3

One Gene is Not Enough: Transgene Detection, Expression and Control

K.D. Wells and R.J. Wall

Gene Evaluation and Mapping Laboratory, USDA-ARS, Beltsville, Maryland, USA

Introduction

It is not surprising that recently reported successes with sheep and cattle nuclear transfer have rekindled interest in examining new methodologies for producing transgenic animals. The proponents of nuclear transfer and other new approaches in part justify their efforts by pointing to the inefficiency of producing transgenic livestock. The techniques for producing transgenic livestock have remained basically unchanged since the first transgenic farm animals were reported in 1985 (Hammer *et al*., 1985). A review of the transgenic livestock literature suggests that the inefficiency of producing transgenic animals, by pronuclear microinjection, can be attributed to poor embryo survival, poor transgene integration rates and unpredictable transgene behaviour (Wall, 1996). At a minimum, performing nuclear transfer with genetically engineered nuclei will eliminate the inefficiency related to poor transgene integration rates. The impact of nuclear transfer on embryo/fetal survival and on the ability to select integrations with predictable transgene expression is still unknown. Therefore, until the new approaches are validated and are shown to increase the overall efficiency, we believe it is prudent to review some of the proposed strategies for improving the efficiency of current technology. Furthermore, no matter which way genes are introduced, their behaviour will be dictated by still ill-defined transgene design criteria. Though we are unable to offer a resolution to the vagaries of transgene design, it is our hope that some of the molecular approaches discussed will stimulate readers to re-evaluate their transgene design guidelines.

The title of this chapter was chosen in recognition of that fact that from the very beginning of transgenic animal technology, transgenes (fusion

 Transgenic Animals in Agriculture
(eds J.D. Murray, G.B. Anderson, A.M. Oberbauer and M.M. McGloughlin)

genes) were generated from sequences borrowed from more than one gene. Also, many of the more promising techniques proposed to improve efficiency rely on using two genes: a selectable reporter gene and a gene of interest. Furthermore, the strategies for more precisely controlling transgene behaviour, discussed in this review, utilize several pieces of DNA from a variety of sources.

Identification of Transgenic Embryos as a Means of Improving Efficiency

No validated tool has been developed to precisely measure the frequency at which microinjected transgenes integrate into mammalian genomes. In the absence of a direct measure, we have been forced to estimate that parameter based on the proportion of transgenic offspring born. Published estimates of the percentage of transgenics born range from about 4 to 30%. The true integration frequency may be higher if one assumes that a greater proportion of transgenic embryos or fetuses is lost during development than their non-transgenic counterparts (Canseco *et al*., 1994). Even though these estimates suggest that transgene integration is not a particularly rare event, maintaining an adequate recipient population (especially for monotocous species) is a costly proposition. In an attempt to reduce recipient costs a significant effort has been placed on identification of transgenic embryos before they are transferred to recipients. Primarily, two tactics have been employed, one based on direct detection of the transgene and the other based on detection of transgene expression. Unfortunately, there are rational theoretical grounds to question the wisdom of most approaches tested to date. A third approach, evaluating the methylation status of transgenes has received less attention.

Direct transgene detection

Direct transgene detection methods are, by necessity, PCR-based (there is simply not enough DNA in an embryo, or part thereof, to employ other techniques). The most serious drawback of PCR in this context is its inability to distinguish between transgenes that have integrated into the genome and copies of the transgenes that have not. The lack of distinction is problematic because it appears that unintegrated DNA can persist in the developing embryo for a significant time (Burdon and Wall, 1992; Cousens *et al*., 1994). Modified PCR strategies that are dependent on ligation are capable of distinguishing between integrated and unintegrated transgenes (Ninomiya *et al*., 1990; MacGregor and Overbeek, 1991; Jones and Winistorfer, 1992). Unfortunately, such techniques cannot be applied to this problem because the amount of DNA available in a few cells is insufficient.

However, it has been shown that following microinjection, transgenes frequently segregate during the early stages of development, resulting in a mosaic distribution. The mosaic nature of microinjected embryos has been demonstrated by indirectly monitoring transgene expression (Hattman *et al.*, 1978) and more directly by detection of transgenes in individual blastomeres (Burdon and Wall, 1992; Cousens *et al.*, 1994). Due to this mosaicism, detection of a transgene in two or three cells harvested from a developing eight- to 16-cell embryo could easily generate a misleading result.

It would appear that the inadequacy of PCR for this application, coupled with the mosaic distribution of transgenes in the developing embryo, has thus far thwarted attempts to select transgenic embryos based on direct detection of transgenes. None of the published accounts have succeeded in increasing the proportion of transgenic offspring or have not conducted the necessary experiments to fully validate this approach (Ninomiya *et al.*, 1989; Horvat *et al.*, 1993; Bowen *et al.*, 1994; Sparks *et al.*, 1994; Page *et al.*, 1995; Hyttinen *et al.*, 1996; Seo *et al.*, 1997).

Detection of transgene expression

Techniques based on monitoring transgene expression in preimplantation embryos are also hindered by lingering unintegrated transgenes which are likely expressed as well as, if not more strongly than, integrated copies. The phenomenon of diminished expression levels of integrated transgenes has clearly been demonstrated in tissue culture where levels of expression of stably transfected (integrated genes) cells are usually much lower than those of transient (unintegrated genes) transfected cells. Therefore, an expression-based selection system requires the potentially interfering background expression of unintegrated DNA to be eliminated or rendered non-detectable at the time of screening. If a regulatory element could be found that 'turns on' a reporter gene after unintegrated DNA is degraded, or if a transgene could be designed that only expressed after integration, this selection approach might work. However, even if unintegrated DNA has been degraded or rendered non-functional, background expression might still be a problem unless the mRNA and protein generated by the unintegrated transgenes have relatively short half-lives.

A number of different strategies to monitor transgene expression in preimplantation embryos have been proposed. They all depend on two genes, a reporter (selectable gene) and the gene of interest. One could devise non-destructive methods for detecting expression in the whole embryo or analyse expression in blastomere biopsies. Wisely, the expression-monitoring experiments that have been published thus far are based on evaluating the whole embryo. Each of the three approaches that have been tested relies on monitoring the expression of a reporter gene, but methods of detection differ.

In one case, expression monitoring was based on detecting a secreted gene product, luciferase (Thompson *et al*., 1995). Surprisingly, the authors claimed that the expression of their reporter construct increased upon integration. This promising approach has not yet been used to demonstrate that selection based on high luciferase expression will yield a higher proportion of transgenic animals.

