Plant Molecular Biology **32:** 393–405, 1996.
© 1996 *Kluwer Academic Publishers. Printed in Belgium.*

Optimizing expression of transgenes with an emphasis on post-transcriptional events

Michael G. Koziel*, Nadine B. Carozzi & Nalini Desai
*Ciba Agricultural Biotechnology Unit, 3054 Cornwallis Road, Research Triangle Park, NC 27709, USA (*author for correspondence)*

Key words: gene expression, endotoxin, untranslated leader, intron, splicing, synthetic genes

Contents

Abstract 393
Introduction 393
5′-Untranslated leaders 394
Bacterial genes expressed in plants 395
The use of introns to increase gene expression 396
Synthetic genes 398
Chloroplast sequestering and targeting 401
3′ regions 402
Effects of protein folding 402
Conclusions 402
References 403

Abstract

Introducing a foreign gene into a new plant host does not always result in a high level of expression of the incoming gene. Numerous promoters have been used to express foreign genes in different plant tissues, but there are sometime various features of the new gene which are deleterious to expression in the new host. There are a number of post-transcriptional steps in the expression of a gene and sometimes sequences present in a particular coding region can resemble the signals which initiate these processing steps. When aberrantly carried out, these steps diminish the level of expression. By removing such fortuitous signals, one can dramatically increase expression of a transgene in plants. Ensuring proper protein folding and/or targeting the protein product to a particular cellular compartment can also be used to increase the level of protein obtained. The various methods used to optimize expression of a foreign gene in plants by concentrating on post-transcriptional events are discussed.

Introduction

The introduction of foreign genes into a variety of plant species is becoming increasingly routine. However, consistent levels of expression of the transgenes is not a foregone conclusion of a successful transformation process. Certain of the 'rules' for obtaining good expression are becoming clear. For instance choice of a proper 5′-untranslated leader or use of an intron in a gene that lacks its own intron, such as a bacterial gene or a cDNA can increase the level of expression obtained from a given gene. Nevertheless, in addition to large variations in expression levels due to the positional insertion site, a large number of potential problems with a deleterious impact on gene expression exist, especially when using a gene from a heterologous source. Improper splicing, improper polyadenylation, improper nuclear transport, or instability of the resulting cytosolic mRNA can result in accumulation of only a low level of both mRNA and the resulting protein. Additionally, the form of a particular protein chosen to be expressed may play an important role in the steady state level of the protein which is obtained. For instance, if a protein requires a stepwise folding with

subsequent processing of the properly folded form to produce a mature active protein, attempting to express only the mature protein may result in a protein which is not properly folded and is either inactive, unstable and rapidly degraded, or both. In this chapter we will review what is known about optimizing expression of transgenes by focusing on post-transcriptional events. Typically, optimization of expression means obtaining a high level of expression. Inducible, developmental, and tissue specific patterns of expression are obtained by using specific promoters with the desired pattern of expression and will not be dealt with here.

5′-Untranslated leaders

The optimization of transgene expression in plants must include signals for efficient initiation of protein translation. The 5′-untranslated leader (5′-UTL) sequence of eukaryotic mRNA plays a major role in translational efficiency. Its role is presumed to influence the efficiency with which the bound 40S ribosomal subunits migrate and recognize the translational start site. The design of chimeric promoter gene fusions, particularly those which involve a fusion between a plant promoter and a bacterial coding region, allow for great latitude in the composition of 5′-UTL sequences. Many early chimeric transgenes expressed using the cauliflower mosaic virus (CaMV) 35S promoter used an arbitrarily chosen length of viral sequence after the 35S start of transcription and fused this artificially defined 5′-UTL to the AUG of the coding region. Several studies have now shown that the 5′-UTL sequence and sequences directly surrounding the AUG can have a large effect in translational efficiency in plants and that this effect can be different depending on the plant. Although a great deal remains to be understood in the role of the 5′-UTL it is helpful to be aware of the importance of this region.

In most eukaryotic mRNAs, translational initiation occurs at the AUG codon closest to the 5′ cap of the mRNA transcript. Mutagenesis studies of specific mRNAs have demonstrated that the sequences surrounding the AUG of the 5′-untranslated leader are involved in translational efficiency. Comparison of vertebrate mRNA sequences and the results of site-directed mutagenesis experiments have demonstrated the existence of a preferred nucleotide context surrounding the initiation codon [42, 43]. Kozak has defined an optimal AUG context for vertebrates as GCC(A/G)CCAUGG. 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 in that position. The −3 position appears to mediate the efficiency of initiation of translation in animal systems, allowing about a five-fold higher rate of translation with an A or purine than with a pyrimidine [42]. A consensus sequence can also be found among plant genes, UAAACAAUGGCU [40, 47]. In position +4 the preference for a guanine is significantly greater in plants (85%) compared with animals (38%). The preference for G in position +4 (85%) and C at +5 (77%) in plants resembles the preference for A in position −3 (80%) in animals [47]. In plant systems, the −3 position does not appear to effect translational efficiency and the sequence requirements do not appear to be as stringent as in vertebrates. A compilation of sequences surrounding the AUG from 85 maize genes yields a consensus of (C/G)AUGGCG [46]. In contrast to mammalian translation, there is no apparent advantage to having an A at −3 in maize. *β-glucuronidase* (GUS) gene expression was enhanced 4-fold in tobacco protoplasts when the start codon was ACCAUGG rather than UCCUAUGG [22]. An 8-fold increase in expression of bacterial chitinase was observed when ACCAUGG was used in place of the native UUUAUGG [39]. In these instances the U at −1 may be the problem since this is rare in both vertebrates and plants.

The construction of chimeric genes, particularly bacterial genes, has included the use of 5′-UTL sequences from plant viruses. Chimeric mRNAs containing plant viral RNA leaders have been shown to act as efficient enhancers of translation in several systems. Plant viral mRNAs, in particular those encoding the coat protein, are efficiently translated and have UTLs which must compete effectively with plant cellular mRNAs for available ribosomes. The alfalfa mosaic virus (AMV) coat protein and brome mosaic virus (BMV) coat protein UTLs have been shown to enhance mRNA translation 8-fold in electroporated tobacco protoplasts [22]. A 67 nucleotide derivative (Ω) of the 5′-UTL of tobacco mosaic virus RNA (TMV) fused to the chloramphenicol acetyltransferase (CAT) gene and GUS gene has been shown to enhance translation of the reporter genes *in vitro* [22, 69, 70]. Electroporation of tobacco mesophyll protoplasts with transcripts containing the TMV leader fused to reporter genes CAT, GUS and LUC produced a 33-, 21-, and 36-fold level of enhancement, respectively [21, 25]. At least in the case of the TMXV Ω 5′-UTL sequences, the effect was independent of the coding region. An 83 nt 5′-UTL of

potato virus X RNA was shown to enhance expression of the neomycin phosphotransferase II (NptII) gene in *Nicotiana tabacum* protoplasts 4-fold [65]. The 5′-UTL sequences from other highly efficient viral mRNAs do not always enhance translational efficiency of a chimeric mRNA as shown in studies with the turnip yellow mosaic virus coat protein mRNA and black beetle virus RNA 2 [22].

The effect of a 5′-UTL may be different depending on the plant, particularly between dicots and monocots. The TMV 5′-UTL has been shown to be less effective in maize protoplasts, producing ca. 2.6-fold enhancement [26] as opposed to a 30-fold enhancement in tobacco protoplasts [24]. In maize protoplasts, we have demonstrated that 5′-UTLs from the CaMV 35S transcript, both a 73 nucleotide leader [67] and a longer hybrid between the CaMV 35S and luciferase native 5′-UTL, function to strongly enhance expression of the luciferase reporter gene 35-fold and 18-fold respectively compared to the native luciferase 5′-UTL. The 5′-untranslated leaders from the maize genes glutelin [6], PEP-carboxylase [35] and ribulose bisphosphate carboxylase increased expression in maize protoplasts 12-fold, 3.7-fold, and 3.4-fold respectively. The 5′-UTLs from TMV-Ω [23] and three plant viral coat protein genes AMV-coat [27, 38], TMV-coat [30], and BMV-coat [18] worked poorly in maize and inhibited expression of the luciferase gene in maize relative to its native leader.

