gene	Abbreviation	Source of gene	Detection/assay
β-Glucuronidase	*gus/uidA*	*E. coli*	Fluorimetric (quantitative) or histochemical (*in situ*), non-radioactive
Green fluorescent protein	*gfp*	*Aequorea victoria* (jellyfish)	Fluorescence, non-destructive
Chloramphenicol acetyltransferase	*cat*	*E. coli*	Radioactive assay of plant extract, sensitive, semi-quantitative
Luciferase	*luc*	*Photinus pyralis* (firefly)	Luminescence
Luciferase	*luxA*, *luxB*	*Vibrio harveyi*	Luminescence

Green fluorescent protein (GFP)

Nowadays, the green fluorescent protein gene (*gfp*) is rapidly becoming a very widely used reporter gene. GFP has the advantages that it is even easier to assay than GUS (given the correct equipment) and the assay is non-destructive (although non-destructive assays for GUS activity have been reported, they have not been adopted widely). This means that GFP can be used in situations where GUS cannot, for example in screening primary transformants, in time-course experiments, or for analysing segregation in small seedlings.

The gene was isolated from the jellyfish *Aequorea victoria*, which are brightly luminescent organisms. To work efficiently in plants it was found that the *gfp* gene had to be modified significantly in order to: (1) remove a cryptic intron (i.e. *gfp* mRNA is efficiently *mis*-spliced in some plants, resulting in the removal of 84 nucleotides); (2) make the codon usage more 'plant-like'; and (3) prevent accumulation in the nucleoplasm.

Luciferases

The firefly luciferase gene (*luc*) encodes an enzyme that catalyses the oxidation of D-luciferin in an ATP-dependent fashion. This oxidation results in the emission of light. Highly sensitive assays (based on photomultipliers, luminometers, or film exposure) have been developed to detect the extremely rapid emission of light. Other assay systems are also available commercially. The firefly luciferase gene is not a widely used marker gene as assaying the gene product is difficult, but it is useful for the detection of low-level or highly localized expression.

The bacterial luciferase genes (*luxA* and *luxB* from *Vibrio harveyi*) are found in a few plant transformation vectors. They catalyse the oxidation of long-chain fatty aldehydes resulting in the emission of light.

Chloramphenicol acetyltransferase

Although widely used as a reporter gene in mammalian cells, the availability of the GUS and GFP reporter systems has generally limited the use of the chloramphenicol acetyltransferase (CAT) system in plants, although it was the first bacterial gene to be expressed in plants. Some plant transformation vectors do, however, carry this reporter gene as it can be assayed very sensitively, but it requires a radioactive assay procedure.

Origins of replication

As both *E. coli* and *Agrobacterium* are Gram-negative bacteria, a single origin (*ori*) with a broad host range can be used to allow vector replication in both *E. coli* and *Agrobacterium*. However, the broad-host-range origins used in many plant transformation vectors (e.g. the RK2 origin used in pBIN19 and derivatives) often results in a low plasmid copy number in *E. coli*, making manipulation of the DNA more difficult or time-consuming. An alternative approach is to use two separate origins of replication, one to enable replication in *Agrobacterium* and a second to direct replication in *E. coli*. This second origin is usually derived from a high-copy-number cloning vector (such as the *Col*E1 *ori* from pBR322 or the pUC *ori*). This results in a higher copy number in *E. coli*, therefore facilitating manipulation, but it also results in a larger vector (as two separate origins are present) with the associated difficulties.

Co-integrative and binary vectors

Initially, vectors used for *Agrobacterium*-mediated plant transformation were deletion derivatives of Ti plasmids that were termed co-integrative vectors. The DNA to be introduced into the plant transformation vector is cloned into an intermediate vector (which is small and easily manipulated *in vitro*) and recombination within the bacterium is used to clone the sequences. In most systems of this type, the modified T-DNA (containing the gene of interest) is on the same molecule as the transfer apparatus as part of one Ti plasmid. Vectors of this type have become less common as binary vector systems are more easily manipulated.

In binary vectors, the transfer apparatus (i.e. the *vir* genes) and the T-DNA are located on separate plasmids (Figure 4.3). As only the border sequences are needed to define the T-DNA region and the *vir* region is absent, binary vectors are relatively small (and therefore easily manipulated). Box 4.2 describes the development of binary vectors.

Families of binary vectors

It soon became apparent that, although vectors such as pBIN19 (which offer one selectable marker) were sufficient for routine transformation studies, the application of transgenic technology to an increasing number of plant species required a more flexible set of binary vectors.