A selection system based on resistance to the toxic effects of the neomycin analogue G418, commonly used in tissue culture transfection studies, has been evaluated in mouse (Tada *et al*., 1995) and bovine (Bondioli and Wall, 1996) embryos. G418 blocks translation, but embryos are protected if they express the neomycin resistance gene (*neo*). In the mouse study, the authors reported a 48% survival rate to the blastocyst stage after microinjection with the *neo* gene and 2 days of culture in G418 at a concentration which otherwise would kill 100% of non-injected embryos. However, these authors did not present evidence to demonstrate that *neo* selection increased the proportion of transgenic mice born. In the bovine embryo study the *neo* gene was fused to the *LacZ* gene in a bicistronic construct. Since bicistronic constructs generate a single mRNA, it can be reasonably assumed that β-galactosidase-positive blastomeres (those expressing the *LacZ* gene) also were producing the neomycin resistance protein. Staining for LacZ after 8 days of culture to the expanded blastocyst stage revealed that 15% of microinjected, but unselected, embryos were LacZ-positive (contained 80% or more β-galactosidase-positive blastomeres). Of the microinjected embryos treated with G418, 60% were LacZ-positive. It remains to be seen if the enrichment of expression-positive embryos translates into more transgenic calves. However, preliminary results using the same construct in mouse embryos suggest that G418 compromises development beyond the blastocyst stage (K.R. Bondioli, K.D. Wells and R.J. Wall, unpublished data).

Possibly the most obvious and straightforward approach for identifying transgenic embryos based on expression of a transgene, was the utilization of green fluorescent protein (GFP) as the reporter (Takada *et al*., 1997). In that very encouraging study an elongation factor promoter ligated to GFP was injected into mouse embryos and green fluorescence was monitored at the morula/blastocyst stage. Eight out of 12 fetuses or pups resulting from green fluorescing embryos were transgenic as assessed by Southern blot analysis. The authors also reported that GFP could be detected in bovine embryos injected with the same construct. This is the first report that we are aware of that clearly demonstrates an increase in the proportion of transgenic animals based on a selection system. This approach warrants further investigation.

Even if GFP selection or another method can be found to enhance the proportion of transgenic animals born, the use of a selectable reporter gene could introduce some problems. Expression of the reporter could compromise embryo viability, thus reducing pregnancy rates. It is also

possible that expression of the reporter transgene might interfere with the function of the co-selected transgene of interest. There is precedence for one transgene influencing expression of another following co-injection. John Clark and colleagues have shown that constructs that are expressed poorly can be rescued by co-injecting them with the ovine β-lactoglobulin gene (Clark *et al.*, 1992). However, they have also demonstrated that co-injected gene constructs can silence or diminish expression of otherwise highly expressed sequences (Clark *et al.*, 1997). The implication of those observations is that it might be prudent to isolate the reporter construct from the co-selected transgene by using DNA sequences such as matrix attachment regions (MAR) or scaffold attachment regions (SAR) which will be discussed later.

A selection concept that has yet to be tested in transgenic animals is based on designing a transgene that can only be expressed after integration. It may be possible to take advantage of the unique characteristics of peptide nucleic acids (PNA) to achieve this goal (Nielsen *et al.*, 1991; Almarsson *et al.*, 1993; Betts *et al.*, 1995; Norton *et al.*, 1995; Corey, 1997). The PNA molecule has a peptide backbone with purines and pyrimidines as the functional side groups (Nielsen and Haaima, 1997). The peptide backbone maintains the same distance between bases as nucleic acids but is uncharged. In the absence of the charged phosphodiester backbone, PNA annealed to DNA through normal base pairing is much more stable than other nucleic acids hybridized in the same fashion (Egholm *et al.*, 1993). The PNA/DNA interaction is sufficiently strong to prevent polymerases from displacing the PNA. The annealed PNA can therefore block progression of transcription. In theory, this ability to block expression could be used in microinjected embryos to silence transient expression of the transgene. Upon replication (assumed for this argument to be integration-dependent), the PNA/DNA hybrid is removed by the editing machinery of the cell. Once this block to expression is removed, an appropriately designed transgene can be detected by its expression.

Detection of methylation status

One approach to selection that does not neatly fit in either the direct detection or the expression monitoring paradigms is based on changes in the methylation state of specific bases in transgenes. Methylation at the N^6 of the adenine within the nucleotide sequence GATC is a characteristic of *dam*$^+$ *Escherichia coli*. Mammalian cells do not normally methylate this adenine and consequently the absence of methylation following introduction of DNA of bacterial origin (transgenes isolated from plasmids) into mammalian cells is taken to indicate that the DNA has been replicated within the host cell (Hattman *et al.*, 1978; Peden *et al.*, 1980). Since replication would be expected to occur following integration in dividing

cells, the absence of methylation would be indicative of transgene integration. The recognition site of the restriction enzyme *Dpn*I includes an adenine, but *Dpn*I only cuts if the adenine within the recognition site is methylated. Attempts to take advantage of the changing state of methylation pattern to monitor integration of transgenes in embryos have failed (Burdon and Wall, 1992; Cousens *et al*., 1994). The failure might result from some specific inhibitory substance in embryos that blocks the action of *Dpn*I, modification of the methylation state of the *Dpn*I site independent of replication or possibly because transgenes are replicated before they are integrated (not a commonly accepted hypothesis).

Transgene Behaviour

Several strategies have been employed to regulate transgene expression patterns. Initial attempts utilized promoter elements from cloned mammalian genes to control tissue specificity and temporal expression. This strategy has been invaluable for targeting expression to particular tissues. However, it has proven to be of little use for temporal control due to limited availability of inducible promoters with low basal activity and the common occurrence of ectopic expression. To overcome this difficulty, strategies have been developed that utilize ligand-dependent sequence-specific DNA binding proteins to initiate or perturb transcription. The 'switches' thus far designed require a minimum of two genes. One gene encodes the 'switch protein' and the second gene responds to the switch protein (Fig. 3.1). Generally the switches fall into three categories. Those that inhibit expression, those that induce expression and those that utilize a 'double switch.'

