This is to certify that the

dissertation entitled

**B.t. TOXIN GENE EXPRESSION AND
DIFFERENTIAL UTILIZATION OF
POLYADENYLATION SIGNALS IN GRAMINEOUS
AND DICOTYLEDONOUS PLANTS**

presented by

Scott Henry Diehn

has been accepted towards fulfillment
of the requirements for

_____ Ph.D. ___ degree in _Botany and_ Plant Path

_____
Major professor

Date __4/13/98__

PLACE IN RETURN BOX
to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE

1/98 c/CIRC/DateDue.p65-p.14

B.t. TOXIN GENE EXPRESSION AND DIFFERENTIAL UTILIZATION OF POLYADENYLATION SIGNALS IN GRAMINEOUS AND DICOTYLEDONOUS PLANTS

By

Scott Henry Diehn

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Graduate Program in Botany and Plant Pathology

1998

ABSTRACT

B.t. TOXIN GENE EXPRESSION AND DIFFERENTIAL UTILIZATION OF POLYADENYLATION SIGNALS IN GRAMINEOUS AND DICOTYLEDONOUS PLANTS

By

Scott Henry Diehn

The ability to transform plants with foreign genes has been an invaluable tool in plant biology as well as in biotechnology. However, not all genes are expressed at expected levels when introduced into plants. One example is the family of *B.t.* toxin genes from *Bacillus thuringiensis* which encode highly specific insecticidal proteins. The mRNA transcribed from these genes does not accumulate in plant cells even when expressed under the control of strong promoters. Because this problem is not unique to *B.t.* toxin, studying the expression of these genes may enhance our understanding of the mechanisms which may limit the expression of other foreign genes in plants. To this end, a typical *B.t.* toxin gene, the *crylA(c)* gene, was expressed in tobacco cells. Poly(A)$^+$ RNA isolated from these cells showed two short transcripts accumulate in addition to the full-length transcript. Hybridization with different segments of the *crylA(c)* coding region indicated the short transcripts were the result of premature polyadenylation of the *crylA(c)* transcript. RT-PCR and oligo-directed RNaseH cleavage experiments were used to identify and confirm the poly(A) addition sites in the *crylA(c)* coding region. One of the

polyadenylation sites was observed to be more efficiently utilized in maize than in tobacco. This, together with other observations in the literature, indicated a monocot/dicot difference in poly(A) signal utilization and that monocots may have a less stringent requirement for poly(A) signal utilization than dicots. To investigate this possibility, a set of chimeric genes containing the polyadenylation signals from monocot (Gramineae) and dicot plant species were expressed in maize, wheat, and tobacco cells as well as in Arabidopsis seedlings. Maize and wheat efficiently utilized both the monocot (Gramineae) and dicot poly(A) signals. However, tobacco and Arabidopsis could not efficiently utilize all the monocot (Gramineae) polyadenylation signals. These results support a differential utilization of plant polyadenylation signals in monocots (Gramineae) and dicots. In addition, they argue monocots (Gramineae) have a less stringent requirement for poly(A) signal utilization than dicots.

ACKNOWLEDGEMENTS

Although there is only one author on this thesis, this work would not have been

possible without the assistance of many people. Each person contributed in his or her

way by providing intellectual conversation or contributing experimental materials. While

this list of people is quite extensive, I would like to acknowledge those which contributed

the most to my project.

First and most importantly, I would like to acknowledge my advisor, Dr. Pam Green.

I have worked in four different laboratories at three different universities and never have I

seen anyone so concerned about their students and their future. Dr. Green provides each

of her students many unique opportunities that allows them to obtain the skills necessary

for their professional development. It is these opportunities which has allowed me bypass

a post-doctoral position and obtain a permanent position in industry. In addition to the

opportunities, Dr. Green has created an environment in the laboratory which makes it

easy to come to the laboratory every day. She has a knack for attracting highly motivated,

bright individuals to the laboratory which are just as concerned about your research as

their own. I will certainly miss her and the laboratory, but look forward to continued

interactions in the future.