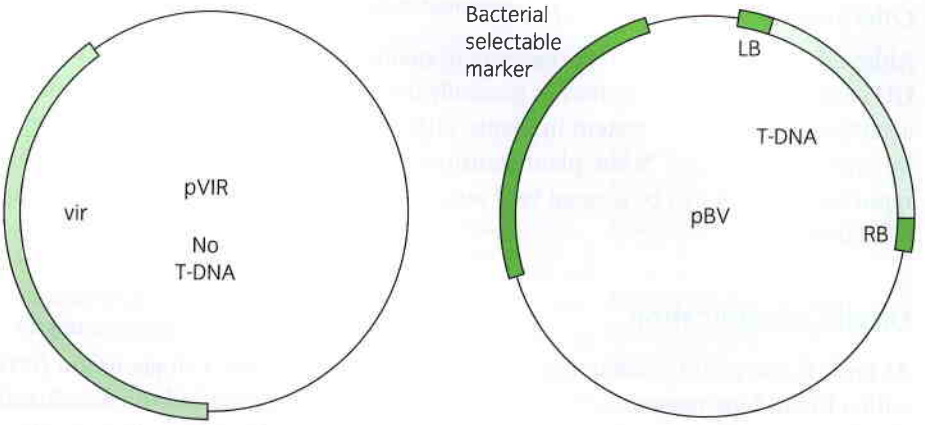


Figure 4.3 The binary-vector strategy illustrated by a pair of hypothetical plasmids. One plasmid (pVIR in this case, a derivative of a Ti plasmid) carries the *vir* genes. This plasmid has no T-DNA region, including border sequences, and therefore no T-DNA can be transferred to the plant genome. A smaller binary vector (pBV, which is over 20 times smaller than pVIR) has a T-DNA region defined by the presence of the left and right borders (LB, RB). As there are no *vir* genes on this plasmid (making the binary vector pBV relatively small), their function is supplied *in trans* by pVIR. pVIR would normally be maintained in suitable strains of *Agrobacterium* and is not generally isolated or manipulated. Binary vectors are designed to be easily manipulated using standard *in vitro* techniques and contain selectable markers for selection in bacteria (*E. coli* and *Agrobacterium*) outside the T-DNA region. The T-DNA region will, in all probability, contain a selectable marker for use in plants. pBV is normally maintained in *E. coli*, the gene of interest (the transgene) being cloned into it. Recombinant *E. coli* colonies are identified after bacterial transformation, and the pBV vector is checked to ensure that the transgene is correctly inserted and has not been mutated. pBV can then be transferred, by electroporation, into the *Agrobacterium* strain used for plant transformation (which also harbours pVIR).

Thus, modern binary vectors tend to come in families. These offer a choice of selectable marker (both antibiotic and herbicide resistance in many cases) and, in some cases, the opportunity for dual selection by incorporating two selectable markers on the one vector. This dual selection approach allows a much more rigorous selection pressure to be applied to plants, in which it has traditionally proved difficult to avoid false positives during the selection process.

Optimization

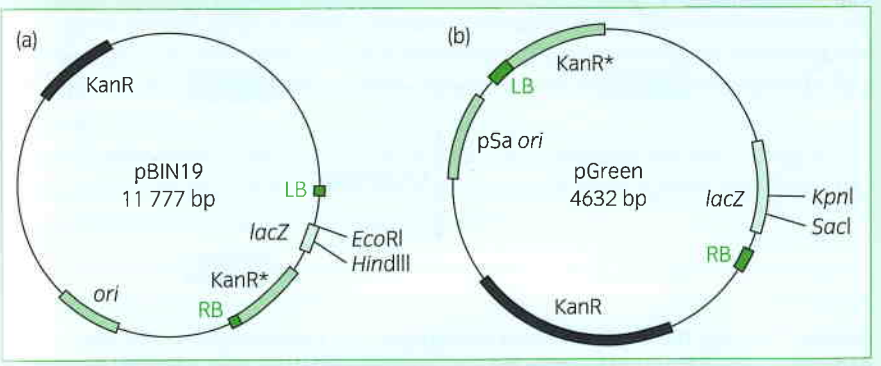
Many features of transformation-vector design influence transgene expression.

Considerable advances have been made in the design of plant transformation vectors (see Box 4.2) and the ease with which these vectors can be produced (see Box 4.3 for an example of how plant transformation vector construction has been simplified). However, despite the improvements made in vector design and advances in our understanding of both the mechanisms of transgene integration and plant gene expression, plant transformation is still in many ways an imprecise art. Several aspects of vector design are known to influence the efficiency of transgene expression.

BOX 4.2 *The evolution of binary plant transformation vectors*

Having looked at the basic requirements for plant transformation vectors, we can look at how the design of plant transformation vectors has improved over the years. The binary plant transformation vectors pBIN19 and pGreen can be compared (see figure).

pBIN19 was one of the first binary plant transformation vectors (developed in the early 1980s) and has been one of the most widely used. pGreen is a new binary vector (one of a family available with different promoters, selectable markers, and reporter genes). As can be seen, pBIN19 (11 777 bp) is much larger than pGreen (4632 bp) and is therefore that much more difficult to work with. This reduction in size is, at least in part, due to the unusual origin of replication present in pGreen. The pSa origin is in itself small, but in pGreen it has been further divided so that only the replication origin is present on the plasmid. The replicase gene is present on another plasmid and functions *in trans*. The restriction sites in pBIN19 available for use in cloning are limited in comparison with pGreen, although in both cases blue/white screening is possible. The selectable marker, which in both cases is *nptII*, is situated close to the right border in pBIN19, but in pGreen it is situated close to the left border. The advantage of having the selectable marker next to the left border becomes apparent if one considers that there is a polarity in T-DNA transfer (Chapter 3). Transfer originates from the right border and therefore the selectable marker will be transferred before the gene of interest if the selectable marker is next to the right border. If T-DNA transfer is interrupted, this could result in plants expressing the selectable marker (and therefore being isolated as genuine positive transformants) but not containing, or containing a truncated version of, the gene of interest.