Inhibition of basal expression

To inhibit expression of a transgene that would otherwise be expressed at some basal level, a strategy has been devised which is based on conditional binding of a switch protein to a specific DNA sequence within the transgene. The binding of the switch protein interferes with initiation or elongation of transcription. For example, transgenes can be designed that allow conditional binding of the *E. coli lac* repressor. The repressor binding sequences are placed in untranslated regions of the transgene and interfere with the production of functional transgene mRNA (Fig. 3.2a; Biard *et al*., 1992). Similarly, the coding region of a silencing domain from a transcription factor such as KRAB (Margolin *et al*., 1994) can be fused to the coding region of the DNA binding domain of *E. coli tet* repressor to produce a synthetic transcription factor which silences expression when bound to the transgene promoter (Fig. 3.2b; Deuschle *et al*., 1995). In either of these cases, to provide complete repression, expression of the switch protein must be in every cell

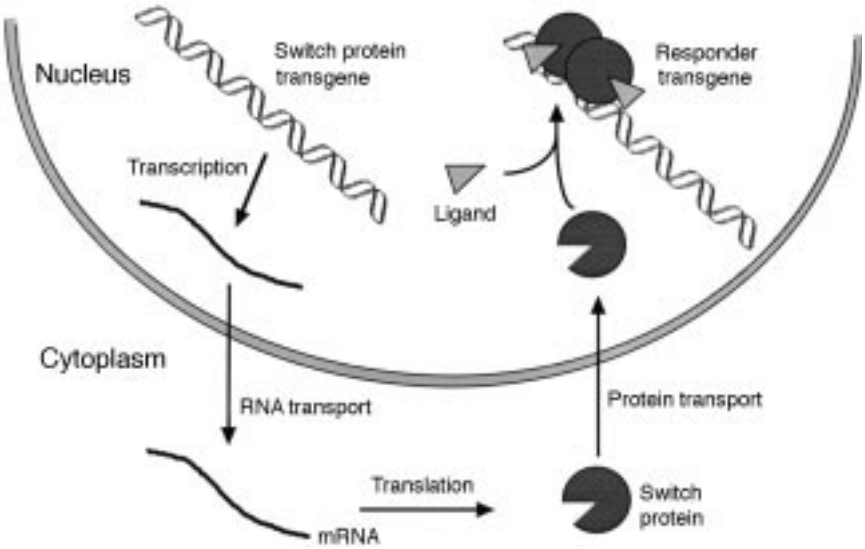

Fig. 3.1. A typical two-gene transgene switch. The transgene encoding the switch protein is controlled by a tissue-specific promoter sequence as necessary. Upon transcription of the switch protein gene and subsequent translation of its mRNA, the switch protein enters the nucleus. In this example, a ligand must interact with the switch protein for the protein switch to bind to its target sequence on the responder transgene. Supplying and withdrawing ligand affords temporal control. Switch systems can be designed that are ligand independent. They can be controlled by developmentally regulated promoters.

that would otherwise express the transgene. Expression would be induced by administration of a ligand (i.e. galactose or tetracycline analogues) which prevents sequence-specific binding of the switch protein and therefore allows the transgene to be expressed. One could predict that these strategies would have limited success in animals due to different ectopic expression patterns between the two transgenes. Switch protein would be required to express not only in the tissue targeted for transgene control but also in every cell that may otherwise express the gene of interest ectopically.

This type of switch may be useful in any situation that requires external control of transgene expression. For example, if one wanted to express a circulating hormone, such as somatotropin, and a dependable promoter with no ectopic expression was available, then that promoter could be used to drive expression of both the switch protein and somatotropin. The promoter driving expression of somatotropin would also include a binding site for the switch protein. When the switch protein was bound to the somatotropin transgene promoter, then only endogenous somatotropin would be produced. In breeding animals or very young animals this state would be desirable. When the switch protein was not bound to the somatotropin transgene promoter, the

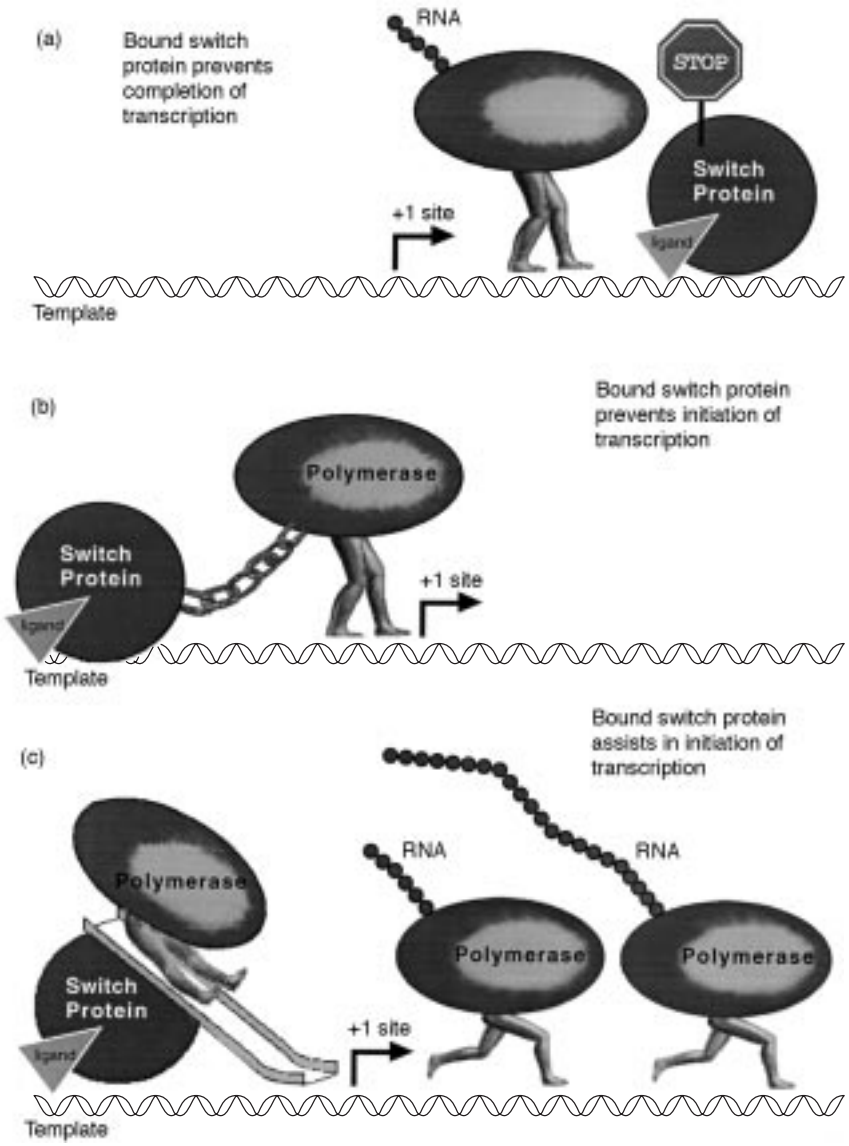

Fig. 3.2. Three ligand-dependent ways to control gene expression. (a) Ligand-dependent DNA binding proteins can perturb progression of RNA polymerase to prevent functional mRNA from being produced. In the presence of bound 'switch protein', the transgene is down-regulated. In the absence of bound switch protein, the transgene would be regulated by the transgene promoter. (b) Ligand-dependent DNA binding proteins can be fused to repressor domains to prevent initiation of transcription. The repressor domain interacts with a series of co-repressors to induce

peptide hormone would be made and secreted into the circulation. This state would be desirable during the growth phase of meat production animals.

