

Chapter 1

Design of Gene Constructs for Transgenic Maize

Dong Liu

Summary

The first step of any maize transformation project is to select gene expression elements that will make up an effective construct. When designing a gene construct, one must have a full understanding of the different expression elements that are currently available and of the strategies that have been successfully used to overcome obstacles in past. In this chapter, we discuss several major classes of expression elements that have been used for maize transformation, including promoters, introns, and untranslated regions. We also discuss several strategies for further improving transgene expression levels, such as optimization of codon usage, removal of deleterious sequences, addition of signal sequences for subcellular protein targeting, and use of elements to reduce position effects. We hope that this chapter can serve as a general guideline to help researchers, especially beginners in the field, to design a gene construct that will have the maximum potential for gene expression.

Keywords: Transgenic maize, Gene construct, Promoters, Introns, UTRs, Codon bias

1. Introduction

To successfully express a foreign gene in transgenic maize, it is critical to properly select the expression elements that are used to control both the “gene of interest” and the selectable marker gene. The expression level of an introduced gene in transgenic maize can be affected by many factors, including promoter activity, stability of mRNA, protein translation efficiency, protein stability in a particular cellular compartment, the chromosomal position of the integrated gene, and degree of gene silencing. This chapter will discuss how these factors should be considered and how to properly select expression-controlling elements when designing a gene construct for maize transformation.

Because data reported in the literature are not always consistent, care must be taken when incorporating some features into the gene construct. The selection criteria for the expression elements discussed here may be used only for a general maize transformation project. More sophisticated techniques, such as generation of marker-free transgenic plants, and the use of homologous recombination systems for site-specific gene integration, are beyond the scope of this chapter and can be found in some review articles if one is interested (1, 2).

2. An Overview of Plant Gene Expression

In plant genomes, a functional gene unit generally consists of a DNA segment which encodes a protein, and regulatory DNA elements which control the expression of the protein coding region. **Figure 1** illustrates the steps that are involved in the process of plant gene expression. The first step of gene expression is the transcription of DNA to pre-mRNA, which is initiated by the regulatory element called the promoter. It is usually located at the 5' end of the gene, upstream of the transcription start site. The promoter contains various sequence elements that function in the recruitment of protein factors that facilitate transcription of the protein coding region of the gene. The most basic elements are the TATA box and CAAT box, the names of which are based on their nucleotide sequences. Transcription termination is determined by a terminator sequence, which is located downstream of the translation termination site.

Once the pre-mRNA is transcribed from its DNA template, a few steps remain before the mRNA is mature. Splicing removes intron sequences that do not encode protein. A 7-methylguanosine nucleotide cap is added to the 5' end. Near the 3' end, a specific sequence, typically AAUAAA, is recognized by cellular machinery, generating a mature 3' end by cleavage about 20–30 nucleotides away from the motif and the addition of a poly (A) tail. Since the terminator contains the sequence motif for polyadenylation and the whole 3' untranslated region, or 3' UTR, the terms terminator, 3' polyadenylation sequence, and 3' UTR have been used interchangeably in the context of vector construction.

Thus, the mature mRNA molecule consists of a protein coding region and UTR at both the 5' and 3' ends, with a 5' cap and 3' poly A tail. Finally, the mature mRNA will be translated into a protein. The resulting protein will be further modified, folded into the proper configuration, and targeted into a specific cellular compartment.

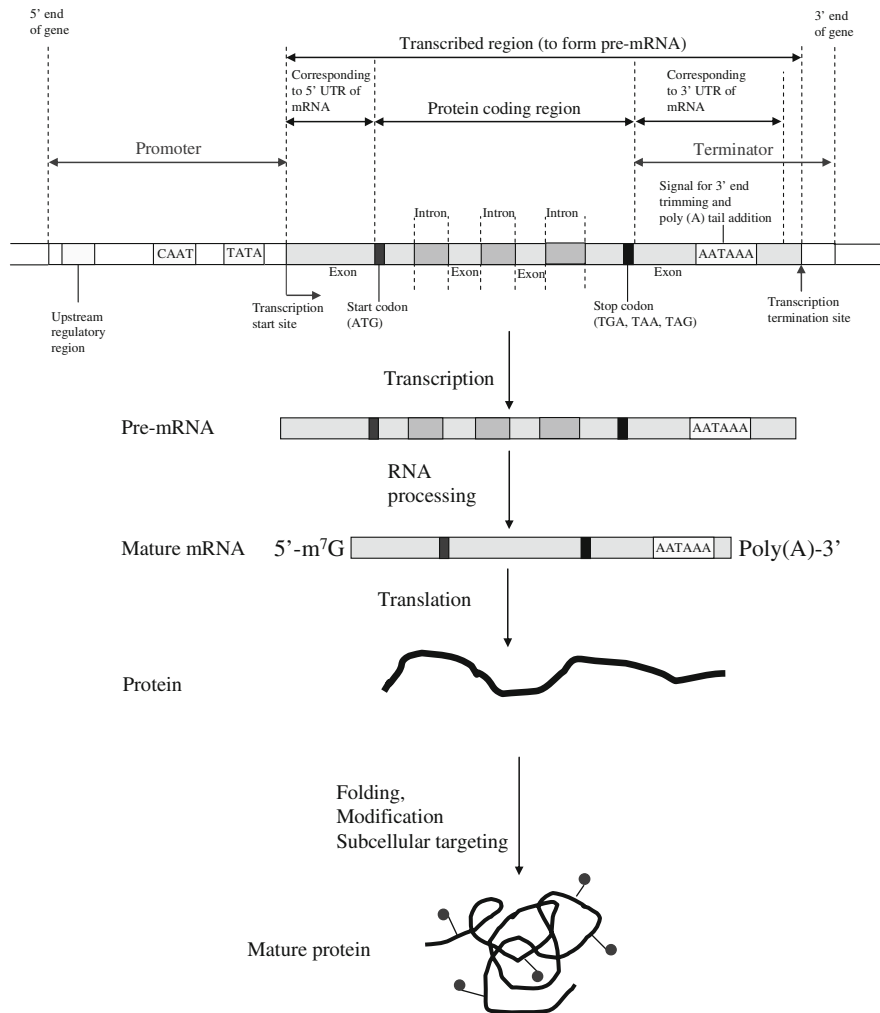


Fig. 1. An overview of the process of plant gene expression.