In tobacco, the effects of these 5′-UTLs is strikingly different, suggesting distinct differences in mechanisms of translational initiation between tobacco and maize. In contrast to maize, the TMV Ω 5′-UTL and the AMV coat protein 5′-UTL enhanced expression in tobacco 5.4-fold and 3.0-fold respectively, whereas the glutelin, maize PEP-carboxylase, and maize ribulose-1,5-bisphosphate carboxylase 5′-UTLs showed no enhancement relative to the native luciferase 5′-UTL. Only the CaMV 35S/Luc hybrid and the CaMV 35S 5′-UTLs enhanced luciferase expression in both maize and tobacco. The TMV and BMV coat protein 5′-UTLs were inhibitory in both maize and tobacco protoplasts. The 5′-UTLs derived from the TMV Ω and AMV coat protein genes produced translational enhancement in tobacco whereas the 5′-UTLs from highly expressed maize cellular genes worked poorly in tobacco. Significant differences exist between maize and tobacco translational systems in their response to the various leaders. The differences in leader activity appear to be a function of the dicot or monocot nature of the cells, and may represent differences in translational protein factors and binding regions.

Many plant 5′-UTLs are considered to be A-rich but this alone does not appear to be significant. Plant genes generally have a high frequency of A and T nucleotides in their 5′-UTL region which may create a relatively unstructured region accessible to the scanning ribosomal 40S subunit [33]. With respect to percent A+T, although TMV omega at 72% and AMV at 70% are the highest, clearly this alone is not critical particularly for maize. A long 5′-UTL does not necessarily confer good activity, as evidenced by the low activity of the TMV Ω 5′-UTL (81 nucleotides) in maize and the poor activities of 5′-UTLs glutelin [6] (87 nucleotides), maize PEP-carboxylase [35] (103 nucleotides), and ribulose-1,5-bisphosphate carboxylase (82 nucleotides) in tobacco.

Bacterial genes expressed in plants

Antibiotic resistance genes of bacterial origin such as neomycin phosphotransferase (NptII), hygromycin phosphotransferase (Hpt) and phosphinothricin acetyltransferase (*bar*) have been used successfully for the selection of transgenic plants. These bacterial genes have been introduced into the monocots rice [4, 12, 68], maize [19, 32, 44, 66], and wheat [76], under control of the CaMV 35S promoter. Improvements in bacterial gene expression have been achieved in monocots through the use of strong constitutive promoters such as the rice actin and maize ubiquitin promoters [8, 72, 77]. The G+C content of *bar*, Hpt, and NptII are 68%, 58%, and 59% respectively.

Bias towards high G+C has made these genes good candidates for expression in plants. The β-glucuronidase (GUS) gene encoded by the *uidA* locus of *Escherichia coli* and the luciferase gene encoded by the firefly, *Photinus pyralis*, are two of the most widely used reporter genes in both dicots and monocots. Expression of luciferase (44% G+C) has been demonstrated in maize [19, 41] whereas expression of the GUS gene (52% G+C) has been reported to be lower and more sporadic [32]. Higher levels of GUS expression have been achieved in the monocots rice [80] and wheat [77] using either the rice actin promoter or the maize ubiquitin promoter.

The use of introns to increase gene expression

Including introns in the transcribed portion of a gene has been found to increase heterologous gene expression in both animal [5] and plant systems [7, 48, 49, 51, 75]. The mechanism of enhancement produced by introns in improving gene expression is not clear. Not all introns produce a stimulatory effect and the degree of stimulation varies, most probably because the effect depends on many factors. However, the effect appears to be accompanied by an increase in the steady-state levels of mRNA lending support to the hypothesis that introns somehow improve efficiency of mRNA processing. In plants the enhancing effect of introns is more apparent in monocots than in dicots. Tanaka *et al.* [71] reported that the presence of the catalase intron 1 isolated from castor beans resulted in an increase in gene expression in rice but not in tobacco when using GUS as a marker gene. Working with a transient expression system in *Arabidopsis* leaf Norris *et al.* [55] found the polyubiquitin intron to have little effect on expression of luciferase using either a CaMV 35S or ubiquitin promoter. Genschick *et al.* [29] found that in a transient expression system derived from tobacco the presence of an intron in the 5′-UTR of the ubiquitin promoter had no effect on expression of the GUS gene.

The first report that the presence of introns could enhance gene expression in transformed monocot cells used the maize alcohol dehydrogenase1 (*Adh1*) gene [7, 13]. When a genomic clone of the *Adh1* gene with its endogenous promoter was used to transform maize cells, the expression of the Adh-s protein was about 100-fold higher than if the coding region came from cDNA clones. To study how the various introns in the *Adh* gene affected expression, regions of the genomic clone were replaced with corresponding regions from the cDNA clone. The level of expression obtained with the genomic clone was restored fully when the sequence downstream of the first intron was replaced by the cDNA sequence. Addition of more genomic sequences did not produce a further increase in expression. In the absence of intron 1, inclusion of introns 8 and 9 restored only 1/3 of the original activity of the genomic clone. This led the authors to conclude that the first intron of *Adh1* was capable of producing up to a 100-fold increase in gene expression. Expression of the *NptII* gene driven by the CaMV 35S promoter was also enhanced 14.6-fold when the region containing the first exon and the first intron of the *Adh1* gene was included in the construct, demonstrating that the stimulatory effect could be observed with a heterologous gene.

The degree to which gene expression is affected in a heterologous system is variable. An important factor appears to be the strength of the promoter being used in the expression cassette [7]. A chimeric gene consisting of an enhanced version of the 35S promoter driving expression of the genomic *Adh1* coding sequences resulted in a 6-fold higher expression compared to that obtained with the native Adh promoter. With the *Adh* cDNA and the CaMV 35S promoter expression was about 500 fold lower then when using the genomic clone [7]. Addition of the first intron from the *Adh1* gene to this construct increased the expression to 1/5 of the level obtained when the complete genomic clone was used. When a shorter 5′ leader was used in the CaMV 35S promoter cassette, there was no apparent loss of expression using the genomic clone but replacing the region between the end of intron 1 and the beginning of intron 8 with cDNA resulted in a two fold increase in expression over that obtained with the complete genomic clone.

Luehrsen and Walbot [45] found that addition of the maize *Adh1* intron 1 at the 5′ end of luciferase or GUS reporter genes driven by Adh1 promoter stimulated expression 2–4-fold. In the presence of a longer leader than that used in the above experiments and the CaMV 35S promoter, expression was reduced to 0.6× of the control construct without introns, mainly due to a significantly higher level of expression from the CaMV 35S promoter and leader combination. Mascarenhas *et al.* [49] found that when using an 'improved' version of the CaMV 35S promoter to drive CAT expression in BMS cells, adding the *Adh1* intron 1 increased expression only 1–2-fold. They reported that under the same conditions *Adh1* introns 2 and 6 produced a 12- and 20-fold increase in expression, respectively. They concluded from their work and results of others [7, 45] that the degree of enhancement by a particular intron depends on the strength of the promoter. This is primarily reflected in the effect of the leader and presence of enhancer elements. Indeed, this appears to be supported by the results of Tanaka *et al.* [71]. Using the catalase intron in an expression cassette containing the CaMV 35S promoter found in pBI221 [37] to drive GUS expression in transformed rice protoplasts, they observed a 80–90-fold increase in expression compared with a construct lacking introns.