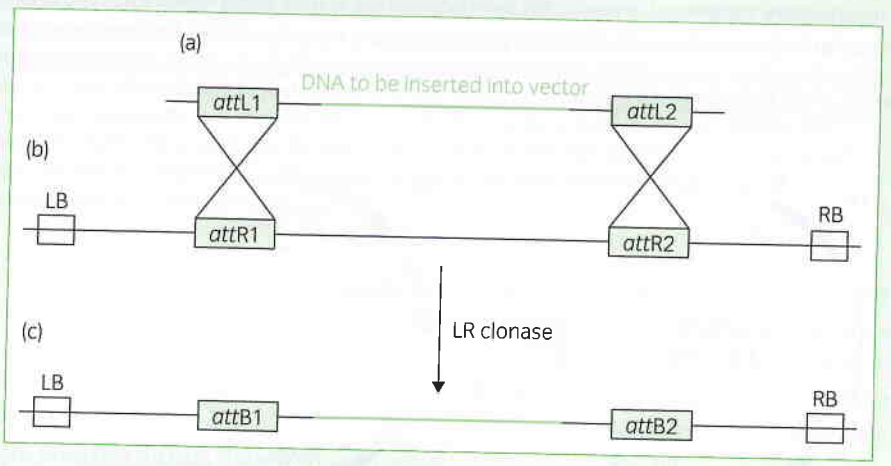
Simplified representation of the binary plant transformation vectors (a) pBIN19 and (b) pGreen (in this case pGreen 0029), showing the pertinent features. pBIN19 is a large plasmid, making manipulations using this plasmid difficult. The selectable marker for selecting transformed plants (KanR*) is located close to the right border (RB), leading to the possibility of false positives (due to truncation of the T-DNA during transfer, kanamycin-resistant plants that do not express the transgene might be isolated). The MCS, situated between the *Eco*RI and *Hin*dIII sites, has eight restriction enzyme sites. The MCS is in the *lacZα* fragment so blue/white screening for recombinant colonies is possible. pGreen 0029 is one of a large family of plant transformation vectors. It is much smaller than pBIN19 and has a much larger MCS (with over 15 restriction enzyme sites) situated in *lacZα* (between the *Kpn*I and *Sac*I sites), so blue/white screening is possible. The pSa origin is small and contributes to the small overall size of pGreen 0029. (Redrawn with permission from figures on the pGreen website, www.pgreen.ac.uk/a_hom.htm.)

BOX 4.3 *Gateway™ vectors*

In the introduction to this chapter we discussed how many of the developments in plant transformation vector design were aimed at making manipulations involving the vector simpler. Binary vectors, even modern ones such as pGreen (shown in Box 4.2), tend on the whole to be quite large and the cloning of sequences into these vectors is time-consuming and laborious. Recently though, a new technology developed by Invitrogen in the USA, termed Gateway™ conversion technology, has opened up the possibility of rapid, facile cloning of sequences into suitably designed binary vectors. Key advantages of Gateway technology are that the cloning of DNA into plant transformation vectors is not dependent on the constraints of sequence that are normally imposed by the need to use restriction enzymes in conventional cloning strategies, and so manipulations are eased and less laborious. Once a piece of DNA is flanked by suitable sites that allow Gateway technology to be used, it can be easily cloned into a large and expanding range of suitable Gateway-compatible plant transformation vectors with the minimum of manipulation.

Gateway cloning

Gateway technology depends in essence on recombination technology, as opposed to the more standard cloning technology based on restriction enzymes and DNA ligase. Later in this chapter



Gateway™ cloning. The DNA to be introduced into the plant transformation vector (the destination vector) is flanked by *att* (in this case *attL*) recombination sites, which are small (25-bp) sequence elements. The destination vector contains similar *attR* sites between the left (LB) and right (RB) borders that define the T-DNA. The LR clonase™ enzyme causes exchange of the DNA fragments between the recombination sites as indicated. The destination vector now contains the DNA fragment flanked by two *attB* recombination sites, which means that it can be further mobilized by recombination into a vector containing *attP* sites (the reaction being catalysed by another clonase enzyme). The efficiency of the system is further improved by enabling positive selection of destination vectors that have the DNA of interest inserted. This positive selection comes from the inclusion of the *ccdB* gene, which prevents the growth of *E. coli*, between the *attR* sites of the destination vector. Only *E. coli* harbouring destination vectors in which the *ccdB* gene has been replaced by the DNA of interest through recombination will grow.

BOX 4.3 *Continued*

the role of site-specific recombinases in facilitating clean-gene technologies is explained; here though a different site-specific recombinase system is exploited. In Gateway cloning, the site-specific recombinase system used by bacteriophage lambda (phage λ) is utilized.

Any DNA fragments that need to be cloned into a plant transformation vector that are flanked by recombination sites will, in the presence of an enzyme, recombine with recombination sites in the vector, leading to incorporation of the fragment into the plant transformation vector. The precise way in which the overall cloning scheme works varies slightly depending on which group has developed the vectors, but all are based on this simple principle.

The recombinase sites that flank the DNA that is to be cloned into the plant's transformation vector can be introduced simply in a variety of ways. One common method is to use topoisomerase-mediated cloning, developed by Invitrogen. In this approach, the polymerase chain reaction (PCR) is used to amplify the target sequence using a forward primer that has a CACC motif at its 5′ end. This 5′ CACC motif mediates directional incorporation of the PCR product into a suitable vector that results in the PCR-amplified target DNA being flanked by *att*L recombination sites. The technology can be used with more traditional approaches as well. Recombination sequences can be introduced into vectors so as to flank MCSs and can be included in PCR primers. These different variations require subtle modification of the technology but have all been used successfully.