3. Promoters and Introns

The promoter is the most critical element for controlling gene expression. Promoters can be divided into three general categories: constitutive, tissue-specific, and inducible. A constitutive promoter drives expression of a gene in all plant tissues and throughout all developmental stages. A tissue-specific promoter expresses the gene in only a specific type of tissue. Tissue-specific promoters may or may not cause expression in all developmental stages. An inducible promoter only initiates gene expression under certain external conditions, such as light, temperature, nutrient levels, or in response to application of a particular chemical.

In the early days, the 35S promoter from Cauliflower Mosaic Virus (*CaMV* 35S) was most commonly used to drive high-level gene expression in plant tissue (3). However, further study showed that the activity of *CaMV* 35S promoter in monocot plants is much lower than the activity in dicot plants. To improve its activity in monocots, several modifications have been made, including the duplication of the enhancer region of the promoter (4). Another is the insertion of an intron found in various promoters of cereal plants. The intron is usually inserted between the 35S promoter and the initial codon of the transgene coding region.

Introns that have been tested include those from the *Adh1*, *Sh1*, *Ubi1*, and *Act1* genes from maize, and the *Chs* gene from petunia (5). Among them, the intron from the maize *Ubi1* promoter gives the highest level of enhancement. The intron from the *Chs* gene gives a surprising 100-fold enhancement. These introns are all quite long (about 1 kb) and located within the 5' UTRs in their native genes. The mechanism of enhancement produced by these introns in improving gene expression is still not well understood. It has been generally thought that inclusion of an intron can improve efficiency of mRNA processing and increase the steady-state levels of mRNA (6). Even though the activity of unmodified *CaMV* 35S promoter is lower in monocots than in dicots, it has still been routinely used in maize transformation, when combined with an intron or some UTR sequences, for expressing screenable or selectable marker genes (7).

So far, the constitutive promoter most commonly used to drive a "gene of interest" in transgenic maize is the promoter of maize *Ubi1* combined with the first intron located in its 5' UTR region (8). As with the examples mentioned earlier, the inclusion of this intron greatly enhances the level of transgene expression. ProdiGene, a plant biotechnology company in Texas, has used the *Ubi1* promoter to express avidin and beta-glucuronidase (GUS) genes in transgenic maize, both of which are used as diagnostic agents in molecular biology. In highly expressive transgenic lines, avidin constitutes over 2% of the aqueous protein extracted from dry maize seeds while the recombinant GUS protein can accumulate to 0.7% of total soluble proteins (9, 10). These two products from transgenic maize are now commercially available from Sigma-Aldrich. It has been estimated that the cost of producing avidin in transgenic maize is about only 10% of the cost of extraction from egg white.

ProdiGene also used the *Ubi1* promoter to drive four transgenes that are the components of a secretory antibody – heavy chain, light chain, joining chain, and a secretory component (11). In this construct, each gene is expressed from the same *Ubi* promoter. In these transgenic maize seeds, they could detect the assembled antibodies with expression levels of up to 0.3% of

total soluble protein in T1 seeds. Based on ProdiGene's success with other products, the selection of high-performance lines and backcrossing should allow this yield to be increased as much as 70-fold over six generations.

Besides *CaMV 35S* and *Ubi1* promoters, the promoter of the maize histone *H2B* gene has also proven capable of driving constitutively high-level expression of a transgene, based on the recovered rate of transformants and histochemical analysis of reporter gene expression (12). However, there are no data reported on the quantitative side-by-side comparison of *H2B*, *Ubi1*, and other promoters in maize plants.

In transgenic rice, rice *Act1* is the most commonly used constitutive promoter to drive high-level expression of a transgene (13, 14). The structure and function of this promoter has been well characterized in rice plants. However, for unknown reasons, this promoter has not been extensively exploited in transgenic maize.

Constitutive promoters are very useful in directing expression of selectable or screenable marker genes, because screening requires the expression of these genes being initiated as soon as possible after integration into plant genome. This ensures that selection for the transformants can be effective and can reduce the false positive rate. There are other uses for a strong constitutive promoter besides marker genes. Herbicide-resistance genes allow breeders to easily follow the presence of the transgene in a plant population and create the valuable trait of crop resistance to weed-control chemicals. Constitutive expression of insecticidal proteins is useful if the target pest is one that feeds on all or most plant parts.

In some cases, specific promoters have advantages over strong constitutive promoters. For example, confining the production of a pharmaceutical recombinant protein to a certain tissue type could prevent unwanted dispersal of the protein in pollen grains. Tissue-specific expression could also aid in easier harvesting and extraction of the protein of interest. Localized expression via a tissue-specific promoter can prevent unnecessary depletion of plant nutrients when the protein of interest is not needed in all parts of the plant. This is applicable when using insecticidal proteins to deter pests that attack only certain parts of the plant.

There has been some research into using tissue-specific promoters for pest control in transgenic maize. Johnson et al. (15) used a putative silk-specific promoter to drive the expression of a maize Myb transcription factor in transgenic maize, enhancing resistance to corn earworm, which feeds on corn silks. Corn leaf aphid is a sap-sucking pest, which feeds on phloem tissues. Using a phloem-specific promoter of rice sucrose synthase gene, Wang et al. (16) expressed snowdrop lectin in transgenic maize. The expression level of snowdrop lectin ranged from 0.13 to 0.28%

of total soluble protein. Three lines with expression levels above 0.22% showed significant resistance to aphids.

In transgenic maize, the most commonly used tissue-specific promoters are for the expression of recombinant protein in seeds. Advantages of expressing protein in seeds include confinement of the transgene product, easy storage prior to protein extraction, and high-level expression. Various seed-specific promoters have been used in maize, many derived from seed storage protein genes. These include promoters from maize 27-kDa zein (*zmZ27*), maize *waxy* (starch synthase) genes, rice *glutelin-1*, and the small subunit of ADP-glucose pyrophosphorylase gene, all of which are maize endosperm-specific (17, 18). The embryo-specific *globulin-1* promoter has also been used to express recombinant proteins in maize (19).

Recently, Yu et al. (20) used maize seed storage protein promoter *P19z* to drive the expression of the lysine-rich pollen-specific protein Sb401 from potato in maize. Their results showed that the expression level of Sb401 was correlated with increased levels of lysine and total protein content in maize seeds. In their R1 maize seeds, compared with the nontransgenic maize control, the lysine content increased by 16.1–54.8% and the total protein content increased by 11.6–39.0%. Further analysis showed that the levels of lysine and total protein remained high for six continuous generations, indicating that the elevated lysine and total protein levels were heritable. However, care must be taken because although these promoters are often described as seed-specific promoters, a low level of activity can also be detected in other plant tissues. For example, maize *waxy* promoter that is endosperm-preferred also shows low activity in pollen (17).