Not all introns appear to be capable of increasing gene expression. However, introns other than the first intron of Adh1 have been shown to have a favorable

effect. Callis *et al.* [7] demonstrated that intron 1 of the maize bronze1 gene had a stimulatory effect on CAT expression in maize BMS cells using a 35S promoter and a nos terminator. Enhanced expression has also been reported in protoplasts of bread wheat using the CaMV 35S promoter and the *Adh1* intron 6 [56] and in maize using the *Adh1* intron 2 and intron 6 driven by an enhanced 35S promoter [49]. Vasil *et al.* [75] reported that the presence of the maize sucrose synthase intron 1 provides a 10-fold enhancement of CAT expression over that obtained by the *Adh1* intron 1 in a maize transient expression assay. Intron 3 of the rice actin gene can stimulate reporter gene expression 2- to 6-fold although the overall level of expression was much lower than under the same conditions using the *Adh1* 5′ region [45]. Plant transformation vectors utilizing the rice actin promoter [51] and the maize ubiquitin promoter [11] have retained the first intron in the 5′-UTR when optimizing for expression on monocots. The presence of these introns is required for expression.

Monocots and dicots have been shown to have different requirements for intron recognition and efficient splicing [31]. According to Goodall and Filipowicz, maize protoplasts are more efficient at processing introns with secondary structure and therefore have less of a dependence on AU-rich sequences which, in addition to other effects, are presumed to lower intron secondary structure (see also chapter by Simpson and Filipowicz in this volume).

Gallie and Young [26], working with aleurone and endosperm protoplasts found that the presence of the Adh1 intron1 within a transcript enhances the expression of GUS but the degree of stimulation may vary due to cell type. Mascarenhas *et al.* [49] found that the expression enhancing activity of an intron was affected by the length of the flanking exon sequences. With *Adh1* intron 6 as the stimulatory element, inclusion of 56 nucleotides of the 5′- and 6 nucleotides of the 3′-flanking region in the construct, gave the greatest enhancement. In the case of the *Adh1* intron 6 the largest effect was seen when the construct contained 76 nucleotides of the 5′ and 53 nt of the 3′ sequences flanking the intron. The combination of the maize sucrose synthase exon 1 and intron 1 gave a 1000-fold increase in CAT activity using the CaMV 35S promoter for expression in maize and rice protoplasts [48]. Taken together, these data indicate that the degree of enhancement of gene expression by a particular intron depends on the strength of the promoter, cell type, and flanking exon sequences.

Although the quantitative effect of enhancement due to the presence of introns depends on many factors, there appears to be a requirement for the placement of the intron in the 5′ transcriptional unit in the correct orientation with respect to the splice junction sequences [7, 48, 49, 56, 71, 75]. Introns placed at the 5′ end of the promoter are not active suggesting that the introns do not contain sequences that act as transcriptional enhancers [7, 48, 49, 56, 71, 75]. A consensus has emerged that splicing per se is important to increase stability of the mRNA resulting in a subsequent enhancement of gene expression. [45, 49] This view does not explain the observation that not all introns increase expression in heterologous systems and even the ones that do seem to increase expression do not show uniformity in effect. Results of experiments to correlate splicing efficiency with expression enhancing activity are ambiguous as complete splicing of the transgene is not observed in transient assays [45]. However, the requirement for the presence of sequences flanking the intron for activity suggest that splicing is probably necessary for the enhancement of gene expression [49]. The placement of a particular intron may also be important. For example, when the *Adh1* intron 9 was placed at the 5′ end of a GUS gene expressed using the CaMV 35S promoter, no increase in expression was observed [49]. However, its inclusion in at the 3′ end of the GUS gene increased expression about 3-fold [7].

Of the introns found to positively influence expression four, rice actin intron1, maize ubiquitin intron1, maize bronze1 intron1 and maize sucrose synthase intron1, are present naturally in the 5′-untranslated regions. Perhaps the presence of the intron in the 5′-untranslated region increases the length and strength of the untranslated leader in the resulting processed mRNA. This speculation is supported by the observation that in studies with these introns the highest enhancement has been found when using the native promoter and the first exon [7, 48]. When stronger promoters or those with a strong 5′-untranslated leader were used the degree of enhancement of the *Adh1* intron 1 drops from 100-fold to 2–6-fold [45, 49, 75]. The stimulatory effect resulting from inclusion of maize *Adh1* intron 2, maize *Adh1* intron 6 and the rice actin intron 3 do not fit this speculation unless one assumes that these introns have sequence motifs that enhance gene expression in some unknown manner which has not yet been supported by other evidence.

When optimizing a gene for expression, the role of introns is best evaluated on a case-by-case basis since the size of increase in expression depends on many

factors. In most cases the use of a promoter with a strong 5′-untranslated leader would appear to be more effective in increasing expression than simply including an intron with a weaker promoter. This phenomenon should be studied further in transgenic plants before definitive conclusions can be made about the impact of an intron on increasing levels of transgene expression. Other factors such as the nucleotide composition of the coding region and the 3′-untranslated region can have significant effects on gene expression. For example, the effect of the nucleotide composition of the coding region can far outweigh any enhancement achieved by the inclusion of an intron and/or a strong promoter if one is dealing with a A+T-rich coding region.

Synthetic genes

Heterologous genes may contain fortuitous processing and/or instability signals that have a deleterious impact on gene expression. This is especially true when introducing a prokaryotic gene into a eukaryotic host. The large differences in the gene expression mechanisms between prokaryotes and eukaryotes provides the opportunity for prokaryotic coding regions to contain eukaryotic processing signals which block or diminish expression in a eukaryotic host. Precise removal or alteration of processing signals can be difficult due to the imprecise nature of these signals and the affects of the surrounding nucleotides. It can be difficult to make a few minor, but precise, changes which result in a large increase in expression of the desired gene. The minor changes may each contribute only a minor improvement in expression. Rather than carrying out multiple rounds of mutagenesis to change several scattered sites, it may be easier to construct a synthetic gene of a desired sequence. Synthesis of a gene provides the ability to remove all potentially deleterious processing signals and produce a gene which may more nearly resemble genes of the new host than does the heterologous coding sequence. The greatest amount of experience using synthetic genes to increase expression of a desired protein is derived from expressing the insecticidal proteins, δ-endotoxins, from *Bacillus thuringiensis* (Bt). Attempts to express these genes in plants over several years has provided a basis for comparing levels of expression obtained from modified and unmodified coding regions in a number of plant hosts.

The first attempts to express δ-endotoxins in plants used the native coding sequences derived from Bt. Vaeck *et al.* [73] used the mannopine synthetase promoter and 3′ polyadenylation region of T-DNA gene 7 to express the *cryIA(b)* δ-endotoxin gene. Constructs examined in tobacco included a NH_2-terminal fragment encoding 610 amino acids and two constructs encoding translational fusions between the NH_2-terminal fragment of CryIA(b) and the *NptII* gene. Levels of Bt δ-endotoxin expressed in the plants containing the truncated proteins or the fusion proteins ranged from 2.6–190 ng CryIA(b)/mg soluble protein, or 0.0002–0.02% of total soluble protein. Barton *et al.* [2] analyzed expression of the native *cryIA(a)* gene in tobacco using both a full-length gene and a truncated gene encoding a 644 amino acid protein. Both versions used the CaMV 35S promoter, the 5′-UTL of alfalfa mosaic virus RNA 4, and the nopaline synthase 3′ polyadenlyation region. No plants transformed with the full-length gene produced detectable levels of Bt protein or mRNA. Northern blot analysis of these plants showed mRNA species shorter than the expected full-length transcript with distinct shorter fragments. The existence of incomplete mRNA in transgenic plants was attributed to inefficient post-transcriptional processing or rapid turnover of the full-length transcript.