Although every member of the laboratory past and present has contributed to my

research in some shape or form, two individuals which have contributed the most are Dr.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

It has been more than ten years since the first transgenic plants expressing foreign genes were regenerated following *Agrobacterium*-mediated transformation. During this time, significant progress has been made toward crop improvement by genetically engineering plants with a variety of desirable traits. New developments in the area of plant transformation and the isolation of agronomically important genes have contributed to this advancement. In some cases, endogenous plant genes have been re-engineered to produce advantageous characteristics, whereas in others, genes from non-plant sources have been used. Prominent examples include the antisense and sense configurations of several genes that have been used to manipulate ripening and senescence processes (1). In addition, genes from bacteria, fungi, plants, and viruses have been expressed in plants to achieve resistance to various pathogens (2) or herbicides (3), to alter male fertility (4,5) or lipid composition (6), or to produce new compounds in plants such as biodegradable plastics (7,8). As a result, many exciting transgenic plants and their products should soon reach the market place.

Despite the many success stories, there are instances where genetically engineered plants fail to adequately produce the desired gene product, particularly when the gene is introduced from a foreign source. This can occur even when the problematic gene is introduced into plants under the control of strong plant promoters, implying that the coding region or protein itself is responsible for the poor expression. Although most of these difficulties have not been published, the problems associated with one prominent group of genes encoding insecticidal proteins, the *B.t.* toxins, have been well documented.

B.t. toxins, also known as insecticidal crystal proteins (ICPs), *B.t.k.* insect control proteins, δ-endotoxins, crystal proteins, and *cry* gene products, are a family of insecticidal proteins produced by the gram-positive soil bacterium *Bacillus thuringiensis*. They have been divided into six classes based on their insecticidal spectra and structural homologies: CryI (effective against Lepidoptera), CryII (effective against Lepidoptera and Diptera), CryIII (effective against Coleoptera), CryIV (effective against Diptera) , and CryV and CryVI (both effective against nematodes) although the CryV designation has also been used to classify *B.t.* toxins effective against both lepidopteran and coleopteran larvae (9), (see references 10 and 11 for a more detailed description of these classes and additional subclasses). Upon sporulation, *Bacillus thuringiensis* generates crystalline parasporal inclusions which can be composed of one or more protoxin species. When ingested by the larvae of a susceptible insect, the crystalline inclusions are solubilized in the midgut releasing the protoxins. Proteases in the larval midgut cleave the protoxins into their active form. These toxic polypeptides act by binding to specific receptors in the membrane of epithelial cells in the midgut of susceptible insect larvae. There the toxin forms a cation-selective channel that disrupts the K^+ balance in the cell. As a result, water enters the epithelial cell causing it to swell and lyse (10,12).

Spore-crystal mixtures of these toxins have been used for many years as an alternative to conventional chemical pesticides because they are very potent and highly specific. They are not known to be toxic to vertebrates or non-targeted insects (13). However, widespread commercial use of these toxins as insecticidal sprays has been limited, in part due to high production costs and poor persistence in the field. An attractive approach that avoids these problems is the genetic engineering of plants with

B.t. toxin genes. Plants that express *B.t.* toxin genes should be resistant to insect feeding damage throughout the growing season, reducing the need for pesticide applications. Considerable effort has been devoted to achieving this goal, particularly in the private sector, because of its high potential for commercial and ecological benefit.

Unfortunately, the path to success was much more challenging than it was first thought. *B.t.* toxin genes have proven to be the most difficult foreign genes reported to date to express at high levels in plants. As such, the work on *B.t.* toxin genes provides an excellent case study of problematic gene expression in higher plants. An informal poll of several biotechnology companies indicated that this problem is not unique to *B.t.* toxin genes; nearly all reported that a proportion of the foreign genes they had examined were not expressed to expected levels in transgenic plants. At present, the most common approach to increase the expression of a foreign gene is to resynthesize the gene to make it more plant-like, a rather time-consuming and expensive operation. This approach can also be somewhat risky if the mechanisms posing the original limitation are unknown.

In this review, we highlight the research done on *B.t.* toxin gene expression in plants and discuss the specific strategies that have been used successfully and unsuccessfully to circumvent the problems that were encountered. Emphasis will be placed on those studies that address the specific mechanisms that limit *B.t.* toxin gene expression in an effort to glean the greatest insight towards the effective engineering of new *B.t.* toxin genes and other problematic foreign genes for future crop improvement efforts. We will not address other aspects associated with *B.t.* toxin such as classification, structure-function relationships, or the evolution of insecticidal resistance, as these topics have been reviewed elsewhere (10-12,14).