An alternative way to produce recombinant proteins in transgenic maize is to use maize cell culture. Cell culture confers the advantages of confinement of the transgene and its product. The short life cycle in cell culture allows for faster harvesting of the protein of interest. In addition to constitutive promoters, inducible promoters may be considered for this purpose. In rice cell culture, it has been demonstrated that the rice *alpha-amylase* promoter can drive high-level gene expression during sugar starvation (21). Recently, the same authors used this promoter combined with the signal peptide sequence of the *amylase* gene to express human serum albumin (HSA) in rice cell culture (22). Upon sugar starvation, the highest yield of HSA that is secreted into the culture medium can reach 74 mg/L. This expression system may be adopted for the production of recombinant pharmaceutical proteins in maize cell culture. To date, no research has been conducted on this subject.

In some cases, expression of a transgene can have detrimental effects on the host cells. Inducible promoters have the advantage of tight control over the timing of transgene expression. These

promoters allow expression, only after exposure to the proper chemical. As with tissue-specific promoters, inducible promoters avoid or reduce unnecessary consumption of nutrients and energy in transgenic plants or cultured cells. Several chemically regulated plant promoters have been reported in the literature (23). None of them has been adopted for field use, mainly due to government regulation of chemicals released into the environment. Some chemicals, such as the antibiotics tetracycline (24) and the steroid dexamethasone, are not suitable to be used in large quantities in the field, though dexamethasone has been widely used as an inducer in laboratory conditions (2).

4. Untranslated Regions

Regulation of mRNA stability and protein translation efficiency is just as important as promoter choice when considering levels of gene expression. Many studies have shown that the 5' and 3' UTRs on the transcript can play an important role in gene expression levels (25). The presumed role of the 5' UTR is to influence the efficiency with which the bound ribosomal subunits migrate and recognize the translation start codon. A useful 5' UTR for transgene expression must be able to efficiently assemble ribosomes and present the start codon in an appropriate configuration. Generally, an AT-rich 5' UTR allows the ribosomal complex to easily scan to the start codon to initiate translation.

When the 5' UTR from CaMV 35S was used with its promoter in maize protoplasts, it enhanced reporter gene expression about 40-fold (26). The 5' UTRs from the maize genes *glutelin* (27) and *PEP-carboxylase* (28) have also been tested in maize protoplasts, resulting in 12-fold and 3.7-fold increased expression, respectively. Effects of a given 5' UTR on the transgene expression can be very different in monocots and dicots, and not every 5' UTR has an enhancing effect. For example, the 5' UTRs from genomic RNA of tobacco mosaic virus and alfalfa mosaic virus coat protein have been particularly effective in enhancing transgene expression in tobacco (29, 30), but they are much less effective in maize cells (31). In contrast, the 5' UTR from a tobacco etch virus has been effectively used to drive a selectable marker gene in maize plants when combined with the *CaMV 35S* promoter (7, 32).

There is evidence that the 3' UTR determines stability of mRNA as well as facilitates polyadenylation of the mRNA. A 3' UTR sequence with a strong stem-loop structure can effectively block degradation of mRNA by RNase that starts from the 3' end of the transcript. Ingelbrecht et al. (33) examined the effects of 3'

UTRs from several different genes, finding that their effects on transgene expression vary widely. The different 3' ends used for testing were from the octopine synthase gene on the Ti plasmid of *Agrobacterium tumefaciens*, the 2S seed protein gene from Arabidopsis, extensin gene from carrot, the gene of the small subunit of *rbcS* from Arabidopsis, and chalcone synthase gene from Antirrhinum. In stable tobacco transformants, there was about a 60-fold difference between the best-expressing construct (with the 3' UTR from small subunit of *rbcS*) and the lowest-expressing construct (with the 3' UTR from a chalcone synthase), with the other 3' UTRs producing different expression levels between these two extremes.

A popular choice of the 3' UTR for both monocot and dicot plants, including maize, is the one from the nopaline synthase (*nos*) gene of *Agrobacterium* Ti plasmid (34). This small unit (about 260 bp) is probably the most widely used 3' UTR today in transgenic plants. Other commonly used 3' UTRs include those from *CaMV* 35S transcripts (7), the potato *pinII* gene (35), and the soybean vegetative storage protein gene (36). To date, no comparative studies on the effects of different 3' UTRs on the transgene expression in transgenic maize plants have been reported.

Inclusion of the UTR sequences from plant genes in constructs for expression of bacterial genes in transgenic plants is particularly useful. In our previous work, we attached the 5' and 3' UTR sequences from a tobacco osmotin gene to both ends of the full-length *Toxin A* gene from bacterium *Photobacterium luminescens*, and several of its derivatives (37). The *Toxin A* gene is 7.5-kb long, encoding a 283-kD insecticidal protein. Achieving expression of *Toxin A* in plant cells is very difficult. With the attachment of the UTR sequences from the osmotin gene, expression levels of *Toxin A* protein in Arabidopsis plants were increased 10-fold and the recovery rate of insect-resistant plants increased 12-fold. We chose the UTR sequences from tobacco osmotin because both osmotin mRNA and protein are very stable in plant cells. Computer modeling of its 3' UTR showed many stem-loop structures, which may effectively block degradation by RNase (Liu et al. unpublished). The attachment of plant UTR sequences makes this prokaryotic gene behave more like a "plant gene," so corresponding mRNA can be easily recognized by plant translation machinery. Homologs of osmotin exist in almost every monocot plant, including maize (38). It will be interesting to test if the UTR sequences from the maize osmotin homolog gene will have the same enhancing effect on gene expression in transgenic maize.

The sequence that immediately surrounds the translation start codon is another important factor that influences the translation efficiency. Kozak (39) compared vertebrate mRNA sequences to

the results of site-directed mutagenesis, and demonstrated the existence of a preferred nucleotide context surrounding the initial codon. Kozak has defined an optimal AUG context for vertebrates as GCC(A/G)CCAUGG (the start codon is underlined), named the Kozak consensus sequence. The most highly conserved position of the consensus is position -3 (the A of the AUG codon being +1) where 97% of vertebrate mRNAs have a purine. The -3 position appears to mediate the efficiency of translation initiation in vertebrate systems. Such a consensus sequence has also been found among plant genes: UAACAAUGGCU (40, 41). In position +4, the preference for a guanine is significantly high in plants (85%). The preference for G in position +4 (85%) and C at +5 (77%) in plants resembles the preference for A in position -3 in animals. Further compilation of sequences surrounding the AUG from 85 maize genes yields a consensus of (C/G)AUGGCG (42). This consensus sequence should be taken into consideration when making gene constructs for use in maize.