Fischhoff *et al.* [15] transformed tomato with a native *cryIA(b)* gene. Two truncated versions of the gene were used, one encoding a protein of 646 amino acids and the other a protein of 725 amino acids. Each version was driven by the CaMV 35S promoter and the 3′ polyadenylation region from the nopaline synthetase gene. Expression of only the NH_2-terminal fragment of a δ-endotoxin gene produced plants with insecticidal activity. Likewise a native *cryIA(c)-NptII* fusion in potato showed little insecticidal activity [10], indicating a low level of expression [14]. An active *cryIIIA* gene starting at amino acid 48 was introduced into tomato and potato plants [58]. The expression levels in these plants were very low, less than 0.001% of total soluble protein. In 1989, Vaeck *et al.* published results of studies on the transformation of tomato and potato plants with the NH_2-terminal Bt *cryIA(b)* gene [74]. Transgenic tomato plants produced 60–80 ng CryIA(b) per gram leaf tissue and potato plants produced 90–150 ng CryIA(b) per gram leaf tissue. Expression of a native *cryIA(b)* endotoxin gene in field grown tobacco was characterized by Carozzi *et al.* [9]. Six transgenic tobacco lines, both homozygous and hemizygous, expressing a 645 amino acid CryIA(b) protein from a truncated native gene present under control of the CaMV 35S promoter and 35S 3′ polyadenylation sequences were studied. Bt δ-endotoxin

levels increased throughout the course of plant development, with a substantial increase at the time of flowering. CryIA(b) levels at flowering ranged from 400 to 1000 ng per gram fresh weight or up to 0.01% of the total soluble protein. Bt δ-endotoxin mRNA of the expected size was readily detected, but there were also distinct truncated RNA forms, perhaps resulting from incomplete transcripts or cleavage products.

In all the above reports, native truncated Bt genes could be expressed in dicots but only at low levels. Analysis of mRNA from the native genes showed transcripts of less than the expected size. There are no reports of expression of native Bt genes in monocots. While it is difficult to precisely compare the expression levels reported by the various groups because different promoters and 3′ ends were used, the observation of poor expression is consistent. Various explanations were put forth to explain this observation, but none was readily proven. To study mRNA stability, Murray *et al.* [53] examined expression of *cryIA(b)*, *cryIA(c)*, and *cryIIIA* genes in both transgenic tobacco and electroporated carrot protoplasts. The study examined full-length and truncated *cryIA(b)* and *cryIA(c)* genes and a full-length *cryIIIA* gene. The *cryIA(b)* gene was under control of the CaMV 35S promoter and 3′ polyadenylation sequences from ORF 26 of the TR-DNA. The *cryIA(c)* gene was under control of the mannopine synthetase promoter. Northern analysis of mRNA isolated from transgenic tobacco plants transformed with either full-length or truncated genes from both *cryIA(b)* and *cryIA(c)* showed only truncated transcripts. Northern analysis of *cryIA(b)* and *cryIA(c)* constructs electroporated into carrot protoplasts showed full-length undegraded forms during the first 8 h after electroporation; but by 18 h the *cryIA* mRNA was degraded.

These results suggested that the truncated transcripts observed in plants were the result of message instability rather than truncation due to improper processing. A series of 3′ deletion constructs in electroporated carrot cells showed that deletion of sequences in the 3′ end of the *cryIA(b)* gene did not increase message stability. mRNA instability was retained in the first 570 bases of the gene. The *cryIIIA* gene was likewise poorly expressed in electroporated carrot cells.

Codon usage in the native δ-endotoxin genes is considerably different from that found in typical plant genes, which have a higher G+C content. Native endotoxin genes tend to have a very low G+C content, around 37%. Plants genes in general tend to have a higher G+C content, with maize showing a strong preference for G+C-rich coding regions [52]. Truncated δ-endotoxin gene transcripts in transgenic plants could result from a number of events relating to their high A+T content. These include premature transcriptional termination or polyadenlyation in regions of high A+T content or inappropriate splicing or cleavage. Instability of the mRNA could be the result of endonucleolytic or exonucleolytic degradation at specific sequences that destabilize the message during transcription or create pausing due to the formation of secondary structures. Instability of the mRNA could also be the result of inefficient translation due to poor codon usage. To solve these problems and increase expression of δ-endotoxins in plants, synthetic genes were constructed using different strategies which emphasized eliminating different problems. These strategies all alter the overall G+C content of the genes. Several laboratories have now made partially or completely modified Bt δ-endotoxin genes that have resulted in significant improvements in expression of endotoxins in cotton [59, 78], tomato [60], and potato [61]. Synthetic Bt genes are a requirement to obtain expression of δ-endotoxins in monocots such as maize [44].

Transgenic cotton expressing truncated forms of two Bt endotoxin genes including native and modified sequences showed the efficacy of modified sequences. In 1990, Perlak *et al.* [59] published reports on the performance of transgenic cotton plants, *Gossypium hirsutum* cv. Coker 312, expressing truncated forms of two Bt δ-endotoxin genes. Constructs included genes expressing amino acids 1–612 of the *cryIA(b)* and amino acids 1–640 or 1–615 of the *cryIA(c)* gene. Both genes used a CaMV 35S promoter containing a duplicated enhancer region. Regions targeted for sequence modification included those with potential roles as regulatory sequences, or sequences with predicted mRNA secondary structure. These modifications increased the levels of both CryIA(b) and CryIA(c) δ-endotoxins to 0.05–0.1% of the total soluble protein, or what the authors estimated to be a 100-fold increase in expression compared to the truncated wild-type gene.

Perlak *et al.* [60] examined several versions of modified *cryIA(b)* and *cryIA(c)* genes in both transgenic tobacco and tomato to further analyze the increased expression associated with various sequence modifications. All genes were under control of a CaMV 35S promoter with a duplicated enhancer region. Two types of modified genes were used. One was partially modified, while the second was fully modified. The partially modified *cryIA(b)* gene had 62 of 1743 bases changed to eliminate regions with potential polyadenylation signals and A+T-rich regions. The partially modified

cryIA(b) gene had 97% homology with the wild-type gene and a G+C content of 41%, compared to 37% in the native *cryIA(b)* gene. The fully modified *cryIA(b)* gene had 390 of 1845 bases changed to remove all ATTTA sequences and regions of potential mRNA secondary structure and also replaced bacterial codons with plant-preferred codons. The fully modified gene had 79% homology with the wild type gene with a G+C content increased to 49%. The majority of the transgenic tomato and tobacco plants expressing the partially modified gene produced δ-endotoxin at levels of 1–200 ng CryIA(b) per mg total protein. Over 10% of the fully modified *cryIA(c)* and *cryIA(b)* transgenic tomato and tobacco plants expressed between 600–2000 ng CryIA per mg. Compared with the native truncated gene, the most highly expressing transgenic plants containing the partially modified gene and the fully modified gene increased expression 10- and 100-fold, respectively. Levels of expression with the modified *cryIA* genes were now up to 0.2% of the total protein in the best plants. Constructs with different combinations of changes were tested in tobacco to study the effects of particular sequences. Modifications in the 5′ one-third of the *cryIA(b)* gene were sufficient to produce expression levels comparable to the partially modified gene whereas modifications in the 3′ half of the gene had no effect. The 5′ one-third of the *cryIA(b)* gene and nucleotides 246–283, which contain three potential polyadenylation signals, were identified as important, but no single region could be identified as being critical to expression levels. Northern analysis of the modified *cryIA(b)* genes in transgenic plants indicated that, although the levels of mRNA were increased, the mRNA increase was not proportional to the level of protein increase. This observation led the authors. to conclude that the low gene expression originated at the level of protein translation and not at the level of transcription.