Activation of induced expression

A second strategy attempts to mimic mammalian transcription factors to induce expression from an otherwise silent promoter. Two genes and an inducing molecule are required. One gene encodes the switch protein and the other responds when it binds the switch protein. The switch protein in these cases bind unique DNA sequences located upstream of a minimal promoter (Fig. 3.2c). The minimal promoter is typically composed of little more than a TATA box and the switch protein binding sequence. Minimal promoters can support initiation of transcription but require the switch protein to be bound to promote elevated expression. These switch proteins are again ligand-dependent conditional binding proteins but now include transactivation domains. In the first scenario the intention was to turn off a gene, whereas this strategy is designed to turn genes on. For example, a mutant *E. coli tet* repressor that binds a specific DNA sequence only in the presence of a tetracycline analogue, doxycycline, can be fused to the transactivation domain of herpes VP16 to make a protein referred to as rtTA (Gossen *et al.*, 1995). In the absence of doxycycline, the switch protein, rtTA, does not specifically bind to the responder transgene minimal promoter and the transgene is silent. Upon administration of doxycycline, the switch protein binds to the DNA included in the minimal promoter of the responder transgene and transgene expression is induced.

Similarly, transcription factors from other eukaryotes such as the *Drosophila melanogaster* ecdysone receptor can be used as the switch protein. The ecdysone receptor normally functions as a heterodimer with the

---

presumptive conformational changes represented here as a chain. In the presence of bound switch protein and endogenous co-repressor protein, the transgene is down-regulated. In the absence of the bound switch protein, the transgene is regulated by the transgene promoter. (c) Ligand-dependent DNA binding proteins can be fused to transactivation domains to enhance initiation of transcription. The transactivation domain interacts with a series of co-activators to stimulate assembly and initiation shown here as a slide. In transgenes with no other enhancers, the minimal promoter is insufficient to drive expression. In the presence of bound switch protein, the transgene is induced to express from a minimal promoter.

gene product ultraspiracle (Yao *et al.*, 1992). Fortunately, the mammalian homologue, retenoic acid receptor X (RXR), can serve in the place of ultraspiracle to heterodimerize with the ecdysone receptor to produce a functional transcription factor (Thomas *et al.*, 1993). Thus, the ecdysone receptor can function as a switch protein in cells that express RXR (Christopherson *et al.*, 1992). A mutant progesterone receptor has also been used as a switch protein (Wang *et al.*, 1997). This mutant does not induce expression in the presence or absence of endogenous steroids but is active upon administration of very low levels of the progesterone antagonist RU486. To prevent induction of endogenous progesterone responsive genes by the switch protein, the native DNA binding domain was replaced with the yeast Gal4 DNA binding domain (Wang *et al.*, 1994). Similar modifications have also been made to the ecdysone receptor (No *et al.*, 1996).

This type of switch may be useful in same scenarios as the first strategy. However, when low background is a major concern this switch is less dependent on a predictable promoter for the switch protein. Instead, this type of switch is most dependent on the construction of a responder transgene that has very low background expression. The level of induction is dependent on both the promoter of the switch protein and the inducibilty of an otherwise silent responder gene.

Cross-breed for expression

The third category of switches does not perturb or mimic mammalian transcription. Instead, this strategy finds an alternative method for producing an active gene. It still requires a switch protein gene in addition to the transgene of interest. One example would be to use a phage RNA polymerase to drive expression of a transgene (Lieber *et al.*, 1989). In this strategy, the gene of interest need not contain any mammalian promoter element thus ensuring little or no basal expression. In the presence of the phage RNA polymerase the gene is active. In the absence of the phage RNA polymerase the gene is inactive. The tissue specificity is provided by the promoter used to drive the phage RNA polymerase expression. The timing of the expression is also provided by the promoter used to drive polymerase expression. Therefore, a second switching mechanism is required to control the polymerase (cross-breeding serves as the second 'switch' in this example). The use of a double switch has been demonstrated to provide tighter control (Aubrecht *et al.*, 1996) but this binary strategy is primarily applicable to situations where transgene expression needs to be regulated in a particular generation and requires the production of two types of transgenic animals. For example, one might want to express a transgene that could prevent sexual maturation or gonadal development as an alternative to castration. If an animal carried both the phage RNA polymerase gene and the responder gene, those individuals could not reproduce. However, if breeding stock were

developed so that one line carried the responsive gene of interest and another line harboured the switch protein gene, then only the offspring generated by crossing those lines would have this phenotype. Similar arguments could be made for transgenic fish farms to prevent a phenotype from being transferred to wild populations even if the transgenes themselves are transferred to wild populations. Like the phage RNA polymerase, recombinases which remove inhibitory sequence or invert a transgene coding sequence, can be used to control gene expression. Some success using both the yeast FLP recombinase and the bacterial Cre recombinase has been demonstrated in tissue culture and transgenic animals (O'Gorman *et al*., 1991; Barlow *et al*., 1997).

The strategy utilizing synthetic transcription factors as the switch protein will continue to be improved. Already the tetracycline repressor switch has incorporated DNA binding mutants such that the wild-type version can be induced by ligand withdrawal and the mutant version by ligand administration. In many livestock scenarios breeding strategies required to use the binary approach are impractical. However, for livestock breeding companies that want to control proprietary genetics, a binary system may provide a means of protecting their unique genetic stock. For other situations where efficient transmission of transgenes is important, the genes of the binary system may be combined into a single gene construct incorporating both coding regions in the same construct (Schultze *et al*., 1996). However, these systems are greatly affected by integration position. Integration sites that provide adequate transgene induction tend to have greater basal expression. If the effect of integration site can be removed or reduced, the strategy may become very useful.

Integration Site Effects

In most studies about half of the transgenic founder animals do not express their transgene. The efficiency of producing transgenic animals could be doubled if a way could be found to correct that problem. Most transgenes are composite sequences originating from a variety of genes, species or even kingdoms. The regulatory elements are often less than fully characterized and the structural gene can contain sequences that interfere with the intended regulation. In addition, the sequences surrounding the integration may be as varied as the genome itself. The surround sequence may therefore also interfere with the intended regulation of the transgene. Several strategies have been used to overcome some of these problems.