B.t. TOXIN GENE EXPRESSION AND INSECTICIDAL ACTIVITY IN TRANSGENIC PLANTS

Initial Studies with *B.t.* Protoxins in Plants

The first *B.t.* toxin genes to be transformed into plants were members of the *cryIA* class of genes. Initially, the entire protoxin coding region was introduced into tobacco and tomato by *Agrobacterium*-mediated transformation. Unfortunately, regenerated plants displayed little or no insecticidal activity in insect feeding assays. A determination of the *B.t.* toxin levels within these plants showed that very low levels of the protein (2 ng/mg protein) were synthesized (15-17). Higher levels of protoxin (10-50 ng/mg protein) were reportedly produced in tobacco calli, but plants could not be successfully regenerated from these calli. This prompted the suggestion that the protoxin is deleterious to plant cells (18).

Truncated *B.t.* Toxin Genes

In an attempt to circumvent the problems encountered with the intact protoxin genes, truncated *B.t.* toxin genes were engineered for expression in plants. Extensive deletion analyses performed in bacteria (reviewed in 10) and plants (16) have demonstrated that the N-terminal portion of the protoxin is effective in killing susceptible insect larvae. The function of the C-terminal portion is unknown, although in the case of the *cryI* genes, it is thought to be involved in the co-assembly of protoxins into the crystal (10,12).

The truncated genes, in comparison to full-length *B.t.* toxin genes, were expressed at significantly higher levels and protected plants to a much greater degree. Tobacco hornworm (*Manduca sexta*) mortality rates in isolated leaf bioassays generally ranged from 75 to 100% within 6 days, with most transformants above 90% (15,16,18). Up to 14 ng of *B.t.* toxin per mg total protein typically accumulated in these plants (15,18). The degree of truncation also appears to have an impact on the level of expression. Tobacco plants transformed with a gene fusion corresponding to the first 683 amino acids of a *cryIA(b)* gene linked to the kanamycin resistance coding region, *nptII*, exhibited a higher level of insecticidal activity than did plants expressing a similar gene fusion containing the first 724 amino acids (15). These results indicate that *B.t.* toxin genes containing only those sequences necessary for toxicity may provide the highest levels of *B.t.* toxin protein in plants.

Although plants expressing a truncated *B.t.* toxin gene are able to control the tobacco hornworm, they are often not able to manifest this effect against less susceptible insects. The tobacco hornworm is among the insects that are the most sensitive to *B.t.* toxin having a LC_{50} of approximately 0.04 mg/ml for both CryIA(b) and CryIA(c) proteins (19). Other lepidopteran insect pests such as the tobacco budworm (*Heliothis virescens*), *Helicoverpa zea* (corn earworm, cotton bollworm, and tomato fruitworm, formerly known as *Heliothis zea*), and beet armyworm (*Spodoptera exigua*) are increasingly difficult to kill with both CryIA(b) and CryIA(c) proteins in this order, although the sensitivities between the two toxins do vary for some species (19). Thus, transgenic plants exhibiting high levels of resistance to the tobacco hornworm were only partially toxic to the tobacco budworm and even less resistant to *Helicoverpa zea* in laboratory and field tests (16,20).

As a result of these studies, it was apparent that higher levels of *B.t.* toxin expression were still required to protect plants from a great majority of the agronomically important insect pests that are susceptible to *B.t.* toxin. Nonetheless, it was realized that the C-terminal portion was dispensable and most subsequent plant transformation experiments exploited this fact. All *B.t.* toxin genes referred to in the remainder of this review should be assumed to be truncated unless otherwise indicated.

Attempts to enhance or control expression with 5' and 3' flanking sequences. In many cases, *B.t.* toxin genes have been expressed under the control of the strong and relatively constitutive 35S promoter of the Cauliflower Mosaic Virus (35S) (16,18,21-27). Yet, these genes usually gave rise to very low levels of *B.t.* toxin mRNA in transgenic plants. One exception was a 35S-driven *cryIA(b)* gene that, for unknown reasons, produced readily detectable levels of mRNA and correspondingly high quantities of protein (~100 ng *B.t.* toxin/mg protein) in tobacco plants (28). However, efforts to achieve improved levels of insect control continued, in part through the construction of *B.t.* toxin genes with alternative 5' flanking sequences. For example, in one study phaseolin, 19S, 35S, soybean ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) small subunit, and mannopine synthase promoters were each used to control the transcription of either a protoxin or truncated *cryIA(c)* gene in transgenic tobacco (21). Transcripts were detected only in young plants containing the mannopine synthase promoter-driven genes. Interestingly, expression of the genes disappeared with plant maturation (21).