Expression levels of CryIIIA were improved by using synthetic *cryIIIA* genes. A modified version of the *cryIIIA* gene was engineered to remove potential polyadenylation sites and A+T-rich regions, resulting in a final G+C content of 49% as compared to 37% in the native gene [61]. Expression of the *cryIIIA* gene in transgenic potatoes was under control of the CaMV 35S promoter with a duplicated enhancer region. Expression levels attained were between 0.002–0.3% of total soluble protein. Adang *et al.* [1] made a synthetic *cryIIIA* gene and examined its expression in carrot and maize protoplasts as well as transgenic potato plants. The codon pattern of this synthetic gene was altered to match the codon pattern of a dicot. The G+C content of the modified *cryIIIA* gene in this study was 45%. Only the synthetic gene, not the native gene, produced the expected size mRNA in electroporated carrot or maize protoplasts. CryIIIA protein was detected in these protoplasts only with the synthetic gene. The levels of CryIIIA obtained in carrot and maize protoplasts was similar, 0.001 to 0.005%, or 10–50 ng/mg total protein. The native gene did not produce CryIIIA. Transgenic potatoes containing the synthetic *cryIIIA* gene were estimated to produce CryIIIA up to 0.025% of total protein.

Barton *et al.* [3] constructed a synthetic *cryIA(a)* gene which resembled plant genes more closely, using a series of constructs with blocks of synthetic DNA representing 50 codons. Each of the constructs was studied to examine the effect of the modified block. Large increases (100-fold) in gene expression were seen with as little as 10% of the peptide coding region modified in the NH_2-terminal region. A modified truncated *cryIA(c)* gene under the control of the *Arabidopsis thaliana* ribulose-1,5-bisphosphate carboxylase small subunit promoter was tested in tobacco [79]. The small subunit promoter with its own 5′-untranslated leader and chloroplast transit peptide provided a 10–20-fold increase in CryIA(c) expression levels compared to the CaMV 35S promoter with a double enhancer. The increase in expression was the result of a combined effect due to both the 5′-untranslated leader and the transit peptide. A similar increase in expression of the CryIA(c) protein was observed when the 5′-untranslated leader and transit peptide were fused behind the CaMV 35S promoter. In this case, optimizing the translational leader improved expression and so did sequestering the final protein to an intracellular compartment.

The first example of a cereal plant expressing a δ-endotoxin used a synthetic gene [44]. A synthetic gene encoding the first 648 amino acids of the 1155 amino acid CryIA(b) protein produced by Bt var. *kurstaki* HD-1 [28] was constructed using the most preferred codon from maize for each amino acid [44, 52]. The synthetic gene had 65% homology with the native gene and a G+C content of 65%, compared to 37% for the native gene. When compared with the native gene, the synthetic gene produced significantly higher levels of CryIA(b) protein in both tobacco and maize. In transgenic tobacco plants, the synthetic gene produced about five times the level of CryIA(b) protein as did the native truncated gene using the CaMV 35S promoter and terminator regions. Transgenic maize lines

expressing this gene were characterized. Plants containing the synthetic gene under control of the maize PEP-carboxylase [35] and pollen-specific promoters produced up to 4000 ng CryIA(b) per mg soluble protein in certain plants. Since native Bt genes are not expressed in maize, the magnitude of increase cannot be calculated, but 1–5 ng of CryIA(b) per mg total protein can be detected so this level of expression could be taken as the baseline. Transformation of rice with a native *cryIA(b)* gene produced no detectable protein so a highly modified *cryIA(b)* gene was designed based on the codon usage of rice genes [20]. The modified *cryIA(b)* gene had 66.6% of the codons changed to produce an overall G+C content of 59.2%. Bioassays of transgenic rice showed 10–50% mortality against the striped stemborer (*Chilo suppresalis*) and 45–55% mortality against the leaffolder (*Cnaphalocrosis medinalis*) indicating expression of the synthetic gene. The success of synthetic *cry* genes in improving gene expression facilitates studies on transcriptional and translational regulation of the *cry* genes, as well as other genes, in plants.

Synthetic genes encoding proteins other than δ-endotoxins have been made and tested for expression in plants. Hightower *et al.* [34] introduced two versions of a gene encoding a cecropin, a protein with anti-bacterial activity, into tobacco. The first synthetic gene used the same nucleotide sequence present in the source organism, the *Cecropia* moth. Cecropin protein was barely detectable in the transgenic plants. A second cecropin gene was synthesized using plant preferred codons, but transgenic tobacco plants showed no increased level of cecropin from this gene. The authors showed that synthetic cecropin protein is rapidly degraded in plant extracts and suggest that the low levels of protein observed with both versions of the cecropin genes was due to rapid protein degradation. In this case, it would appear to be more important to be concerned with the final stability or perhaps cellular location of the protein product than to use different coding regions to optimize expression. Florack *et al.* [17] also tested expression of a cecropin gene in tobacco. They used different versions of cDNA clones. The short version, lacking the amino terminal signal peptide, showed the poorest accumulation of mRNA. The cDNA containing the amino terminal signal peptide showed higher mRNA levels, while a version of the cecropin cDNA fused with a plant signal peptide had the highest levels of mRNA accumulation. In all cases, the cecropin protein could not be detected, apparently due to rapid degradation. It is interesting to note that the protein being produced appears to have an affect on the level of steady-state mRNA, even when the same final protein is being expressed. Addition of protein processing signals to proteins normally processed appears to have an affect on mRNA levels. The reason for this is unclear.

A gene encoding an insect specific scorpion toxin, I_5A, was synthesized and expressed in bacteria, yeast, and plants by Pang *et al.* [57]. Expression in tobacco was driven by the CaMV 35S promoter and the nopaline synthase 3′ end was used for the polyadenylation site. mRNA derived from the synthetic gene was observed in transgenic plants and appeared to be the expected size. The I_5A peptide was purified from the transgenic plants, but it lacked biological activity, presumably due to improper folding.

Chloroplast sequestering and targeting

Another approach for optimizing expression of A+T-rich Bt genes in plants has been to transform the chloroplast of tobacco. McBride *et al.* [50] have reported very high levels of expression of an unmodified Bt *cryIA(c)* gene in tobacco. The tobacco chloroplast genome has a relatively high A+T content. The introduced Bt gene, present at up to about 10 000 copies per cell in the chloroplast genome, is not improperly processed as it is in the nucleus but is expressed at a high level, yielding up to 3–5% of total leaf protein as CryIA(c). The limitation at present for this approach is the availability of a broadly applicable chloroplast transformation system and the result of having the trait transmissible from only the female parent.

Chloroplast targeting has also been used to increase the expression of polyhydroxybutyrate (PHB) in plants. PHB is the product of a three enzyme pathway starting with acetyl-CoA. When these three proteins were expressed in the cytoplasm of *Arabidopsis*, PHB accumulated at a low level, 20–100 μg per gram fresh weight, and the plants were stunted [62–64]. Nawrath *et al.* [54] targeted these enzymes to the chloroplast using a pea chloroplast transit peptide. The resulting transgenic *Arabidopsis* plants accumulated PHB in plastids up to a level of 14% of the dry weight without any apparent deleterious effects on the plant. The targeting of the enzymes to an intracellular compartment with a high flux of acetyl-CoA increased accumulation of PHB about 100-fold. Choice of the proper intracellular compartment was critical in providing the required substrates for the introduced enzymes. Optimization of

product yield in this case did not require modification of the coding region other than to add an appropriate targeting sequence. This underscores the importance of understanding the biochemistry underlying the mechanisms being introduced into a plant to achieve optimal expression of the desired product.