Insulators

The location in which a transgene integrates can contain enhancers, silencers or other elements that may interact with the transgene promoter. Transgene expression therefore can be greatly affected by site of integration. In

D. melanogaster, sequences have been identified on the basis of their ability to functionally separate enhancers from their corresponding promoters (Cal and Levine, 1995; Gdula and Corces, 1997). These regions therefore insulate the sequence on one side of the insulating element from effectors located on the other side of the element. One example of a *D. melanogaster* insulator is located on a P-element, *Gypsy* (Roseman *et al*., 1995). The *Gypsy* insulator has been demonstrated to reduce the effect of integration site on transgenes. A vertebrate sequence has been shown to be functionally similar to *Gypsy* by insulating transgenes from position effects in transgenic flies (Chung *et al*., 1993). This vertebrate sequence was characterized by its location as the boundary of the DNase hypersensitive region of the expressing locus of chicken lysozyme gene (Stief *et al*., 1989). This boundary element was further demonstrated to have sequences which bind to the nuclear matrix (MAR) and are thought to define the border of the lysozyme locus. In transgenic mice, this sequence was also shown to insulate heterologous transgenes from position effects, to allow more appropriate developmental timing of expression, and to reduce ectopic expression (McKnight *et al*., 1992, 1996; Greenberg *et al*., 1994). Sequences from other genes have been described as locus control regions (LCRs) (Stamatoyannopoulos *et al*., 1997), as chromosomal SARs (Allen *et al*., 1996) and MARs. At some level these DNA fragments have related functions since they all can decrease position effects. However, these sequences are not identical nor are the effects of these regions universal among different transgenes (Barash *et al*., 1996a).

MARs, as mentioned above, have sequences that can bind the nuclear matrix and/or chromosomal scaffold. As such, these sequences are thought to anchor chromatin loops to the central organizing matrix of each chromosome. This function alone cannot explain the insulating ability of MARs. This function however is associated with maintenance of an open chromatin region once a gene is transactivated (Ciejek *et al*., 1983; Tewari *et al*., 1996). Regions within some MARs may assist in transactivator access to promoters. The chicken lysozyme MAR does contain intrinsically bent DNA – a structure associated with transactivator accessibility (Phi-Van and Stratling, 1996).

An insulating function such as that found in the globin LCR or the *Gypsy* insulator may not be associated with matrix binding yet can interact with the basal transcription machinery to prevent interactions with distal enhancers. Such a function has been implicated in the endogenous regulation of globin genes to prevent fetal globin enhancers from activating adult globin promoters (Boulikas, 1994). Given the functional similarity between *Gypsy* and the chicken lysozyme MAR, sequences within or near some MARs may be binding sites for insulator proteins such as those described for *Gypsy*, suppressor of *Hairy-wing* (Smith and Corces, 1992).

The mechanism of protection from integration position effects is unknown. In many cases MAR-like sequences may provide some protection from position effects. Providing locus boundaries, scaffold attachment, topoisomerase substrates or direct interaction between enhancers and the

basal transcription machinery may be involved in the action of MARs (Boulikas, 1995). The use of MARs in specific scenarios has proven to be less than predictable (Barash *et al*., 1996a). However, as another tool in the arsenal of DNA elements available for transgene design, MARs will prove to be very useful in many cases.

Rescue

In the above discussions, the availability of tissue-specific promoters has been assumed to be present. However, very few loci have been sufficiently described as to assure that a particular promoter region contains sufficient elements to provide the predicted expression pattern. As the globin locus has illustrated, regulatory elements can be located at distances impractical for inclusion in most plasmids (Slightom *et al*., 1997). Regulatory elements have been described upstream and downstream of transcriptional start sites, within coding regions and within introns. When spatial relationships between regulatory elements are important it may not be possible to create a transgene which maintains those relationships. As an alternative, non-manipulated configurations have been coinjected with transgene constructs to increase the fidelity of regulation (Barash *et al*., 1996b). This strategy has been called 'rescue' and has, in fact, rescued correct expression patterns from a transgene which otherwise does not express as predicted. Again, the mechanism and universal application is unknown. However, in the absence of full understanding of the elements required for correct expression, this strategy allows a spatially correct pattern of regulatory elements to be included within a transgene locus.

Native regulation

As an alternative to transgene rescue, inclusion of transgenes with native context offers the potential to generate spatially correct transgenes with all or most regulatory elements for the endogenous gene. One such strategy is the insertion, by homologous recombination, of the coding region of interest within the native gene whose expression pattern meets all requirements. This strategy is similar to the proven technologies of enhancer trapping and promoter trapping but relies on precise placement of the transcoding region within an endogenous locus. If the disturbed allele is required for some other function, this strategy may create unanticipated problems. These unknowns may be circumvented by including an internal ribosome entry site for the transcoding region or the endogenous coding region. One still risks perturbing expression at the targeted allele by changing spatial relationships or by inserting within important sequences. Likewise the introduction of regulatory elements within the transcoding region or its introns may also

create an unpredicted expression pattern. However unproven, this strategy remains potentially useful.

When the transgene needs to retain its endogenous expression pattern, very large fragments of DNA such as those found in yeast or bacterial artificial chromosomes (YACs or BACs, respectively) may prove useful (Peterson *et al.*, 1996). By using such large regions of the genome, inclusion of the entire locus is more likely. However, this strategy will generally prevent known sequence from being integrated unless the entire artificial chromosome has been sequenced. The lack of complete sequence information would be of serious concern for commercial projects in the US because the Federal Drug Agency currently requires such information.

Codon bias

As suggested above, transgene coding regions may include regulatory elements that generate unpredicted expression patterns when combined with well described promoters. The unexpected pattern of expression may be no detectable expression at all. Regulatory elements have been described within both coding regions and introns. However, since prokaryotes have had no genetic selection against sequence which have functions in eukaryotes, it is inevitable that silencers, enhancers, splice sites and degradation signals will be present in some genes of bacterial origin. Such sequences have been described (Artelt *et al.*, 1991; Kushner *et al.*, 1994). In addition, the base composition of DNA between divergent species is sometimes very different. Those differences are reflected in the codons that are favoured by particular species. Some tRNAs are relatively rare in species that have a codon bias and the codon for those tRNAs are also rare. The extreme sequence differences that cause codon bias between species may cause inefficiencies in translation, but those differences may also represent sequences that have been selected against in the target organism. For example, attempts to express an insect resistance gene (*Bacillus thuringiensis*, cryIA) in transgenic plants failed until the coding region was chemically synthesized to generate a coding region that statistically represented the DNA sequence most likely to encode the desired amino acid sequence (Perlak *et al.*, 1991). By altering the codon bias, three poly(A) cleavage signals were removed. The resulting coding region, now codon biased for the host plant, expressed at very high levels. This strategy remains to be proven in transgenic vertebrates but codon bias has been shown to be important in many other systems (Perlak *et al.*, 1993; Alexandrova *et al.*, 1995).