5. Codon Bias

During early attempts to express a foreign gene in transgenic plants, it was frequently found that only a very low amount of the foreign protein accumulated. This is particularly true when introducing a prokaryotic gene into plant cells. The example that best illustrates this point is of the expression of insecticidal crystal proteins of *Bacillus thuringiensis* (*Bt*). When Vaeck et al. (43) first introduced *cryIA(b)* gene into plant cells, they found that the expressed foreign protein was only 0.0002–0.02% of total soluble protein. Similar results were reported by other researchers. Subsequent analysis of these transgenic plants showed that mRNA species were shorter than the expected full-length transcript, suggesting inefficient posttranscriptional processing or rapid turnover of full-length transcripts (44).

Close examination of many bacterial native endotoxin genes has found that they tend to have a very low G+C content, around 37%. In contrast, plant genes in general tend to have a higher G+C content, with maize showing a strong preference for G+C-rich coding regions (45, 46). The A+T-rich sequence in bacterial genes may provide many putative motifs, which will function in plant cells as splice sites, polyadenylation signal sequences (40, 44, 47), and RNA destabilizing elements, such as ATTTA. The short motif of ATTTA was first found in animal cells, causing mRNA instability when present in the the 3' UTR of the transcript (48). This phenomenon has been confirmed in plant cells (49). Since all plant introns contain A/T-rich sequences (70), the presence of AT-rich

elements, such as four or more consecutive A or T bases, may facilitate aberrant pre-mRNA splicing. With the presence of these deleterious sequence motifs, early constructs for Bt in plant cells produced highly unstable mRNA, resulting in very little translation of intact protein.

Codon bias is another major obstacle preventing expression of bacterial genes in transgenic plants. Different species may exhibit a preference or bias for one of many synonymous codons for a given amino acid. For the amino acid arginine in alfalfa, the codon CGU is 50 times more likely to occur in the genome than the rarest codon, CGG. In maize, AGG and CGC are the preferred codons for arginine, but the codon CGG is also frequently used. **Table 1** lists the codon frequencies in the maize genome.

Many codons commonly found in maize are rare in bacteria, and vice versa. Codon frequency for two insecticidal protein

Table 1
Codon usage of 2,280 coding sequences of maize genes

Amino acid	Codon	Freq	Amino acid	Codon	Freq	Amino acid	Codon	Freq	Amino acid	Codon	Freq
Ala	GCU	21.1	Gln	CAA	13.3	Leu	TTA	5.7	Ser	TCG	10.5
Ala	GCC	31.2	Gln	CAG	23.5	Leu	CTG	25.8	Ser	TCA	11.2
Ala	GCA	16.7	Glu	GAG	40.9	Leu	CTA	7.3	Thr	ACC	16.5
Ala	GCG	23.1	Glu	GAA	20.0	Lys	AAG	39.6	Thr	ACT	10.8
Arg	AGG	14.8	Gly	GGT	14.3	Lys	AAA	15.0	Thr	ACA	10.5
Arg	CGC	14.3	Gly	GGC	30.2	Met	ATG	24.1	Thr	ACG	10.8
Arg	AGA	8.8	Gly	GGA	13.3	Phe	TTC	25.1	Trp	TGG	12.9
Arg	CGT	6.1	Gly	GGG	15.3	Phe	TTT	12.6	Tyr	TAC	19.3
Arg	CGG	9.4	His	CAC	14.8	Pro	CCA	13.9	Tyr	TAT	9.4
Arg	CGA	4.3	His	CAT	10.1	Pro	CCT	12.6	Val	GTC	21.1
Asn	AAC	22.2	Ile	ATC	23.0	Pro	CCC	13.5	Val	GTG	25.6
Asn	AAT	13.5	Ile	ATT	14.0	Pro	CCG	15.4	Val	GTT	15.8
Asp	GAC	32.2	Ile	ATA	8.4	Ser	TCC	16.2	Val	GTA	6.3
Asp	GAT	23.0	Leu	CTC	25.5	Ser	TCT	12.1	Ter	TGA	1.1
Cys	TGC	12.1	Leu	TTG	13.2	Ser	AGC	16.1	Ter	TAA	0.5
Cys	TGT	5.6	Leu	CTT	15.9	Ser	AGT	7.8	Ter	TAG	0.7

This table is adopted from the following website with some modifications: [http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=Zea+mays+\[gbpln\]](http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=Zea+mays+[gbpln]) *Freq*: occurred frequency per thousand codons

genes (*tcdA* and *tcbA*) from the bacterium *P. luminescens* can be found in an issued patent (Patent No. US 6,590,142 B1). The codons GCG for alanine, GAA for glutamate, and CTG for leucine were not used in these two bacterial genes, but their frequencies in maize are about 23, 20, and 26% in each amino acid group, respectively.

The low expression levels correlated with codon bias are probably due to the small pool size of the specific tRNAs that are encoded by the rare codons. Presence of these rare codons could cause ribosomes to stall during translation of the transcripts. Stalled ribosomes may cause mRNA destabilization by leaving the transcripts exposed to components of the RNA degradation machinery.

Besides the average codon usage for amino acids, attention should also be paid to some other unique features in coding sequences of plant genes. Intercodon nucleotide doublets CG and AT are rarely seen in plant genes (46). In contrast, CT and TG doublets frequently occur. The potential RNA polymerase II termination sequence CAN7-9AGTNNA, though not extensively studied, is believed to be important for transcription termination (50). Furthermore, although maize genes usually have codons with G or C in the third position, a majority of codons in endotoxin genes have A or T in the third position. Thus, the presence of sequences deleterious to mRNA stability in plant cells, coupled with differences in codon bias, can make bacterial genes extremely difficult to express in plant cells.