Gateway-compatible plant transformation vectors

A variety of laboratories have developed destination vectors for use in plants, many of which are available, through either a material transfer agreement or freely. These destination vectors have been developed with a range of applications in mind, from simple constitutive expression of transgenes, through inducible expression, to protein interactions, protein localization, and protein purification.

The key advantages of Gateway technology in facilitating the quick and easy cloning of DNA into plant transformation vectors means that such vectors are ideally suited to high-throughput applications associated with genomics and proteomics.

Arrangement of genes in the vector

Analysis of the efficiency of transgene expression allows a few simple 'rules' about vector design to be drawn up. If more than one gene is in the vector, different promoters and terminators should be used for each. The use of the same promoter and/or terminator can lead to an increase in gene silencing (Box 4.4 discusses how vectors designed specifically to induce gene silencing can be produced). Multiple genes on one vector should not be immediately adjacent to each other, and should be in the same orientation. This avoids adjacent inverted repeats that cause plasmid instability in bacteria and increased gene silencing in plants.

BOX 4.4 *Gene silencing*

In Chapter 1 we saw how early experiments with transgenic plants resulted in the discovery of gene silencing, and instability of transgene expression is still a problem encountered in experiments involving transgenic plants. There are a number of practical steps in vector design that can be taken to reduce the risk of gene silencing. Reducing the number of transgenes inserted (the copy number), avoiding the use of multiple copies of the same promoter or terminator (particularly in repeat structures), and avoiding using promoters and transgenes with a high degree of similarity to endogenous ones can help to reduce the probability of gene silencing.

However, despite the problems presented by unintended gene silencing, no discussion of plant transformation vectors would be complete without looking at how vectors deliberately designed for gene silencing can be designed.

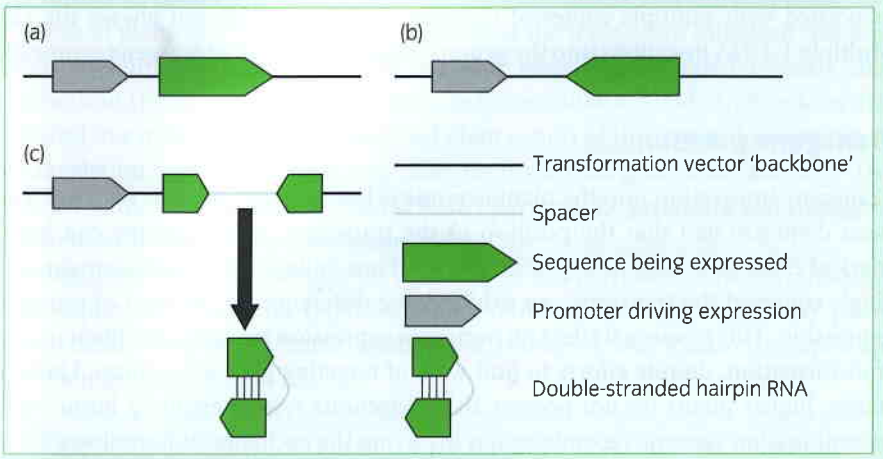
A full explanation of gene silencing is given in Chapter 1, but a short introduction to the pertinent points as far as vector design is concerned is given here. Two observations were key in leading to the development of gene-silencing strategies for the down-regulation of gene expression in transgenic plants: attempts to achieve high-level transgene expression could paradoxically result in very low or no transgene expression and the expression of sense-orientation transgenes as controls for antisense RNA experiments resulted in gene silencing. In both cases, the silencing was post-transcriptional, and dependent on the production of double-stranded (ds) RNA that is subsequently processed by nucleases into small interfering RNAs (siRNAs), which direct the degradation of homologous RNA by a different nuclease. This post-transcriptional dsRNA-mediated gene silencing is known as RNA interference (RNAi). By constructing plant transformation vectors that indirectly or directly lead to the production of dsRNA, RNAi can be deliberately induced to silence the expression of specific genes in transgenic plants.

RNAi vectors

Normally, when expression of a transgene is desired, the arrangement of promoter and sequence to be expressed mimics the usual arrangement of the gene, so that the promoter directs expression of sense RNA that is in the correct orientation to be translated (panel a in the figure). The simplest RNAi-related technology is to produce so-called antisense RNA. In vectors designed to produce antisense RNA, the sequence to be expressed (which can be all or a part of the coding sequence of an endogenous gene) is placed in the reverse orientation with respect to the promoter (panel b in the figure). The promoter therefore directs the production of antisense RNA, which is complementary to the endogenous mRNA. The complementarity between antisense RNA and mRNA results in the production of dsRNA.

However, antisense RNA production is not the most effective way of directing RNAi; far more efficient is the production of hairpin RNA (hpRNA). In a vector designed to produce hpRNA, the transgene sequence is cloned as an inverted repeat, with the repeat units separated by a spacer (panel c in the figure). The expression of the inverted repeat is driven by a strong promoter such as the 35S promoter for dicotyledonous plants or the ubiquitin 1 promoter for monocotyledonous plants. The inverted-repeat structure allows the two complementary regions to double-back on each other, pivoting at the spacer region. This double-stranded hpRNA is extremely efficient in RNAi-induced destruction of any homologous mRNA.