Conclusions

The search for alternative approaches for producing transgenic animals is well justified in light of the low efficiency of the current process which

results in high-cost scientific projects that are often beyond the capacity of granting agencies to fund. Recent developments in nuclear transfer promise to eliminate at least one of the low efficiency steps, namely poor integration rates. The heightened interest in nuclear transfer may also result in identifying new embryo culture systems that address another of the low efficiency steps, that of poor embryo/fetal survival. Nevertheless, it may not be time to abandon microinjection as a method for producing transgenic animals, especially in light of the encouraging results with at least one scheme for selecting transgenic embryos. If GFP expression can be used as effectively for selection of transgenic livestock embryos as it was in mice, producing 50–60% transgenic founders would make current technology more appealing.

Furthermore, new approaches to gene transfer are going to have little immediate impact on correcting transgene function. Even if genes can be readily introduced by homologous recombination, the transgenic mouse literature suggests that homologous recombination will not always be an appropriate strategy and will not be a panacea for achieving precise control of gene behaviour. Use of isolating elements and transgene switches hold the best promise for controlling the action of transgenes. Transgene switches have already been used to control formation and regression of carcinomas (Furth *et al*., 1994), control gene expression in embryonal carcinoma cells (Miller and Rizzino, 1995) and heart muscle (Yu *et al*., 1996) and in plants (Weinmann *et al*., 1994). It is probably time for these switches to be tested in transgenic livestock.

References

Alexandrova, R., Eweida, M., Georges, F., Dragulev, B., Abouhaidar, M.G. and Ivanov, I. (1995) Domains in human interferon α-1 gene containing tandems of arginine codons AGG play the role of translational initiators in *Escherichia coli*. *International Journal of Biochemistry and Cell Biology* 27, 469–473.

Allen, G.C., Hall, G.J., Michalowski, S., Newman, W., Spiker, S., Weissinger, A.K. and Thompson, W.F. (1996) High-level transgene expression in plant cells: effects of a strong scaffold attachment region from tobacco. *Plant Cell* 8, 899–913.

Almarsson, O., Bruice, T.C., Kerr, J. and Zuckermann, R.N. (1993) Molecular mechanics calculations of the structures of polyamide nucleic acid DNA duplexes and triple helical hybrids. *Proceedings of the National Academy of Sciences USA* 90, 7518–7522.

Artelt, P., Grannemann, R., Stocking, C., Friel, J., Bartsch, J. and Hauser, H. (1991) The prokaryotic neomycin-resistance-encoding gene acts as a transcriptional silencer in eukaryotic cells. *Gene* 99, 249–254.

Aubrecht, J., Manivasakam, P. and Schiestl, R.H. (1996) Controlled gene expression in mammalian cells via a regulatory cascade involving the tetracycline transactivator and lac repressor. *Gene* 172, 227–231.

Barash, I., Ilan, N., Kari, R., Hurwitz, D.R. and Shani, M. (1996a) Co-integration of β-lactoglobulin/human serum albumin hybrid genes with the entire β-lactoglobulin gene or the matrix attachment region element: repression of human serum albumin and β-lactoglobulin expression in the mammary gland and dual regulation of the transgenes. *Molecular Reproduction and Development* 45, 421–430.

Barash, I., Nathan, M., Kari, R., Ilan, N., Shani, M. and Hurwitz, D.R. (1996b) Elements within the β-lactoglobulin gene inhibit expression of human serum albumin cDNA and minigenes in transfected cells but rescue their expression in the mammary gland of transgenic mice. *Nucleic Acids Research* 24, 602–610.

Barlow, C., Schroeder, M., Lekstrom-Himes, J., Kylefjord, H., Deng, C.X., Wynshaw-Boris, A., Spiegelman, B.M. and Xanthopoulos, K.G. (1997) Targeted expression of *Cre* recombinase to adipose tissue of transgenic mice directs adipose-specific excision of LoxP-flanked gene segments. *Nucleic Acids Research* 25, 2543–2545.

Betts, L., Josey, J.A., Veal, J.M. and Jordan, S.R. (1995) A nucleic acid triple helix formed by a peptide nucleic acid–DNA complex. *Science* 270, 1838–1841.

Biard, D.S., James, M.R., Cordier, A. and Sarasin, A. (1992) Regulation of the *Escherichia coli lac* operon expressed in human cells. *Biochimica Biophysica Acta* 1130, 68–74.

Bondioli, K.R. and Wall, R.J. (1996) Positive selection of transgenic bovine embryos in culture. *Theriogenology* 45, 345.

Boulikas, T. (1994) Transcription factor binding sites in the matrix attachment region (MAR) of the chicken alpha-globulin gene. *Journal of Cellular Biochemistry* 55, 513–529.

Boulikas, T. (1995) Chromatin domains and prediction of MAR sequences. *International Review of Cytology* 162A, 279–288.

Bowen, R.A., Reed, M.L., Schnieke, A., Seidel, G.E., Jr, Stacey, A., Thomas, W.K. and Kajikawa, O. (1994) Transgenic cattle resulting from biopsied embryos: expression of *c-ski* in a transgenic calf. *Biology of Reproduction* 50, 664–668.

Burdon, T.G. and Wall, R.J. (1992) Fate of microinjected genes in preimplantation mouse embryos. *Molecular Reproduction and Development* 33, 436–442.

Cal, H. and Levine, M. (1995) Modulation of enhancer–promoter interactions by insulators in the *Drosophila* embryo. *Nature* 376, 533–536.

Canseco, R.S., Sparks, A.E., Page, R.L., Russell, C.G., Johnson, J.L., Velander, W.H., Pearson, R.E., Drohan, W.N. and Gwazdauskas, F.C. (1994) Gene transfer efficiency during gestation and the influence of co-transfer of non-manipulated embryos on production of transgenic mice. *Transgenic Research* 3, 20–25.

Christopherson, K.S., Mark, M.R., Bajaj, V. and Godowski, P.J. (1992) Ecdysteroid-dependent regulation of genes in mammalian cells by a *Drosophila* ecdysone receptor and chimeric transactivators. *Proceedings of the National Academy of Sciences USA* 89, 6314–6318.

Chung, J.H., Whiteley, M. and Felsenfeld, G. (1993) A 5′ element of the chicken *beta-globin* domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* 74, 505–514.