Using a completely synthesized coding sequence is an effective way to solve the problem of transgene expression. By reengineering the whole coding sequence, one can not only change the codon preference and elevate the overall G+C content that is close to the level of the targeting plant species, but also can remove all the deleterious sequence motifs that may cause mRNA instability. **Table 2** lists some sequence motifs that should be avoided during the gene rebuilding process. Through gene rebuilding, people have made significant improvement in expression of endotoxin genes in cotton (51), tomato (52), and potato (53).

The first synthetic gene that was successfully expressed in transgenic maize is the *cryIA(b)* gene from Bt var. Kurstaki HD-1 (54). The coding region for the first 648 amino acids of the 1,155 amino acid CryIA(b) was synthesized using the most preferred codons from maize for each amino acid (55). The synthetic gene had 65% homology with the native gene and a G+C content of 65%, compared with 37% of the native gene. When compared with the native gene, the synthetic gene produced significantly higher level of CryIA(b) protein in both tobacco and maize. In maize plants, the protein produced from native gene was not detectable, while the synthetic gene produced up to 4 ug/mg soluble protein in certain plants.

Table 2
Sequence motifs that should be avoided in synthetic bacterial gene expression in plants (44, 47, 48)

RNA destabilizing element	Polyadenylation signal sequence	Splice site	Intercodon doublet	Nucleotide in 3rd position	AT-rich sequence	RNAP II
ATTTA	AATAAA ^a	GT	CG	A and T	(A) 4 or more	CAN7-9-AGTNNNA
	AATTAT ^b	AG	AT		(T) 4 or more	
	AGTATA ^b					
	AATATT ^b					
	AATGAA ^c					
	AAAATA ^c					
	AAGCAT ^c					
	AATAAT ^c					
	AATAAG ^c					
	ATACAT ^c					
	ATTAAT ^c					

RNAP II: RNA polymerase II termination sequence

^aThe most commonly used polyadenylation signal sequence in both monocots and dicots

^bThree sequence motifs found in bacterial endotoxin genes that also act as polyadenylation signals in plants

^cPotential polyadenylation signal sequences found in some monocot genes. There are other polyadenylation signal sequences of dicot plants not listed in this table

With the great success of using synthetic genes to express bacterial proteins in transgenic plants, the gene rebuilding process has become a routine practice in transgenic plant programs.

6. Subcellular Targeting

Protein stability is another key factor in regulating the expression level of a heterologous protein in transgenic plants. The cellular compartment in which a protein accumulates can strongly influence the interrelated process of folding, assembly, and posttranslational modification. Also, when trying to modify a biosynthetic pathway, it is important that the transgene product accumulates in the same compartment as the metabolites that will serve as substrate. For these reasons, subcellular targeting must be considered in order to obtain a desired trait in transgenic plants.

In plants, it has been clear that entry of a protein into the endoplasmic reticulum (ER) is mediated by a signal sequence of about 20 amino acids, generally located at its N-terminus. If no further signaling determinants are present in the protein, it will be loaded into the secretory pathway. The default destination for this pathway is the intercellular matrix or cell culture medium (56). In previously mentioned examples, ProdiGene used the maize *ubi1* promoter combined with a signal peptide from barley alpha-amylase to express avidin and *Trametes versicolor* laccase I isoform in maize seeds. They found that expression levels of each protein could reach up to 2% and 1%, respectively, and expressed recombinant protein accumulated in intercellular space (9, 19). High accumulation of the proteins in the intercellular space may be due to the lack of proteinase in these locations.

Another benign environment for protein accumulation in plant cells is the lumen of the ER. The mechanism by which proteins are retained in the ER has been well characterized and shown to be evolutionarily conserved among plants and animals. A tetrapeptide motif, usually KDEL or HDEL, present at the C-terminus of the protein, is necessary and sufficient to cause ER retention, or at least recycling back to the ER (57). There are various reports that proteins targeted to the ER by addition of a KDEL or HDEL sequence at its C-terminus can accumulate to relatively high levels in tobacco plants (58–60).

Retention of protein in the ER lumen has other advantages as well. The activity of many proteins is dependent on the posttranslational modification, such as the formation of disulfide bonds and glycosylation. These modification processes all occur in the lumen of the ER. Retention of foreign proteins in the ER lumen can also prevent the proteins from being transported into the protease-rich environment of the vacuole.

So far, this ER targeting strategy has not been vigorously exploited in transgenic maize. In an attempt to improve nutrition of maize with porcine alpha-lactalbumin, Yang et al. (61) made three gene constructs with a synthetic protein coding sequence containing maize preferred codons. All three constructs used the *ubi1* promoter and *nos* terminator for controlling gene expression. The first construct has no additional amino acid sequences attached to the porcine alpha-lactalbumin gene. The second and third constructs have signal peptide sequences added at the 5' end of the gene. One of the two has a KDEL sequence attached to its C-terminus. The transgenic plants with the first construct did not produce detectable levels of recombinant protein, indicating the importance of a signal peptide for protein accumulation. Both the second and third constructs produced recombinant protein in callus and kernels, but the data did not allow them to distinguish between the two constructs regarding the level of protein produced or the frequency of expression.

7. Position Effect

It is well known that expression levels of transgenes in primary transformants can vary substantially. The degree of variability differs among gene constructs and plant species, but even expression from the same construct in the same plant host can have considerable variation. Random placement of a transgene, known as the position effect, is the main explanation for high intertransformant variability. Transgene expression can be influenced by the properties of the surrounding genomic DNA: the presence of flanking promoters, enhancing or silencing DNA elements, the copy number of integrated transgene, the chromatin structure or methylation status, etc. Today, there is no transformation technology that can effectively control where the transgene is integrated into the chromosomes of the targeted plants. Thus, it is still a costly and time-consuming process to sort a large number of individual transformants to find one or several events that may have a desired trait with stable inheritability for several generations.

Reduced intertransformant variability has been reported with the use of special elements such as matrix attachment regions (MAR) or scaffold-associated regions (SARs) (62–64). It has also been reported that certain MARs provide expression levels in transgenic rice that are higher and more predictable (65, 66). However, the use of different MARs or SARs, in combination with various promoters, can have different effects on gene expression in different systems (67). In transgenic maize, Sidorenko et al. (68) tested the effects of two MARs, one from the *p1-rr* gene and another from the *Adh1* gene, on the variability of transgene expression when combined with three different promoters. Their results showed that these two MARs have no effects on the reduction of variability of transgene expression; however, the *Adh1*-MAR did modify the spatial pattern of expression driven by one synthetic promoter. In another report, Torney et al. (69) tested two MARs from the *Adh1* gene, revealing new transgene expression patterns in roots and systemically induced variegation, but no effects on reduction of transgene expression variability among the individual transformants. So far, no effective methods have been developed in transgenic maize to eliminate the position effect that causes the variability of transgene expression. The only way to obtain high-performance transgenic lines is to generate large numbers of transformants, and then select individuals with the desired trait through sorting.