The 35S promoter has been modified for greater transcriptional activity by duplicating the enhancer region (29). The activity of this enhanced 35S promoter is approximately

ten-fold higher than the unmodified version and about 100 times as active as a NOS promoter. An enhanced 35S promoter has been used to direct the expression of a *crylA(c)* gene in tobacco cells (30). *B.t.* toxin transcripts from poly(A)-enriched fractions were observed on northern blots, but they were not observed in total RNA fractions, indicating that this promoter would not suffice as a remedy for the low accumulation of *B.t.* toxin in plants. Thus far, no promoters have been demonstrated to be more effective than an enhanced 35S promoter for achieving high levels of *B.t.* toxin gene expression, although most studies lack comparative analyses.

In other attempts to optimize the expression of *B.t.* toxin genes in plants, the 5' untranslated region (5' UTR) from alfalfa mosaic virus (AMV) RNA 4 was inserted between the promoter and the coding regions of several *cryI* genes (18,25,26,28). AMV RNA 4 is a well-translated transcript encoding the viral coat protein. The 5' UTR of this message has been found to increase the translational efficiency of chimeric mRNAs *in vitro* and *in vivo* (reviewed in 31). Whether these sequences actually increase the accumulation of *B.t.* toxin in plants is unclear. In only one study were the levels of *B.t.* toxin produced with and without the translational enhancer directly compared. Inclusion of the AMV 5' UTR in a *crylA(b)* gene construction did not appear to have an effect; however, only one line expressing the AMV-BT chimeric gene was generated, making it difficult to distinguish between position effects and the effect of the leader sequence (28). Otherwise, the AMV 5' UTR has been included in gene constructions under the assumption that it would benefit expression. Incorporation of the AMV leader into *crylA(b) B.t.* toxin constructs gave rise to *B.t.* toxin levels of 2-12 ng/mg protein in experiments carried out in transgenic tobacco (18) and soybean (25) plants. Two to forty

nanograms of *B.t.* toxin protein per milligram protein was detected in transgenic tomato plants (32), however, only 2-10 ng/mg protein was present in leaf tissues (25), sites where many insect larvae feed. Because none of these accumulation levels are of economic relevance, AMV 5' UTR sequences in combination with strong promoters are not sufficient to overcome the expression problems associated with *B.t.* toxin in plants.

Translational fusions consisting of the *nptII* gene coupled to the 3' end of a truncated *B.t.* toxin gene have been utilized in an attempt to select for transformants expressing high levels of *B.t.* toxin (15). These fusions do not increase the expression of the *B.t.* toxin gene, but increase the possibility of finding transformants with high levels of expression. Plants able to survive on growth media containing high concentrations of kanamycin displayed elevated levels of insect mortality. However, still higher levels of expression were needed to fully protect the plants from less sensitive insects (15).

The 5' and 3' flanking sequences from a construct known to direct the expression of a luciferase reporter gene in the interveinal regions of a leaf upon wounding (33) have also been fused to a *cryIA(c)-nptII* coding region in an attempt to increase *B.t.* toxin levels. Several species of plants have been engineered with this chimeric gene, including some woody perennials (22-24,27). However, very few plants expressed the fusion protein to high enough levels to kill larvae (23,24,27). In only one case were insects effectively controlled, but it is not clear whether or not wound induction of the *B.t.* toxin gene occurs (22).

Some genera of Coleoptera and Diptera feed on the roots and root nodules of economically important plants. These tissues are difficult to protect using conventional methods; however, an obvious approach would be to express an appropriate *B.t.* toxin

gene under the control of a root-specific promoter. To the best of our knowledge this has

not been pursued, possibly due in part to the fact that some of these host plants belong to

the families Leguminosae and Gramineae, historically difficult plants to transform and

regenerate. To overcome this obstacle, root-colonizing bacteria have been engineered to

express a variety of *B.t.* toxin genes (34-37). An advantage to this approach is that the

problems of low expression in plants can be avoided. The accumulation of *B.t.* toxin in

these bacteria was significant, but only modest levels of protection were observed in

inoculated plants (35-37). Thus far, the reasons for this poor protection have not been

elucidated nor overcome. Expression of *B.t.* toxin in these plants may be more practical

in the future owing to recent advances in plant transformation technology.