Ciejek, E.M., Tsai, M.J. and O'Malley, B.W. (1983) Actively transcribed genes are associated with the nuclear matrix. *Nature* 306, 607–609.

Clark, A.J., Cowper, A., Wallace, G., Wright, G. and Simons, J.P. (1992) Rescuing transgene expression by co-injection. *Bio/Technology* 10, 1450–1454.

Clark, A.J., Harold, G. and Yull, F.E. (1997) Mammalian cDNA and prokaryotic reporter sequences silence adjacent transgenes in transgenic mice. *Nucleic Acids Research* 25, 1009–1014.

Corey, D.R. (1997) Peptide nucleic acids: expanding the scope of nucleic acid recognition. *Trends in Biotechnology* 15, 224–229.

Cousens, C., Carver, A.S., Wilmut, I., Colman, A., Garner, I. and O'Neill, G.T. (1994) Use of PCR-based methods for selection of integrated transgenes in preimplantation embryos. *Molecular Reproduction and Development* 39, 384–391.

Deuschle, U., Meyer, W.K. and Thiesen, H.J. (1995) Tetracycline-reversible silencing of eukaryotic promoters. *Molecular and Cellular Biology* 15, 1907–1914.

Egholm, M., Buchardt, O., Christensen, L., Behrens, C., Freier, S.M., Driver, D.A., Berg, R.H., Kim, S.K., Norden, B. and Nielsen, P.E. (1993) PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules. *Nature* 365, 566–568.

Furth, P.A., St Onge, L., Boger, H., Gruss, P., Gossen, M., Kistner, A., Bujard, H. and Hennighausen, L. (1994) Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter. *Proceedings of the National Academy of Sciences USA* 91, 9302–9306.

Gdula, D.A. and Corces, V.G. (1997) Characterization of functional domains of the su(Hw) protein that mediate the silencing effect of mod(mdg4) mutations. *Genetics* 145, 153–161.

Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W. and Bujard, H. (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science* 268, 1766–1769.

Greenberg, N.M., DeMayo, F.J., Sheppard, P.C., Barrios, R., Lebovitz, R., Finegold, M., Angelopoulou, R., Dodd, J.G., Duckworth, M.L., Rosen, J.M. and Matusik, R.J. (1994) The rat *probasin* gene promoter directs hormonally and developmentally regulated expression of a heterologous gene specifically to the prostate in transgenic mice. *Molecular Endocrinology* 8, 230–239.

Hammer, R.E., Pursel, V.G., Rexroad, C.E. Jr, Wall, R.J., Bolt, D.J., Ebert, K.M., Palmiter, R.D. and Brinster, R.L. (1985) Production of transgenic rabbits, sheep and pigs by microinjection. *Nature* 315, 680–683.

Hattman, S., Brooks, J.E. and Masurekar, M. (1978) Sequence specificity of the P1 modification methylase (*M.Eco P1*) and the DNA methylase (*M.Eco dam*) controlled by *Escherichia coli* dam gene. *Journal of Molecular Biology* 126, 367–369.

Horvat, S., Medrano, J.F., Behboodi, E., Anderson, G.B. and Murray, J.D. (1993) Sexing and detection of gene construct in microinjected bovine blastocysts using the polymerase chain reaction. *Transgenic Research* 2, 134–140.

Hyttinen, J.M., Peura, T., Tolvanen, M., Aalto, J. and Jänne, J. (1996) Detection of microinjected genes in bovine preimplantation embryos with combined DNA digestion and polymerase chain reaction. *Molecular Reproduction and Development* 43, 150–157.

Jones, D.H. and Winistorfer, S.C. (1992) Sequence specific generation of a DNA panhandle permits PCR amplification of unknown flanking DNA. *Nucleic Acids Research* 3, 595–600.

Kushner, P.J., Baxter, J.D., Duncan, K.G., Lopez, G.N., Schaufele, F., Uht, R.M., Webb, P. and West, B.L. (1994) Eukaryotic regulatory elements lurking in plasmid DNA: the activator protein-1 site in pUC. *Molecular Endocrinology* 8, 405–407.

Lieber, A., Kiessling, U. and Strauss, M. (1989) High level gene expression in mammalian cells by a nuclear T7-phase RNA polymerase. *Nucleic Acids Research* 17, 8485–8493.

MacGregor, G.R. and Overbeek, P.A. (1991) Use of a simplified single-site PCR to facilitate cloning of genomic DNA sequences flanking a transgene integration site. *PCR Methods and Applications* 2, 129–135.

Margolin, J.F., Friedman, J.R., Meyer, W.K., Vissing, H., Thiesen, H.J. and Rauscher, F.J. (1994) Kruppel-associated boxes are potent transcriptional repression. *Proceedings of the National Academy of Sciences USA* 91, 4509–4513.

McKnight, R.A., Shamay, A., Sankaran, L., Wall, R.J. and Hennighausen, L. (1992) Matrix-attachment regions can impart position-independent regulation of a tissue-specific gene in transgenic mice. *Proceedings of the National Academy of Sciences USA* 89, 6943–6947.

McKnight, R.A., Spencer, M., Wall, R.J. and Hennighausen, L. (1996) Severe position effects imposed on a 1 kb mouse whey acidic protein gene promoter are overcome by heterologous matrix attachment regions. *Molecular Reproduction and Development* 44, 179–184.

Miller, K. and Rizzino, A. (1995) The function of inducible promoter systems in F9 embryonal carcinoma cells. *Experimental Cell Research* 218, 144–150.

Nielsen, P.E. and Haaima, G. (1997) Peptide nucleic acid (PNA). A DNA mimic with a pseudopeptide backbone. *Chemical Society Reviews* 26, 73–78.

Nielsen, P.E., Egholm, M., Berg, R.H. and Buchardt, O. (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* 254, 1497–1500.

Ninomiya, T., Hoshi, M., Mizuno, A., Nagao, M. and Yuki, A. (1989) Selection of mouse preimplantation embryos carrying exogenous DNA by polymerase chain reaction. *Molecular Reproduction and Development* 1, 242–248.

Ninomiya, T., Iwabuchi, T., Soga, Y. and Yuki, A. (1990) Direct sequencing of flanking regions of a transgene amplified by inverted PCR. *Agricultural and Biological Chemistry* 7, 1869–1872.

No, D., Yao, T.P. and Evans, R.M. (1996) Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proceedings of the National Academy of Sciences USA* 93, 3346–3351.