In summary, the success of a maize transformation project is largely dependent on a construct design that incorporates the proper expression elements and strategies. Figure 2 is a diagram of the decision-making tree. We hope that this chapter will help

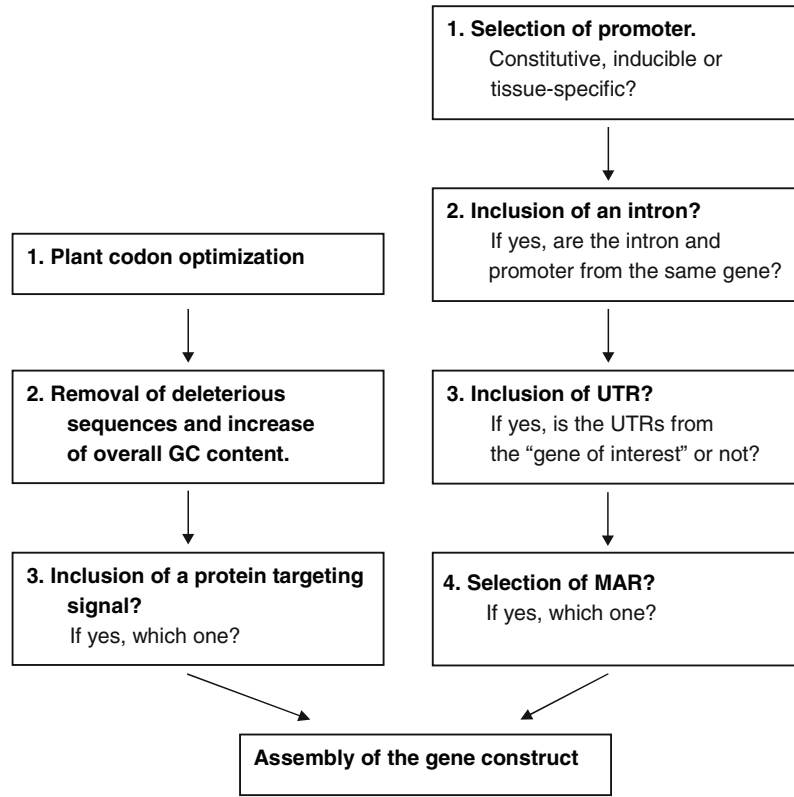


Fig. 2. A flow chart of the decision-making process for the design of a gene construct for maize transformation.

researchers design gene constructs that will have the maximum potential for expression of transgenes in maize.

Acknowledgments

The author sincerely wishes to thank Anastasia Bodnar for her critical reading and help on the manuscript preparation.

Note

The research on insect-resistant Arabidopsis plants with Toxin A gene (ref. 37) was conducted at Dow AgroScience LLC, 9330 Zionsville, IN 46268, U.S.A.

References

1. Darbani, B., Eimanifar, A., Stewart, C.N. Jr., and Camargo, W.N. (2007) *Methods to produce marker-free transgenic plants. Biotechnology* 2(1), 83–90.
2. Zuo, J., Niu, Q.W., Ikeda, Y., and Chua N.H. (2002) *Marker-free transformation: increasing transformation frequency by the use of regeneration-promoting genes. Curr. Opin. Biotechnol.* 13(2), 173–180.
3. Odell, J.T., Nagy, F., and Chua, N.H. (1985) *Identification of DNA sequence required for activity of the cauliflower mosaic virus 35S promoter. Nature* 313, 810–812.
4. Kay, R., Chan, A., Daly, M., and McPherson, A. (1987) *Duplication of CaMV 35S promoter sequence creates a strong enhancer for plant genes. Science* 236, 1299–1302.
5. Vain, P., Finer, K.R., Engler, D.E., Pratt, R., and Finer J. (1996) *Intron-mediated enhancement of gene expression in maize (Zea mays L.) and bluegrass (Poa pratensis L.). Plant Cell Rep.* 15, 489–494.
6. Lueharsen, K.R., and Walbot, V. (1991) *Intron enhancement of gene expression and splicing efficiency of introns in maize cells. Mol. Gen. Genet.* 225, 81–93.
7. Frame, B.R., Shou, H., Chikwamba, R.K., Zhang, Z., Xiang, C., Fonger, T.M., Pegg, S.E., Li, B., Nettleton, D.S., Pei, D., and Wang K. (2002) *Agrobacterium tumefaciens-mediated transformation of maize embryos using a standard binary vector system. Plant Physiol.* 129(1), 13–22.
8. Christensen, A.H., and Quail, R.H. (1996) *Ubiquitous promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledons plants. Transgenic Res.* 5, 231–218.
9. Hood, E., Witcher, D., Maddock, S., Meyer, T., Baszczyński, C., et al. (1997) *Commercial production of avidin from transgenic maize: characterization of transformant, production, processing, extraction and purification. Mol. Breed.* 3, 291–306.
10. Witcher, D.R., Hood, E.E., Peterson, D., Bailey, M., Bond, D., et al. (1998) *Commercial production of β -glucuronidase (GUS): a model system for the production of proteins in plants. Mol. Breed.* 4, 301–312.
11. Hood, E.E., Woodard, S.L., and Horn, M. (2002) *Monoclonal antibody manufacturing in plants: myths and realities. Curr. Opin. Biotechnol.* 13, 630–635.
12. Rasco-Gaunt, L.S., Li, C., Hagemann, K., Riley, A., et al. (2003) *Characterization of the expression of a novel constitutive maize promoter in transgenic wheat and maize. Plant Cell Rep.* 21, 569–576.
13. McElroy, D., Zhang, W., Cao, J., and Ray, W. (1990) *Isolation of an efficient actin promoter for use in rice transformation. Plant Cell* 2(2), 163–171.
14. Wang, Y., Zhang, W., Cao, J., and Ray, W. (1992) *Characterization of cis-acting elements regulating transcription from the promoter of a constitutively active rice actin gene. Mol. Cell Biol.* 12(8), 399–406.
15. Johnson, E., Berhow, M., and Dowd, P. (2007) *Expression of a maize Myb transcription factor driven by a putative silk-specific promoter significantly enhances resistance to Helicoverpa zea in transgenic maize. J. Agric. Food Chem.* 55, 2998–3003.
16. Wang, Z., Zhang, K., Sun, X., Tang, K., and Zhang, J. (2005) *Enhancement of resistance to aphids by introducing the snowdrop lectin gene into maize plants. J. Biosci.* 30, 627–638.
17. Russell, D.A., and Fromm, M.E. (1997) *Tissue-specific in transgenic maize of four endosperm promoters from maize and rice. Transgenic Res.* 6, 157–168.
18. Zheng, Z.W., Kawagoe, Y., Xiao, S.H., Li, Z., and Okita, T. (1993) *5' distal and proximal cis-acting regulator elements are required for developmental control of a rice seed storage protein glutelin gene. Plant J.* 4, 357–366.
19. Hood, E.E., Bailey, M.R., Beifuss, K., Magallane-Lundback, M., and Callaway, E. (2003) *Criteria for high-level expression of a fungal laccase gene in transgenic maize. Plant Biotechnol. J.* 1, 129–140.
20. Yu, J., Peng, P., Zhang, X., Zhao, Q., Zhu, D., Sun, X., Liu, J., and Ao, G. (2005) *Food Nutr. Bull.* 26, 427–431.
21. Chan, M., Chao, Y., and Yu, S. (1994) *Novel gene expression system for plant cells based on induction of alpha-amylase promoter by carbohydrate starvation. J. Biochem. Chem.* 269, 17635–17641.
22. Huang, L., Liu, Y., Lu, C., Hsieh, S., and Yu, S. (2005) *Production of human serum albumin by sugar starvation induced promoter in rice cell culture. Transgenic Res.* 14, 569–581.
23. Gatz, C. (1997) *Chemical control of gene expression. Annu. Rev. Plant Physiol. Plant Mol. Biol.* 48, 89–108.
24. Weinmann, P., Gossen, M., Hillen, W., Bujard, H., and Gatz, C. (1994) *A chimeric transactivator allows tetracycline-responsive gene expression in whole plants. Plant J.* 5, 559–569.