High expression levels achieved with modified *B.t.* toxin genes. C-terminal

truncations, modification of 5' and 3' UTR's, protein fusions, and the use of strong plant

promoters were all insufficient to overcome the low-level expression of otherwise

"unmodified" *B.t.* toxin genes in plants. Much higher levels of expression have been

achieved with the additional step of constructing "modified" *B.t.* toxin genes in which the

coding sequences of wild-type *B.t.* toxin genes have been extensively altered. Fifteen

modified *B.t.* toxin genes and their expression properties have been reported to date and

are listed in Table 1-1. The expression of modified *B.t.* toxin genes has been examined in

ten monocot and dicot species and has been shown to protect against more than ten types

of insects including relatively insensitive species such as beet armyworm and European

corn borer (*Ostrinia nubilalis*). Increased insecticidal activity and increased mRNA

abundance were observed in all cases where these parameters were compared between

modified and unmodified *B.t.* toxin genes. Similar N-terminal regions of the protoxin genes have been used in most instances but the degree of modification has covered a wide range. Sequence changes have altered from as few as 3% to as many as 85% of the codons, and the GC content has been raised from between 34% and 38% for unmodified genes to between 41% and 65% for modified genes.

The rationale for altering the coding sequence is based on the high AT content of *B.t.* toxin genes relative to plant gene coding sequences. Whereas *B.t.* toxin genes typically have an AT content of approximately 65% (Table 1-1), dicot coding regions are approximately 55% AT and monocot coding regions are approximately 45% AT (38). One consequence of the AT-richness is that *B.t.* toxin coding regions are similar to plant 3' UTRs and introns. These properties increase the probability that *B.t.* toxin transcripts may by chance contain sequences recognized in plants as signals for polyadenylation, mRNA decay, splicing, or other processes that can affect the structure and accumulation of the mRNA. Another consequence of the AT-richness of *B.t.* toxin genes is the frequent occurrence of codons that are rare in plant genes. It has been proposed that these rare codons could slow translation, thereby inhibiting *B.t.* toxin protein synthesis (21).

One approach used for sequence modification has been the specific targeting of sequences suspected to inhibit *B.t.* toxin gene expression in plants. Modifications were made in AT-rich regions, especially those containing potential plant polyadenylation signals (39) or the ATTTA instability motif (40,41). Sequence changes that eliminated these possible deleterious elements were designed so as to convert the codons involved to synonymous codons preferred in plant genes (26,42-45). Thorough application of the criteria for sequence modification resulted in extensive changes, with more than half of

Table 1-1. Characteristics of modified *B.t.* toxin genes expressed in plant cells.

a   Amino acid residues from protoxin genes included in modified *B.t.* toxin genes.

b   + indicates increased mRNA accumulation for the modified *B.t.* toxin gene relative
to the unmodified gene.

c   Highest observed level of *B.t.* toxin protein as a percent of total protein.

d   + indicates increased insecticidal activity resulting from expression of the modified
B.t. toxin gene relative to the unmodified gene.

e   Insecticidal activity observed against the indicated insect.

nr   Not reported.

<<   Below limit of detection.

(-)   No insecticidal activity; tested against blackheaded fireworm.

Table 1-1 was contributed by Dr. Jay De Rocher.