3′ regions

Ingelbrecht *et al.* [36] examined 3′-end regions of transgenes in plants and found a large difference in expression could be obtained in stable plants. Interestingly, this difference was not detected in transient expression assays. Neomycin phosphotransferase II (*NptII*) was used as the reporter gene for this study. Chimeric gene expression was driven by the CaMV 35S promoter. A *NptII* gene containing no plant 3′-end sequences could be expressed in transgenic tobacco, but at a level about 12-fold lower, and much more variable, than a comparable construct containing the 3′ end of the octopine synthase (OCS) gene. In transient expression assays, the construct containing the OCS 3′ sequences expressed about 20-fold higher than the construct with no 3′-end sequences. Different 3′ ends were compared in both transient and stable transformants. In transient expression assays, the CaMV 35S promoter/*NptII* gene was expressed at comparable levels regardless of the source of the 3′ end. The different 3′ ends used were obtained form the octopine synthase gene, the 2S seed protein from *Arabidopsis*, the small subunit of *rbcS* from *Arabidopsis*, extensin from carrot, and chalcone synthase from *Antirrhinum*. However, in stable tobacco transformants, there was about a 60-fold difference between the best-expressing construct (small subunit *rbcS* 3′ end) and the lowest-expressing construct (chalcone synthase 3′ end), with the other 3′ ends producing different expression levels between these two extremes. Levels of mRNA corresponded well with the observed differences in protein levels. The authors suggested that the higher level of expression obtained with the small subunit *rbcS* 3′ end was a reflection of the need of the plant to express high levels of the small subunit *rbcS* protein and therefore the mRNA was likely to be more stable than the mRNA encoding proteins present at lower levels. The precise role of 3′-end sequences and how to optimize them for maximal expression in transgenic plants is an area where much remains to be learned.

The 3′-untranslated region has been implicated in determining stability or instability of a mRNA. For a detailed discussion on determinants of mRNA instability and their effect on gene expression in plants, readers should see the review by Green *et al.* in this volume.

Effects of protein folding

The effects of pre- and pro-protein sequences on the level of mature protein obtained in transgenic plants was studied by Florack *et al.* [16]. Small anti-bacterial proteins called hordothionins (HDH) are found in barley. The mature proteins are processed from larger precursors which contain the mature protein, an amino terminal signal peptide (SP), and a carboxy-terminal acidic peptide (AP). Different versions of the mature protein were expressed in transgenic tobacco. The chimeric genes were derived from barley cDNA clones as well as synthetic sequences encoding the mature protein and the signal peptide. Plants containing the synthetic gene encoding only the mature peptide produced the lowest level of mRNA. Addition of sequences encoding the SP or adding both the SP and AP increased the level of mRNA detected. Plants containing the gene encoding only the mature protein did not produce detectable amounts of mature protein even though there was mRNA. Addition of the SP to this synthetic gene yielded plants expressing HDH up to about 0.1% of total protein. Addition of both SP and AP sequences increased maximum expression levels almost ten fold to about 0.7% of total protein. Using either the synthetic or barley cDNA to encode the mature protein did not affect levels of mature protein obtained from the constructs containing the SP and AP sequences. In this case, the coding region itself was not responsible for altering the level of the final protein product. Rather, the initial protein produced had a much more dramatic affect. Attempting to produce only the mature protein produced the poorest results and expressing the entire precursor protein produced the best results. The SP was required to produce detectable HDH, likely because this sequence targets the protein to the endoplasmic reticulum where the needed disulfide bonds can be properly formed. Optimization of HDH peptides requires use of the proper pro- and pre-sequences of the protein rather than changing of the coding region.

Conclusions

There are several methods for increasing expression of a particular gene in a plant host. While several oppor-

tunities exist to increase expression, which ones are likely to produce the best effect need to be determined on a case by case basis. Addition of introns to the transcript, use of a good leader for a particular host plant, removing RNA instability signals, and synthesis of optimized coding regions have all been shown to increase levels of expression. Use of a stable 3′ end can help increase expression. While changes in steady-state RNA levels have proven successful in changing levels of protein obtained, the fate of the encoded protein should also be kept in mind. Using truncated versions containing only the desired activity may omit certain folding events necessary for proper folding, stability and activity. Further, the desired protein might not be stable in the plant cell and be rapidly degraded after expression. In such cases, targeting to intracellular compartments or secretion outside the cell might improve levels of the desired protein. For a given protein, not all of the problems are likely to occur, but at present the best method for determining how to optimize expression lies with empirical observation and trial and error. There are likely further refinements used by cells which remain to be elucidated and put to use in optimizing transgene expression.

References

1. Adang MJ, Brody MS, Cardineau G, Eagan N, Roush RT, Shewmaker CK, Jones A, Oakes JV, McBride KE: The reconstruction and expression of a *Bacillus thuringiensis cryIIIA* gene in protoplasts and potato plants. Plant Mol Biol 21: 1131–1145 (1993).
2. Barton KA, Whiteley HR, Yang N: *Bacillus thuringiensis* delta-endotoxin expressed in transgenic *Nicotiana tabacum* provides resistance to lepidopteran insects. Plant Physiol 85: 1103–1109 (1987).
3. Barton KA, Miller MJ: Production of *Bacillus thuringiensis* insecticidal proteins in plants. In: Kung SD, Wu R (eds) Transgenic Plants, vol 1: Engineering and Utilization, pp. 297–315. Academic Press, New York (1993).
4. Battraw M, Hall TC: Expression of a chimeric neomycin phosphotransferase II gene in first and second generation transgenic rice plants. Plant Sci 86: 191–202 (1992).
5. Borenstein P, McKay J, Morishima Jk, Devarayalu S, Gelinas RE: Regulatory elements in the first intron contribute to transcriptional control of the human a1(I) collagen gene. Proc Natl Acad Sci USA 84: 8869–8873 (1987).
6. Boronat A, Martinez MC, Reina M, Puigdomenech P, Palau J: Isolation and sequencing of a 28 kd glutelin-2 gene from maize: Common elements in the 5′ flanking regions among zein and glutelin genes. Plant Sci 47: 95–102 (1986).
7. Callis J, Fromm M, Walbot V: Introns increase gene expression in cultured maize cells. Genes Devel 1: 1183–1200 (1987).
8. Cao J, Duan X, McElroy D, Wu R: Regeneration of herbicide resistant rice plants following microprojectile-mediated transformation of suspension culture cells. Plant Cell Rep 11: 586–591 (1992).
9. Carozzi NB, Warren GW, Desai N, Jayne SM, Lotstein R, Rice DA, Evola S, Koziel MG: Expression of a chimeric CaMV 35S *Bacillus thuringiensis* insecticidal protein gene in transgenic tobacco. Plant Mol Biol 20: 539–548 (1992).
10. Cheng J, Bolyard MG, Saxena RC, Sticklen MB: Production of insect resistant potato by genetic transformation with a delta-endotoxin gene from *Bacillus thuringiensis* var. *kurstaki*. Plant Sci 81: 83–91 (1992).
11. Christensen AH, Sharrock RA, Quail PH: Maize polyubiquitin genes: Structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. Plant Mol Biol 18: 675–689 (1992).
12. Christou P, Ford T, Kofron M: Production of transgenic rice (*Oryza sativa* L.) plants from agronomically important indica and japonica varieties via electric discharge particle acceleration of exogenous DNA into immature zygotic embryos. Bio/technology 9: 957–962 (1991).
13. Dennis ES, Gerlach WL, Pryor AJ, Bennetzen JL, Inglis A, Llewellyn D, Sachs MM, Ferl RJ, Peackocock WJ: Molecular analysis of the alcohol dehydrogenase (*Adh1*) gene of maize. Nucl Acids Res 12: 3983–4000 (1984).
14. Ebora RV, Ebora MM, Sticklen MB: Transgenic potato expressing the *Bacillus thuringiensis cryIA(c)* gene: effects on the survival and food consumption of *Phthorimea operculella* (Lepidoptera: Gelechiidae) and *Ostrinia nubilalis* (Lepidoptera: Noctuidae). J Econ Entomol 87: 1122–1127 (1994).
15. Fischhoff DA, Bowdish KS, Perlak FJ, Marrone PG, McCormick SM, Niedermeyer JG, Dean DA, Kusano-Kretzmer K, Mayer EJ, Rochester DE, Rogers SG, Fraley RT: Insect tolerant transgenic tomato plants. Bio/technology 5: 807–813 (1987).
16. Florack D, Dirkse WG, Visser B, Heidekamp F, Stiekema W: Expression of biologically active hordothionins in tobacco. Effects of pre- and pro-sequences at the amino and carboxyl termini of the hordothionin precursor on the mature protein expression and sorting. Plant Mol Biol 24: 83–96 (1994).
17. Florack D, Allefs S, Bollen R, Bosch D, Visser B, Stiekema W: Expression of giant silkmoth cecropin B genes in tobacco. Transgenic Res 4: 132–141 (1995).
18. French R, Janda M, Ahlquist P: Bacterial gene inserted in an engineered RNA virus: efficient expression in monocotyledonous plant cells. Science 231: 1294–1297 (1986).
19. Fromm ME, Morrish F, Armstrong C, Williams R, Thomas J, Klein TM: Inheritance and expression of chimeric genes in the progeny of transgenic plants. Bio/technology 8: 833–844 (1990).
20. Fujimoto H, Itoh K, Yamamoto M, Kyozuka J, Shimamoto K: Insect resistant rice generated by introduction of a modified delta-endotoxin gene of *Bacillus thuringiensis*. Bio/technology 11: 1151–1155 (1993).
21. Gallie DR, Sleat DE, Watts JW, Turner PC, Wilson TMA: The 5′-leader sequence of tobacco mosaic virus RNA enhances the expression of foreign gene transcripts *in vitro* and *in vivo*. Nucl Acids Res 15: 3257–3273 (1987).
22. Gallie DR, Sleat DE, Watts JW, Turner PC, Wilson TMA: A comparison of eukaryotic viral 5′-leader sequences as enhancers of mRNA expression *in vivo*. Nucl Acids Res 15: 8693–8711 (1987).
23. Gallie DR, Sleat DE, Turner PC, Wilson TMA: Mutational analysis of the tobacco mosaic virus 5′-leader for altered ability to enhance translation. Nucl Acids Res 16: 883–893 (1988).