25. Gallie, D., and Bailey-Serre, J. (1997) *Eyes off transcription! The wonderful world of post-transcriptional regulation*. *Plant Cell* 9, 667–673.
26. Rothstein, S.J., Lahners, K.N., Lostein, R.J., Carozzi, N.B., and Jayne, S.M. (1987) *Rice DA: promoter cassettes, antibiotic-resistance genes, and vectors for plant transformation*. *Gene* 53, 153–161.
27. Boronat, A., Martinez, M.C., Reina, M., Puigdomenech, P., and Palau, J. (1986) *Isolation and sequencing of a 28 kd glutenin-2 gene from maize: common elements in the 5' flanking regions among zein and glutenin genes*. *Plant Sci.* 47, 95–102.
28. Hudspeth, L.W., and Grula, J.W. (1989) *Structure and expression of the maize gene encoding the phosphoenolpyruvate carboxylase isozyme involved in C4 photosynthesis*. *Plant Mol. Biol.* 12, 579–589.
29. Dowson, D., Ashurst, M.J., Mathias, J.L., Watts, S.F., Wilson, T.M., and Dixon, R.A. (1993) *Plant viral leaders influence expression of a reporter gene in tobacco*. *Plant Mol. Biol.* 23, 97–109.
30. Jobling, S.A., and Gehrke, L. (1987) *Enhanced translation of chimeric messenger RNAs containing a plant viral untranslated leader sequence*. *Nature* 325, 622–625.
31. Gallie, D.R., and Young, T.E. (1994) *The regulation of expression in transformed maize aleurone and endosperm protoplast*. *Plant Physiol.* 106, 929–939.
32. Carrington, J.C., and Freed, D.D. (1990) *Cap-independent of enhancement of translation by a plant poty virus 5' non-translated region*. *J. Virol.* 64, 1590–1597.
33. Ingelbrecht, L.W., Herman, L.M.F., Dekeyser, R.A., Van Montagu, M.C., and Depicker, A.G. (1989) *Different 3' end regions strongly influence the level of gene expression in plant cells*. *Plant Cell* 1, 671–680.
34. Depicker, A., Stachel, S., Dhaese, P., Zambryski, P., and Goodman, H.M. (1982) *Nopaline synthase, transcript mapping and DNA sequence*. *J. Mol. Appl. Genet.* 1, 561–573.
35. An, G., Mitra, A., Hong, K.C., Costa, M.A., An, K., and Thornburg, R.W., et al. (1989) *Functional analysis of the 3' control region of the potato wound-inducible proteinase inhibitor II gene*. *Plant Cell* 1, 115–122.
36. Mason, H.S., DeWald, D., and Mullet, J.E. (1993) *Identification of a methyl jasmonate-responsive domain in the soybean vspB promoter*. *Plant Cell* 15, 241–251.
37. Liu, D., Burton, S., Glancy, T., Li, Z., Hampton, R., Meade T., and Merlo, D. (2003) *Insect resistance conferred by 283-kD Photorhabdus luminescens protein TcdA in Arabidopsis thaliana*. *Nat. Biotechnol.* 21, 1222–1228.
38. Vigers, A.J., Roberts, W.K., and Selitrennikoff, C.P. (1991) *A new family of plant anti-fungal proteins*. *Mol. Plant. Microbe Interact.* 4, 315–323.
39. Kozak, M. (1986) *Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes*. *Cell* 44, 283–292.
40. Joshi, C.P. (1987) *Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis*. *Nucleic Acids Res.* 15, 9627–9640.
41. Lutcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F., and Scheele, G.A. (1987) *Selection of AUG initiation codon differs in plants and animals*. *EMBO J.* 6, 43–48.
42. Lueharsen, K.R., and Walbot, V. (1994) *The impact of AUG start codon context on maize gene expression in vivo*. *Plant Cell Rep.* 13, 454–458.
43. Vaeck, M., Reynaerts, A., Hofte, H., Jansens, S., De Beuckeleer, M., Dean, C., et al. (1987) *Transgenic plants protected from insect attack*. *Nature* 328, 33–37.
44. Diehn, S.H., Chiu, W-L., De Rocher, E.J., and Green, P.J. (1998) *Prematuration polyadenylation at multiple sites within a Bacillus thuringiensis toxin gene-coding region*. *Plant Physiol.* 117, 1433–1443.
45. Cambell, C.H., and Gowri, G. (1990) *Codon usage in higher plants, green algae, and cyanobacteria*. *Plant Physiol.* 92, 1–11.
46. Murray, E.E., Lotzer, J., and Eberle, M. (1989) *Codon usage in plants*. *Nucleic Acids Res.* 17, 477–498.
47. Dean, C., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., et al. (1986) *mRNA transcripts of several plant genes are polyadenylated at multiple sites in vivo*. *Nucleic Acids Res.* 14, 2229–2240.
48. Shaw, G., and Kamen, R. (1986) *A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation*. *Cell* 46, 659–667.
49. Ohme-Takagi, M., Taylor, C. B., Newman, T. C., and Green, P. (1993) *The effect of sequences with high AU content on mRNA stability in tobacco*. *Proc. Natl Acad. Sci. USA.* 90, 11811–11815.
50. Vankan, P., and Filipowicz, W. (1988) *Structure of U2 snRNA genes of Arabidopsis thaliana and their expression in electroporated protoplast*. *EMBO J.* 7, 791–799.