BOX 4.4 *Continued*



A simplified representation of constructs used for RNAi. When expression of a transgene is required, the arrangement of promoter and sequence to be expressed mimics the usual arrangement of the gene, so that the promoter directs expression of sense RNA (a). In antisense constructs, the orientation of the sequence to be expressed is reversed relative to the promoter, so that antisense RNA is produced (b). More efficient down-regulation is achieved by the expression of hpRNA. The expressed sequence is composed of an inverted repeat, with the repeat units separated by a spacer (c). The spacer is usually composed of an intron sequence, which is necessary for construct stability during cloning and the promoter is often one driving high-level constitutive expression. When the sequence is expressed from the promoter, the RNA can form a hairpin structure with a loop formed from the intron spacer region. This hpRNA directs RNAi-mediated destruction of any homologous mRNA.

hpRNA-induced RNAi has many advantages as a technology for down-regulation of gene expression.

- hpRNA-induced RNAi may be stably inherited, unlike some other mechanisms that induce post-transcriptional gene silencing. The stability of any introduced phenotypes in transgenic plants is obviously crucial, particularly if applied to crop plants. hpRNA-induced RNAi is known to be inherited through multiple generations in *Arabidopsis*.
- Down-regulation of gene expression by hpRNA-induced RNAi is also effective when the target RNA is encoded by a multigene family. Down-regulation of multigene families is difficult to achieve with other technologies used for down-regulation.
- The degree of down-regulation can also be fine-tuned by either moderating the expression level by the choice of an appropriate promoter (although very weak promoters are not suitable) or by using sequences with varying degrees of identity to the target RNA.
- The expression of the hpRNA can be restricted to given organs by the use of organ-specific promoters or can be made to respond to various stimuli by the use of inducible promoters.
- hpRNA constructs are relatively simple to make, and can be combined with Gateway™ technology to facilitate rapid vector construction.

Transgene copy number

In any plant transformation protocol, multiple copies of the transgene can be incorporated into the target plant genome. Often high levels of transgene expression are associated with multiple copies of the transgene, but this is not always the case. Multiple T-DNA insertions into the genome may lead to erratic transgene expression.

Transgene position

Transgene integration into the plant genome is basically a random event, and it has been demonstrated that the position of the transgene in the genome can have a marked effect on transgene expression levels. Thus, independent transformants with single copies of the transgene can exhibit large differences in the level of transgene expression. This positional effect on transgene expression remains a problem in plant transformation, despite efforts to find ways of targeting gene integration. Unfortunately, higher plants do not possess an endogenous system enabling homologous recombination (genetic recombination involving the exchange of homologous loci) at a high efficiency (but see Chapter 5) that allows transgenes to be targeted to a particular region of the genome. However, alternatives utilizing bacteriophage or yeast site-specific recombinases have been used with some success.

One approach taken to assuage position effects is the inclusion of matrix-attachment regions (MARs). These AT-rich sequences are thought to be involved in maintaining the chromatin in an open structure allowing gene expression. The inclusion of MARs flanking transgenes has been shown to provide position-independent expression in animal systems, and some, albeit limited, success has been achieved with their use in plants. MARs may also help to stabilize transgene expression.

Transgene features

It is well documented that heterologous genes, particularly those from non-plant species, tend to be expressed poorly in plants, even when driven by a strong promoter. This can be due to a variety of factors, many of which are associated with the structure of the transgene.

Genes from different organisms tend to have different G+C contents, with bacterial genes having a particularly low G+C content. It has been found that a high A+T content in transgenes interferes with mRNA processing in plants, leading to little or no expression of the transgene. Differences in A+T content between the transgene and the isochore (long stretches of DNA with a homogeneous base composition) into which it is integrated may also contribute to the transgene being recognized as foreign. A high A+T content also often results in the presence of sequences (AUUUA) that can destabilize mRNA and may also form potential plant polyadenylation signals. So-called cryptic introns may also be found. These can be efficiently mis-spliced from transgenes (i.e. part of the coding sequence is recognized by the plant as an intron and spliced out) by the plant's mRNA-processing machinery, effectively resulting in a deletion mutation of the transgene.

Even if the transgene is transcribed efficiently it may not be translated efficiently, probably due to the architecture of the translation initiation region and the presence of codons that are used infrequently by plants. Consensus sequences flanking the ATG initiation codon are known to differ between plants and other species. The inclusion of plant-specific sequences upstream of the translation initiation codon is known to improve translatability. It has also been demonstrated recently that modification (by the insertion, after the translation initiation codon, of codons that are found in a number of highly expressed plant genes) of the region downstream of the translation initiation codon may also improve transgene translatability. The inclusion of the tobacco mosaic virus Ω leader between the promoter and transgene has also been shown to improve translatability.

Mutagenesis can be used to alter the codon usage of a particular sequence, although if the sequence is large and requires a large number of changes, this may prove difficult and/or expensive. Some examples of transgene modification to optimize expression are given in Box 4.5.