24. Gallie DR, Lucas WJ, Walbot V: Visualizing mRNA expression in plant protoplasts: factors influencing efficient mRNA uptake and translation. Plant Cell 1: 301–311 (1989).
25. Gallie DR, Feder JN, Schimke, RT, Walbot V: Post-transcriptional regulation in higher eukaryotes: the role of the reporter gene in controlling expression. Mol Gen Genet 228: 258–264 (1991).
26. Gallie DR, Young TE: The regulation of expression in transformed maize aleurone and endosperm protoplasts. Plant Physiol 106: 929–939 (1994).
27. Gehrke L, Auron PE, Quigley GQ, Rich A, Sonenberg N: 5′-Conformation of capped alfalfa mosaic virus ribonucleic acid 4 may reflect its independence of the cap structure or of cap-binding protein for efficient translation. Biochemistry 22: 5157–5164 (1983).
28. Geiser M, Schweitzer S, Grimm C: 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 (1986).
29. Genschick P, Marbach M, Uze M, Feuerman B, Plesse B, Fleck J: Structure and promoter activity of a stress and developmentally regulated polyubiquitin-encoding gene of *Nicotiana tabacum*. Gene 148: 195–202 (1994).
30. Goelet P, Lomonossoff GP, Butler PJG, Akam ME, Gait MJ, Karn J: Nucleotide sequence of tobacco mosaic virus RNA. Proc Natl Acad Sci USA 79: 5818–5822 (1982).
31. Goodall GJ, Filipowicz W: Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. EMBO J 10: 2635–2644 (1991).
32. Gordon-Kamm WJ, Spencer M, Mangano ML, Adams TR, Daines RJ, Start WG, O'Brien JV, Chambers SA, Adams WR, Willetts NG, Rice TB, Mackey CJ, Krueger RW, Kausch AP, Lemaux PG: Transformation of maize cells and regeneration of fertile transgenic plants. Plant Cell 2: 603–618 (1990).
33. Heidecker G, Messing J: Structural analysis of plant genes. Annu Rev Plant Physiol 37: 439–466 (1986).
34. Hightower R, Baden C, Penzes E, Dunsmuir P: The expression of cecropin peptide in transgenic tobacco does not confer resistance to *Pseudomonas syringae* pv. *tabaci*. Plant Cell Rep 13: 295–299 (1994).
35. Hudspeth RL, Grula JW: Structure and expression of the maize gene encoding the phosphoenolpyruvate carboxylase isozyme involved in C4 photosynthesis. Plant Mol Biol 12: 579–589 (1989).
36. Ingelbrecht LW, Herman LMF, Dekeyser RA, Van Montagu MC, Depicker AG: Different 3′ end regions strongly influence the level of gene expression in plant cells. Plant Cell 1: 671–680 (1989).
37. Jefferson RA, Kavanaugh TA, Bevan MW: GUS Fusions: β-glucuronodase as a sensitive and versatile gene fusion marker in higher plants. EMBO J 6: 3901–3907 (1987).
38. Jobling SA, Gehrke L: Enhanced translation of chimaeric messenger RNAs containing a plant viral untranslated leader sequence. Nature 325: 622–625 (1987).
39. Jones JDG, Dean C, Gidoni, Gilbert D, Bond-Nutter D, Lee R, Bedbrook J, Dunsmuir P: Expression of bacterial chitinase protein in tobacco leaves using two photosynthetic gene promoters. Mol Gen Genet 212: 536–542 (1988).
40. Joshi CP: An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucl Acids Res 15: 6643–6653 (1987).
41. Klein TM, Roth BA, Fromm ME: Regulation of anthocyanin biosynthetic genes introduced into intact maize tissues by microprojectiles. Proc Natl Acad Sci USA 86: 6681–6685 (1989).
42. Kozak M: Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44: 283–292 (1986).
43. Kozak M: An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucl Acids Res 15: 8125–8132 (1987).
44. Koziel MG, Beland GL, Bowman C, Carozzi NB, Crenshaw R, Crossland L, Dawson J, Desai N, Hill M, Kadwell S, Launis K, Lewis K, Maddox D, McPherson K, Meghji MR, Merlin E, Rhodes R, Warren GW, Wright M, Evola SV: Field performance of elite transgenic maize plants expressing an insecticidal protein derived from *Bacillus thuringiensis*. Bio/technology 11: 194–200 (1993).
45. Luehrsen KR, Walbot V: Intron enhancement of gene expression and the splicing efficiency of introns in maize cells. Mol Gen Genet 225: 81–93 (1991).
46. Luehrsen KR, Walbot V: The impact of AUG start codon context on maize gene expression *in vivo*. Plant Cell Rep 13: 454–458 (1994).
47. Lutcke HA, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA: Selection of AUG initiation codons differs in plants and animals. EMBO J 6: 43–48 (1987).
48. Maas C, Laufs J, Grant S, Korfhage C, Werr W: The combination of a novel stimulatory element in the first exon of the maize shrunke-1 gene with the following intron enhances reporter gene expression 1000-fold. Plant Mol Biol 16: 199–207 (1991).
49. Mascerenhas D, Mettler IJ, Pierce DA, Lowe HW: Intron mediated enhancement of heterologous gene expression in maize. Plant Mol Biol 15: 913–920 (1990).
50. McBride KE, Svab Z, Schaaf DJ, Hogan PS, Stalker DM, Maliga P: Amplification of a chimeric *Bacillus* gene in chloroplasts leads to an extraordinary level of an insecticidal protein in tobacco. Bio/technology 13: 362–365 (1995).
51. McElroy D, Zhang W, Wu R: Isolation of an efficient promoter for use in rice transformation. Plant Cell 2: 163–171 (1990).
52. Murray EE, Lotzer J, Eberle M: Codon usage in plants. Nucl Acids Res 17: 477–498 (1989).
53. Murray EE, Rocheleau T, Eberle M, Stock C, Sekar V, Adang M: Analysis of unstable RNA transcripts of insecticidal crystal protein genes of *Bacillus thuringiensis* in transgenic plants and electroporated protoplasts. Plant Mol Biol 16: 1035–1050 (1991).
54. Nawrath C, Poirier Y, Somerville C: Targeting of the polyhydroxybutyrate biosynthetic pathway to the plastids of *Arabidopsis thaliana* results in high levels of polymer accumulation. Proc Natl Acad Sci USA 91: 12760–12764 (1994).
55. Norris SR, Meyer SE, Callis J: The intron of *Arabidopsis thaliana* polyubiquitin genes is conserved in location and is a quantitative determinant of chimeric gene expression. Plant Mol Biol 21: 895–906 (1993).
56. Oard JH, Paige D, Dvorak J: Chimeric gene expression using maize intron in cultured cells of breadwheat. Plant Cell Rep 8: 156–160 (1989).
57. Pang SZ, Oberhaus SM, Rasmussen JL, Knipple DC, Bloomquist JR, Dean DH, Bowman KD, Sanford JC: Expression of a gene encoding a scorpion insectotoxin in yeast, bacteria, and plants. Gene 116: 165–172 (1992).
58. Perlak FJ, McPherson SA, Fuchs RL, Macintosh SC, Dean DA, Fischhoff DA: Expression of *Bacillus thuringiensis* protein in transgenic plants. In: Roberts DW, Granados RR (eds) Proceedings of a Conference on Biotechnology, Biological Pesti-