Norton, J.C., Waggenspack, J.H., Varnum, E. and Corey, D.R. (1995) Targeting peptide nucleic acid-protein conjugates to structural features within duplex DNA. *Bioorganic and Medicinal Chemistry* 3, 437–445.

O'Gorman, S., Fox, D.T. and Wahl, G.M. (1991) Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* 251, 1351–1355.

Page, R.L., Canseco, R.S., Russell, C.G., Johnson, W.H., Velander, W.H. and Gwazdauskas, F.C. (1995) Transgene detection during early murine embryonic development after pronuclear microinjection. *Transgenic Research* 4, 12–17.

Peden, K.W.C., Pipas, J.M., Pearson-White, S. and Nathan, D. (1980) Isolation of mutants of an animal virus in bacteria. *Science* 209, 1392–1396.

Perlak, F.J., Fuchs, R.L., Dean, D.A., McPherson, S.L. and Fischhoff, 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.

Perlak, F.J., Stone, T.B., Muskopf, Y.M., Petersen, L.J., Parker, G.B., McPherson, S.A., Wyman, J., Love, S., Reed, G. and Biever, D. (1993) Genetically improved potatoes: protection from damage by Colorado potato beetles. *Plant Molecular Biology* 22, 313–321.

Peterson, K.R., Clegg, C.H., Navas, P.A., Norton, E.J., Kimbrough, T.G. and Stamatoyannopoulos, G. (1996) Effect of deletion of 5′HS3 or 5′HS2 of the human *beta-globin* locus control region on the developmental regulation of *globin* gene expression in *beta-globin* locus yeast artificial chromosome transgenic mice. *Proceedings of the National Academy of Sciences USA* 93, 6605–6609.

Phi-Van, L. and Stratling, W.H. (1996) Dissection of the ability of the chicken lysozyme gene 5′ matrix attachment region to stimulate transgene expression and to dampen position effects. *Biochemistry* 35, 10735–10742.

Roseman, R.R., Johnson, E.A., Rodesch, C.K., Bjerke, M., Nagoshi, R.N. and Geyer, P.K. (1995) A P element containing suppressor of hairy-wing binding regions has novel properties for mutagenesis in *Drosophila melanogaster*. *Genetics* 141, 1061–1074.

Schultze, N., Burki, Y., Lang, Y., Certa, U. and Bluethmann, H. (1996) Efficient control of gene expression by single step integration of the tetracycline system in transgenic mice. *Nature Biotechnology* 14, 499–503.

Seo, B.B., Kim, C.H., Tojo, H., Tanaka, S., Yamanouchi, K., Takahashi, M., Sawasaki, T. and Tachi, C. (1997) Efficient selection of preimplantation transgenic embryos by an improved procedure using *Dpn I-Bal 31* digestion and the polymerase chain reaction. *Reproduction, Fertility and Development* 9, 263–269.

Slightom, J.L., Bock, J.H., Tagle, D.A., Gumucio, D.L., Goodman, M., Stojanovic, N., Jackson, J., Miller, W. and Hardison, R. (1997) The complete sequences of the galago and rabbit *beta-globin* locus control regions: extended sequence and functional conservation outside the cores of DNase hypersensitive sites. *Genomics* 39, 90–94.

Smith, P.A. and Corces, V.G. (1992) The suppressor of *Hairy-wing* binding region is required for *Gypsy* mutagenesis. *Molecular and General Genetics* 233, 65–70.

Sparks, A.E.T., Canseco, R.S., Russell, C.G., Johnson, J.L., Moll, H.D., Velander, W.H. and Gwazdauskas, F.C. (1994) Effects of time of deoxyribonucleic acid microinjection on gene detection and *in vitro* development of bovine embryos. *Journal of Dairy Science* 77, 718–724.

Stamatoyannopoulos, J.A., Clegg, C.H. and Li, Q. (1997) Sheltering of *gamma-globin* expression from position effects requires both an upstream locus control region and a regulatory element 3′ to the A *gamma-globin* gene. *Molecular and Cellular Biology* 17, 240–247.

Stief, A., Winter, D.M., Stratling, W.H. and Sippel, A.E. (1989) A nuclear DNA attachment element mediates elevated and position-independent gene activity. *Nature* 341, 343–345.

Tada, N., Sato, M., Hayashi, K., Kasai, K. and Ogawa, S. (1995) *In vitro* selection of transgenic mouse embryos in the presence of G-418. *Transgenics* 1, 535–540.

Takada, T., Iida, K., Awaji, T., Itoh, K., Takahashi, R., Shibui, A., Yoshida, K., Sugano, S. and Tsujimoto, G. (1997) Selective production of transgenic mice using green fluorescent protein as a maker. *Nature Biotechnology* 15, 458–461.

Tewari, R., Gillemans, N., Harper, A., Wijgerde, M., Zafarana, G., Drabek, D., Grosveld, F. and Philipsen, S. (1996) The human beta-globin locus control region confers an early embryonic erythroid-specific expression pattern to a basic promoter driving the bacterial *lacZ* gene. *Development* 122, 3991–3999.

Thomas, H.E., Stunnenberg, H.G. and Stewart, A.F. (1993) Heterodimerization of the *Drosophila* ecdysone receptor with retinoid X receptor and ultraspiracle. *Nature* 362, 471–475.

Thompson, E.M., Adenot, P., Tsuji, F.I. and Renard, J.P. (1995) Real time imaging of transcriptional activity in live mouse preimplantation embryos using a secreted luciferase. *Proceedings of the National Academy of Sciences USA* 92, 1317–1321.

Wall, R.J. (1996) Transgenic livestock: progress and prospects for the future. *Theriogenology* 45, 57–68.

Wang, Y., O'Malley, B.W.J., Tsai, S.Y. and O'Malley, B.W. (1994) A regulatory system for use in gene transfer. *Proceedings of the National Academy of Sciences USA* 91, 8180–8184.

Wang, Y., DeMayo, F.J., Tsai, S.Y. and O'Malley, B.W. (1997) Ligand-inducible and liver-specific target gene expression in transgenic mice. *Nature Biotechnology* 15, 239–243.

Weinmann, P., Gossen, M., Hillen, W., Bujard, H. and Gatz, C. (1994) A chimeric transactivator allows tetracycline-responsive gene expression in whole plants. *Plant Journal* 5, 559–569.

Yao, T.P., Segraves, W.A., Oro, A.E., McKeown, M. and Evans, R.M. (1992) *Drosophila* ultraspiracle modulates ecdysone receptor function. *Cell* 71, 63–72.

Yu, Z., Redfern, C.S. and Fishman, G.I. (1996) Conditional transgene expression in the heart. *Circulation Research* 79, 691–697.