BOX 4.5 *Examples of optimization of vector components*

Modification of the *Bacillus thuringiensis cry*1A(b) gene

The *cry*1A(b) gene encodes a protein that can be used to control insect pests. The wild-type gene was found to be expressed at relatively low levels in transgenic plants. The effect of modifying the sequence of the gene on expression was investigated. Codon usage and overall G+C content was altered and several potential plant polyadenylation signals and ATTTA sequences were removed. These changes were seen to result in large increases (up to 100-fold) in the levels of the *cry*1A(b) gene product.

	Wild-type gene	Modified gene
No. of altered bases (%)	0	390 (21%)
No. of modified codons (%)	0	356 (60%)
G+C content (%)	37	49
No. of potential polyadenylation sites	18	1
No. of ATTTA sequences	13	0

(Data taken with permission from Perlak *et al.* (1991).)

Modification of the *gfp* gene

The *Aequorea victoria gfp* gene is used as a reporter gene in plant transformation experiments. To improve its use in plants, and in particular in *Arabidopsis*, it has been modified extensively. A cryptic intron that is efficiently mis-spliced in some plant species was removed by altering codon usage (altering the codon usage has also been demonstrated to increase the expression level). Translatability of the gene was improved by introducing an AACA sequence upstream of the translation initiation codon. In order to prevent build-up of GFP in the nucleoplasm (where it may prove to be mildly toxic), a 5′ terminal signal peptide and a 3′ endoplasmic reticulum retention signal (HDEL, His–Asp–Glu–Leu) were introduced to target the GFP to the endoplasmic reticulum. (Data taken with permission from Haselhoff *et al.* (1997).)

Clean-gene technology

New technologies are being developed to reduce the possible environmental impact and increase the acceptability of transgenic plants.

Since the introduction of transgenic crop plants into agriculture, public concern about the safety of such technology, particularly about the safety of antibiotic- and herbicide-resistance genes, has increased to the point where governments have been forced to take action (see Chapter 12). Whatever the scientific merits of the argument, it is a fact that the perceived dangers of transgenic crops are preventing their widespread adoption, particularly in Europe. Even in countries where the growth of such crops is not seen as a major issue (like the USA), the effect is still felt as export markets for foodstuffs derived from genetically engineered crops are affected. There has therefore been a move to so-called clean-gene technologies where the selectable marker is no longer present in the field-grown crop, or, as described earlier, a more acceptable marker gene is used. To develop crops that contain no selectable marker genes, several approaches are available, as follows.

1 Theoretically, the use of selectable marker genes could be avoided altogether by simply inserting the transgene of interest. Molecular biology techniques, such as the polymerase chain reaction (PCR), could be used to screen populations of putative transformed plants for those containing the transgene. In practice, this approach would simply be too expensive and labour-intensive in most situations.

2 A second approach is to introduce the gene of interest and the selectable marker genes on separate (unlinked) T-DNA molecules. Subsequent segregation analysis of the offspring can be used to identify plants that contain the gene of interest but not the selectable marker. Various strategies to deliver separate T-DNAs have been developed, the most effective appearing to be those in which one binary vector contains several T-DNA regions that are integrated at unlinked sites in the plant genome.

3 An alternative approach is to excise the selectable marker from the plant genome. This can be done by utilizing site-specific recombinase systems. Several recombinase systems (phage P1 Cre-*lox*, *Saccharomyces cerevisiae* FLP-*frt*, and the R-*RS* system of *Zygosaccharomyces rouxii*) are available. Vectors in which the selectable marker gene is placed between recognition sites (*lox*, *frt*, or *RS*) for a site-specific recombinase can be used to transform plants. Subsequently, the recombinase (Cre, FLP, R) is expressed (by transient expression, crossing of plants with recombinase-expressing plants, or by using an inducible expression system), resulting in the selectable marker gene being excised from the plant genome.

4 Recently, vectors have been demonstrated in which the selectable marker gene was flanked by sequences that apparently considerably increase the frequency of intrachromosomal recombination leading to excision of the marker gene.

SUMMARY

This chapter has looked at the construction of binary vectors for use in *Agrobacterium*-mediated transformation strategies. These binary vectors are, in effect, highly specialized derivatives of the *Agrobacterium tumefaciens* Ti plasmid looked at in Chapter 3. Key

to their successful use in plant transformation protocols are the modifications made, resulting in removal of the oncogenes (thereby preventing abnormal plant development) and a general improvement in their ease of use. *Agrobacterium*-mediated transformation with binary vectors has been successfully applied to a wide range of crop plants, but improvements to many aspects of vector design are still being made. Some of these improvements are based on purely scientific grounds, to prevent or reduce gene silencing for instance, whereas others are based on widening the applicability of binary vectors (for example, alternative selectable markers). Other modifications are being made in response, at least in part, to concerns (public and scientific) about the safety of such vectors. These concerns, mainly over selectable marker genes, are particularly evident in Europe. In the UK, the Royal Society's 1998 report, *Genetically Modified Plants for Food Use*, questioned the future use of antibiotic-resistance genes and recommended that if they were used they should be removed at an early stage of development of the genetically modified plant. The Advisory Committee on Releases to the Environment (ACRE), in its *Guidance on Principles of Best Practice in the Design of Genetically Modified Plants*, similarly noted the desirability of reducing the amount of 'extraneous' DNA (such as marker genes) in transgenic plants. Clean-gene technologies and alternative selectable markers are therefore becoming increasingly attractive.