51. Perlak, F.J., Deaton, R.W., Armstrong, T.A., Fuchs, R.L., Sims, S.R., Greenplate, J.T., and Fischhoff, D.A. (1990) *Insect resistant cotton plants*. *Bio/Technology* 8, 939–943.
52. 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*. *Proc. Natl Acad. Sci. USA* 88, 3324–3328.
53. Adang, M.J., Brody, M.S., Cardineau, G., Eagan, N., Roush, R.T., and Shewmaker, C.K. et-al. (1993) *The reconstruction and expression of a Bacillus thuringiensis cryIIIA gene in protoplast and potato plants*. *Plant Mol. Biol.* 21, 1131–1145.
54. Geiser, M., Schweitzer, S., and Grimm, C. (1986) *The hypervariable region in the genes coding for entomopathogenic crystal proteins of Bacillus thuringiensis: nucleotide sequence of the kurhd1 gene of subsp. Kurstaki HD1*. *Gene* 48, 109–118.
55. Koziel, M.G., Beland, G.L., Bowman, C., Carrozzi, N.B., Crenshaw, R., et al. (1993) *Field performance of elite transgenic maize plants expressing an insecticidal protein derived from Bacillus thuringiensis*. *Bio/Technology* 11, 194–200.
56. Denecke, J., Botterman, J., and Deblaere, R. (1990) *Protein secretion in plant cells can occur via a default pathway*. *Plant Cell* 2, 531–550.
57. Munro, S., and Pelham, H.R.B. (1987) *A C-terminal signal prevents secretion of luminal ER proteins*. *Cell* 48, 899–907.
58. Napier, R.M., Fowke, L.C., Hawes, C., Lewis, M., and Pelham, H.R.B. (1992) *Immunological evidence that plants use HDEL and KDEL for targeting proteins to the endoplasmic reticulum*. *J. Cell Sci.* 102, 261–271.
59. Schouten, A., Roosien, J., van Engelen, F.A., de Jong, G.A.M., Borst-Vrens, A.W.M. et-al. (1996) *The C-terminal KDEL sequence increases the expression of a single-chain antibody designed to be targeted to both the cytosol and the secretory pathway in transgenic tobacco*. *Plant Mol. Biol.* 30, 781–793.
60. Wandelt, C.I., Khan, M.R.I., Craig, S., Schroeder, H.H., Spencer, D., and Higgins, T.J.V. (1992) *Vicilin with carboxy-terminal KDEL is retained in the endoplasmic reticulum and accumulates to high levels in the leaves of transgenic plants*. *Plant J.* 2, 181–192.
61. Yang, S.H., Moran, D.L., Jia, H.W., Bicar, E.H., Lee, M., and Scott, M.P. (2002) *Expression of a synthetic porcine alpha-lactalbumin gene in the kernels of transgenic maize*. *Transgenic Res.* 11, 11–20.
62. Breyne, P., Van Montagu, M., Depicker, A., and Gheysen, G. (1992) *Characterization of a plant scaffold attachment region in a DNA fragment that normalizes transgene expression in tobacco*. *Plant Cell* 4, 463–471.
63. Mlynarova, L., Loonen, A., Heldens, J., Jansen, R.C., Keizer, P., Stiekema, W.J., and Nap, J.P. (1994) *Reduced position effect in mature transgenic plants conferred by the chicken lysozyme matrix-associated region*. *Plant Cell* 6(3), 417–426.
64. Mlynarova, L., Jansen, R.C., Conner, A.J., Stiekema, W.J., and Nap, J.P. (1995) *The MAR-mediated reduction in position effect can be uncoupled from copy number-dependent expression in transgenic plants*. *Plant Cell* 7(5), 599–609.
65. Oh, S., Jeong, J., Kim, E., Yin, N., Jang, I., et al. (2005) *Matrix attachment region from chicken lysosome locus reduces variability in transgene expression and confer copy number-dependent in transgenic rice plants*. *Plant Cell Rep.* 24, 145–154.
66. Xue, H., Yang, Y.T., Wu, C.A., Yang, G.D., Zhang, M.M., and Zheng, C.C. (2005) *TM2, a novel strong matrix attachment region isolated from tobacco, increases transgene expression in transgenic rice calli and plants*. *Theor. Appl. Genet.* 110, 620–627.
67. Mankin, S., Allen, G., Phelam, A., Spiker, S., and Tompson, W. F. (2003) *Elevation of transgene expression level by flanking matrix attachment region (MAR) is promoter dependent: a study of the interactions of six promoters with the RB7 3' MAR*. *Transgenic Res.* 12, 3–12.
68. Sidorenko, L., Bruce, W., Maddock, S., Tagliani, L., Li, X., Daniels, M., and Peterson, T. (2003) *Functional analysis of two matrix attachment region (MAR) elements in transgenic maize plants*. *Transgenic Res.* 12, 137–154.
69. Torney, F., Partier, A., Saya-lesage, V., Nadaud, I., Barret, P., and Beckert, M. (2004) *Heritable transgene expression pattern imposed onto maize ubiquitin promoter by maize adh-1 matrix attachment regions: tissue and developmental specificity in maize transgenic plants*. *Plant Cell Rep.* 22, 931–938.
70. Goodall, G., and Filipowicz, W. (1989) *The AU-rich sequence present in the introns of plant nuclear pre-mRNA are required for splicing*. *Cell* 58, 473–483.