cides and Novel Plant-Pest Resistance for Insect Pest Management, pp. 77–81. Boyce Thompson Institute for Plant Research, Ithaca, NY (1988).

59. Perlak FJ, Deaton RW, Armstrong TA, Fuchs RL, Sims SR, Greenplate JT, Fischhoff DA: Insect resistant cotton plants. Bio/technology 8: 939–943 (1990).
60. Perlak FJ, Fuchs RL, Dean DA, McPherson SL, Fischhoff DA: Modification of the coding sequence enhances plant expression of insect control protein genes. Proc Natl Acad Sci USA 88: 3324–3328 (1991).
61. Perlak FJ, Stone TB, Muskopf YM, Peterson LJ, Parker GB, McPherson SA, Wyman J, Love S, Reed G, Biever D, Fischhoff DA: Genetically improved potatoes: protection from damage by Colorado potato beetles. Plant Mol Biol 22: 313–321 (1993).
62. Poirier Y, Dennis DE, Klomparens K, Sommerville CR: Polyhydroxybutyrate, a biodegradable thermoplastic, produced in transgenic plants. Science 256: 520–523 (1992).
63. Poirier Y, Dennis DE, Nawrath C, Somerville CR: Progress toward biologically produced biodegradable thermoplastics. Adv Mat 5: 30–37 (1993).
64. Poirier Y, Schechtman LA, Satkowski MM, Noda I, Somerville CR: Synthesis of high-molecular-weight poly([R]-(-)-3-hydroxybutyrate) in transgenic *Arabidopsis thaliana* plant cells. Int J Biol Macromol 17: 7–12 (1995).
65. Poogin MM, Skryabin KG: The 5′ untranslated leader sequence of potato virus X RNA enhances the expression of the heterologous gene *in vivo*. Mol Gen Genet 234: 329–331 (1992).
66. Rhodes CA, Pierce DA, Mettler IJ, Mascarenhas D, Detmer JJ: Genetically transformed maize plants from protoplasts. Science 240: 204–207 (1988).
67. Rothstein SJ, Lahners KN, Lotstein RJ, Carozzi NB, Jayne SM, Rice DA: Promoter cassettes, antibiotic-resistance genes, and vectors for plant transformation. Gene 53: 153–161 (1987).
68. Shimamoto K, Terada R, Izawa T, Fujimoto H: Fertile transgenic rice plants regenerated from transformed protoplasts. Nature 338: 274–276 (1989).
69. Sleat DE, Gallie, DR, Jefferson RA, Bevan MW, Turner PC, Wilson TMA: Characterization of the 5′-leader sequence of tobacco mosaic virus RNA as a general enhancer of translation *in vitro*. Gene 217: 217–225 (1987).
70. Sleat DE, Hull R, Turner PC, Wilson TMA: Studies on the mechaism of translational enhancement by the 5′-leader sequence of tobacco moasaic virus RNA. Eur J Biochem 175: 75–86 (1988).
71. Tanaka A, Mita S, Ohta S, Kyozuka J, Shimamoto K, Nakamura K: Enhancement of foreign gene expression by a dicot intron in rice but not in tobacco is correlated with an increased level of mRNA and an efficient splicing of the intron. Nucl Acids Res 18: 6767–6770 (1990).
72. Toki S, Takamatsu S, Nojiri C, Ooba S, Anzai H, Iwata M, Christensen AH, Quail PH, Uchimiya H: Expression of a maize ubiquitin gene promoter-*bar* chimeric gene in transgenic rice plants. Plant Physiol 100: 1503–1507 (1992).
73. Vaeck M, Reynaerts A, Hofte H, Jansens S, De Beuckeleer M, Dean C, Zabeau M, Van Montagu M, Leemans J: Transgenic plants protected from insect attack. Nature 328: 33–37 (1987).
74. Vaeck M, Reynaerts A, Hofte H: Protein engineering in plants: expression of *Bacillus thuringiensis* insecticidal protein genes. In: Schell J, Vasil IK (eds) Cell Culture and Somatic Cell Genetics of Plants, vol 6. Academic Press, Troy, MO (1989).
75. Vasil V, Clancy M, Ferl RJ, Vasil IK, Hannah LC: Increased gene expression by the first intron of maize *shrunken–1* locus in grass species. Plant Physiol 91: 1575–1579 (1989).
76. Vasil V, Castillo AM, Fromm ME, Vasil IK: Herbicide resistant fertile transgenic wheat plants obtained by microprojectile bombardment of regenerable embryogenic callus. Bio/technology 10: 667–674 (1992).
77. Weeks JT, Anderson OD, Blechl A: Rapid production of multiple independent lines of fertile transgenic wheat (*Triticum aestivum*). Plant Physiol 102: 1077–1084 (1992).
78. Wilson FD, Flint HM, Deaton WR, Fischhoff DA, Perlak FJ, Armstrong TA, Fuchs RL, Berberich SA, Parks NJ, Stapp BR: Resistance of cotton lines containing a *Bacillus thuringiensis* toxin to pink bollworm (Lepidopteran: Gelechiidae) and other insects. J Econ Entomol 4: 1516–1521 (1992).
79. Wong EY, Hironaka CM, Fischoff DA: *Arabidopsis thaliana* small subunit leader and transit peptide enhance the expression of *Bacillus thuringiensis* proteins in transgenic plants. Plant Mol Biol 20: 81–93 (1992).
80. Zhang W, McElroy D, Wu R: Analysis of rice *act1* 5′ region activity in transgenic rice plants. Plant Cell 3: 1155–1165 (1991).