Binary vectors and *Agrobacterium*-mediated transformation have contributed to the considerable successes that plant biotechnology has already achieved, and will no doubt continue to contribute well into the future. In the chapters that follow, we will look at specific examples of plant biotechnology in practice, including ones that have relied on the use of *Agrobacterium*-mediated transformation and binary vectors.

FURTHER READING

Visit www.oxfordtextbooks.co.uk/orc/slaterplants2e/ for hyperlinked references that take you directly to online abstracts for the journal articles listed below:

Vector construction

Croy, R. R. D. (1993) Plant selectable genes, reporter genes and promoters. In *Plant Molecular Biology Labfax* (R. R. D. Croy, ed.), pp. 149–182. BIOS, Oxford.

Early, K., Haag, J. R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C. S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *The Plant Journal* **45**, 616–629.

Guerineau, F. and Mullineaux, P. (1993) Plant transformation and expression vectors. In *Plant Molecular Biology Labfax* (R. R. D. Croy, ed.), pp. 121–147. BIOS, Oxford.

Hellens, R., Mullineaux, P., and Klee, H. (2000) A guide to *Agrobacterium* binary Ti vectors. *Trends in Plant Science* **5**, 446–451.

Hull, G. A. and Devic, M. (1995) The β-glucuronidase (*gus*) reporter gene system. In *Methods in Molecular Biology* (H. Jones, ed.), pp. 125–141. Humana Press, Totowa.

Jefferson, R. A., Kavanagh, T. A., and Bevan, M. A. (1987) GUS fusions: β-glucuronidase as a versatile gene fusion marker in higher plants. *EMBO Journal* **6**, 3901–3907.

Karimi, M., Inze, D., and Depicker, A. (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends in Plant Science* **7**, 193–195.

Karimi, M., De Meyer, B., and Hilson, P. (2005) Modular cloning in plant cells. *Trends in Plant Science* **10**, 103–105.

Selectable markers

Erikson, O., Hertzberg, M., and Nasholm, T. (2005) The *dsdA* gene from *Escherichia coli* provides a novel selectable marker for plant transformation. *Plant Molecular Biology* **57**, 425–433.

Haldrup, A., Noerremark, M., and Okkels, F. T. (2001) Plant selection principle based on xylose isomerase. *In vitro Cellular and Developmental Biology—Plant* **37**, 114–119.

Haselhoff, J., Siemering, K. R., Prasher, D. C., and Hodge, S. (1997) Removal of a cryptic intron and subcellular localisation of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. *Proceedings of the National Academy of Sciences USA* **94**, 2122–2127.

Joersbo, M. (2001) Advances in the selection of transgenic plants using non-antibiotic marker genes. *Physiologia Plantarum* **111**, 269–271.

Kaeppler, H. F., Carlson, A. R., and Menon, G. K. (2001) Routine utilization of green fluorescent protein as a visual selectable marker for cereal transformation. *In vitro Cellular and Developmental Biology—Plant* **37**, 120–126.

LaFayette, R. R., Kane, P. M., Phan, B. H., and Parrott, W. A. (2005) Arabitol dehydrogenase as a selectable marker for rice. *Plant Cell Reports* **24**, 596–602.

Leyman, B., Avonce, N., Ramon, M., Van Dijck, P., Iturriaga, G., and Thevelein, J. M. (2006) Trehalose-6-phosphate synthase as an intrinsic selection marker for plant transformation. *Journal of Biotechnology* **121**, 309–317.

Luo, K., Zheng, X., Chen, Y., Xiao, Y., Zhoa, D., McAvoy, R., Pei, Y., and Li, Y. (2006) The maize *knotted1* gene is an effective positive selectable marker gene for *Agrobacterium*-mediated tobacco transformation. *Plant Cell Reports* **25**, 403–409.

Penna, S., Sagi, L., and Swennen, R. (2002) Positive selectable marker genes for routine plant transformation. *In vitro Cellular and Developmental Biology—Plant* **38**, 125–128.

Reed, J., Privalle, L., Powell, M. L., Meghji, M., Dawson, J., Dunder, E., Suttie, J., Wenck, A., Launis, K., Kramer, C., Chang, Y.-F., Hansen, G., and Wright, M. (2001) Phosphomannose isomerase: an efficient selectable marker for plant transformation. *In vitro Cellular and Developmental Biology—Plant* **37**, 127–132.

Inducible gene expression

Gurr, S. J. and Rushton, P. J. (2005) Engineering plants with increased disease resistance: how are we going to express it? *Trends in Biotechnology* **23**, 283–290.

Kong, H.-G., Fang, Y., and Singh, K. B. (1999) A glucocorticoid-inducible transcription system causes severe growth defects in *Arabidopsis* and induces defence-related genes. *Plant Journal* **20**, 127–133.

Malnoy, M., Reynoird, J., Borejsza-Wysocka, E., and Aldwinckle, H. (2006) Activation of the pathogen-inducible *Gst1* promoter of potato after elicitation by *Venturia inaequalis* and *Erwinia amylovora* in transgenic apple (*Malus* x *domestica*). *Transgenic Research* **15**, 83–93.

Martinez, A., Sparks, C., Hart, C. A., Thompson, J., and Jepson, I. (1999) Ecdysone agonist inducible transcription in transgenic tobacco plants. *Plant Journal* **19**, 97–106.

Moore, I., Samalova, M., and Kurup, S. (2006) Transactivated and chemically inducible gene expression in plants. *The Plant Journal* **45**, 651–683.

Padidam, M. (2003) Chemically regulated gene expression in plants. *Current Opinion in Plant Biology* **6**, 169–177.

Rushton, P. J., Reinstädler, A., Lipka, V., Lippok, B., and Somssich, I. E. (2002) Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen- and wound-induced signaling. *The Plant Cell* **14**, 749–762.

Vreugdenhill, D., Claassens, M. M. J., Verhees, J., van der Krol, A. R, and van der Plas L. H. W. (2005) Ethanol-inducible gene expression: non-transformed plants also respond to ethanol. *Trends in Plant Science* **11**, 9–11.

Zuo, J. and Chua, N.-H. (2000) Chemical-inducible systems for regulated expression of plant genes. *Current Opinion in Biotechnology* **11**, 146–151.

Transgene integration and expression

Butaye, K. M. J., Goderis, I. J. W. M., Wouters, P. F. J., Pues, J.-T. G., Delaure, S. L., Broekaert, W. F., Depicker, A., Cammue, B. P. A., and De Bolle, M. F. C. (2004) Stable high-level transgene expression in *Arabidopsis thaliana* using gene silencing mutants and matrix attachment regions. *The Plant Journal* **39**, 440–449.

Fang, R.-X., Nagy, F., Sivasubramaniam, S., and Chua, N.-H. (1989) Multiple *cis* regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. *Plant Cell* **1**, 141–150.

Gallie, D. R. (1998) Controlling gene expression in transgenics. *Current Opinion in Plant Biology* **1**, 166–172.

Kumar, S. and Fladung, M. (2001) Controlling transgene integration in plants. *Trends in Plant Science* **6**, 155–159.

Miki, B. (2002) Transgene expression and control. *In vitro Cellular and Developmental Biology—Plant* **38**, 139–145.

Perlak, F. J., Fuchs, R. L., Dean, D. A., McPherson, S. L., and Fischoff, D. A. (1991) Modification of the coding sequence enhances plant expression of insect control protein genes. *Proceedings of the National Academy of Sciences USA* **88**, 3324–3328.

RNAi

Byzova, M., Verduyn, C., De Brouwer, D., and De Block, M. (2004) Transforming petals in sepaloid organs in *Arabidopsis* and oilseed rape: implementation of the hairpin RNA-mediated gene silencing technology in an organ-specific manner. *Planta* **218**, 379–387.

Chuang, C.-F. and Meyerowitz, E. M. (2000) Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences USA* **97**, 4985–4990.

Dugas, D. V. and Bartel, B. (2004) MicroRNA regulation of gene expression in plants. *Current Opinion in Plant Biology* **7**, 512–520.

Guo, H.-S., Fei, J.-F., Xie, Q., and Chua, N.-H. (2003) A chemical-regulated inducible RNAi system in plants. *The Plant Journal* **34**, 383–392.

Kusaba, M. (2004) RNA interference in crop plants. *Current Opinion in Plant Biotechnology* **15**, 139–143.

Miki, D., Itoh, R., and Shimamoto, K. (2005) RNA silencing of single and multiple members in a gene family of rice. *Plant Physiology* **138**, 1903–1913.

Clean-gene technology

Ebinuma, H. and Komamoine, A. (2001) MAT (multi-auto-transformation) vector system. The oncogenes of *Agrobacterium* as positive markers for regeneration and selection of marker free transgenic plants. *In vitro Cellular and Developmental Biology—Plant* **37**, 103–113.

Jia, H., Pang, Y., Chen, X., and Fang, R. (2006) Removal of the selectable marker gene from transgenic tobacco plants by expression of Cre recombinase from a tobacco mosaic virus vector through agroinfection. *Transgenic Research* **15**, 375–384.

McCormac, A. C., Elliott, M. C., and Chen, D. F. (1999) pBECKS2000: a novel plasmid series for the facile creation of complex binary vectors which incorporates 'clean-gene' facilities. *Molecular and General Genetics* **261**, 226–235.

McCormac, A. C., Fowler, M. R., Chen, D. F., and Elliott, M. C. (2001) Efficient co-transformation of *Nicotiana tabacum* by two independent T-DNAs, the effect of T-DNA size and implications for genetic separation. *Transgenic Research* **10**, 143–155.

Puchta, H. (2000) Removing selectable marker genes: taking the short cut. *Trends in Plant Science* **5**, 273–274.

Wang, Y., Chen, B., Hu, Y., Li, J., and Lin, Z. (2005) Inducible excision of selectable marker gene from transgenic plants by the Cre/*lox* site-specific recombination system. *Transgenic Research* **14**, 605–614.

Web-links

Best practice in vector design (Advisory Committee on Releases to the Environment (ACRE) report), web-link 4.1: www.defra.gov.uk/environment/acre/bestprac/consult/guidance/bp/index.htm

Gateway™ vectors, Mark Curtis website, web-link 4.2: www.unizh.ch/botinst/Devo_Website/curtisvector/index_2.html

Gateway™ vectors, Plant Systems Biology, University of Ghent, web-link 4.3: www.psb.rug.ac.be/gateway/

pGreen and pBIN19, web-link 4.4: www.pgreen.ac.uk/a_hom.htm

The Royal Society, web-link 4.5: www.royalsoc.ac.uk/