Modified *B.t.* toxin Genes

Gene	Length[a]	Change In GC (%)	Codons Changed	Plant Species	Expression			Active Against[e]	Reference
					Increased mRNA[b]	Maximum Protein[c]	Increased Activity[d]		
cryIA(a)	1-645	nr	141	poplar	nr	nr	nr	forest tent caterpillar	(68)
				cranberry	nr	nr	nr	gypsy moth	(69)
cryIA(b)	1-612	nr	nr	cotton	nr	0.1%	(-)	cabbage looper	(42)
							+	beet armyworm	
							+	cotton bollworm	
"	1-615	37 to 49	356	tomato	+	0.2-0.3%	+	tobacco hornworm	(43)
				tobacco	nr	0.2-0.3%	+	tobacco hornworm	(43)
"	29-607	37 to 41	55	tomato	+	0.02%	+	tobacco hornworm	(43)
				tobacco	nr	0.02%	+	tobacco hornworm	
"	1-648	38 to 65	549	maize	nr	0.4%	nr	european corn borer	(46)
"	1-724	38 to 59	482	rice	nr	0.05%	nr	striped stemborer	(48)
							nr	leaffolder	
"	1-646	nr	21	tomato	+	<<	+	tobacco hornworm	(26)
				tobacco	+	<<	+	tobacco budworm	
							+	tobacco hornworm	
cryIA(c)	1-640	nr	nr	cotton	nr	0.1%	+	cotton bollworm	(42)
"	1-615	nr	306	cotton	nr	0.1%	+	cotton bollworm	(42)
"	29-615	nr	nr	tomato	nr	0.2-0.3%	nr	tobacco hornworm	(43)
				tobacco	nr	0.2-0.3%	nr	tobacco hornworm	(43)
"	10-613	34 to 64	496	maize	+	nr	nr	nr	(49)
				tobacco	+	nr	+	tobacco hornworm	
				Arabidopsis	+	nr	+	tobacco hornworm	
cryIC	1-654	nr	42	tobacco	+	<<	+	beet armyworm	(26)
								tobacco hornworm	
cryIIIA	48-644	36 to 49	nr	tobacco	nr	0.6%	nr	colorado potato beetle	(44)
"	48-645	37 to 49	347	potato	nr	0.3%	+	colorado potato beetle	(45)
"	1-610	36 to 45	190	potato	nr	0.025%	nr	colorado potato beetle	(47)
				carrot	+	0.005%	nr	nr	
				maize	+	0.005%	nr	nr	

the codons altered (see Table 1-1.). This approach was first used successfully to generate

modified *cryIA(b)* and *cryIA(c)* genes expressed in cotton conferring resistance to

cabbage looper (*Trichoplusia ni*), beet armyworm, and *Helicoverpa zea* (42). The

extensive modifications to these genes were estimated to increase expression at least 100-

fold based on insect bioassays (42) and western analysis (1 mg/mg total protein) (43). A

cryIIIA gene that was modified in the same way and transformed into potato yielded

expression levels up to 300-fold higher than unmodified *cryIIIA* as assayed by western

analysis (3 mg/mg total protein)(45).

Targeting AT-rich regions for modification on a limited scale has been attempted as

an alternative to comprehensive modification in an effort to achieve useful levels of

expression with a smaller investment of time and resources. Potential instability and

polyadenylation signals were removed from a *cryIC* gene and a *cryIA(b)* gene with only

42 and 21 altered codons, respectively (26). These modifications increased mRNA levels

from undetectable to detectable, but protein levels remained below the limit of detection

by western blot analysis. These genes did, however, provide insect protection to tobacco

and tomato plants. A second modified *cryIA(b)* gene was constructed with limited

alterations affecting only 55 codons in 9 AT-rich regions (referred to as A through I)

compared to 356 codons changed in a fully modified *cryIA(b)* gene (43). These limited

modifications provided an increase in *B.t.* toxin protein accumulation in tobacco and

tomato that was ten-fold greater than the wild-type gene but ten-fold less than the fully

modified gene. A series of modified genes in which subsets of the nine AT-rich regions

were targeted demonstrated that the changes in one region (region B) provided a large

share of the ten-fold increase in expression over wild-type. From these results, it can be

concluded that specific sequences limiting expression should be identifiable and that the modifications made to the nine targeted regions missed sequences responsible for the further ten-fold improvement in expression found in the fully modified gene. This latter conclusion was supported by the finding that substitution of the 5' one-third of the fully modified gene (containing region B) into the wild-type gene gave only a ten-fold increase in expression. Substitution of the 3' two-thirds of the fully modified gene into the wild-type gene gave nearly a ten-fold increase in expression, indicating that other deleterious sequences must exist in this region. These studies with partially modified genes demonstrate that a degree of improved expression can be achieved without resorting to complete re-engineering of the coding sequence. The fact that improved expression was obtained after specifically targeting potential polyadenylation signals and instability sequences suggests that these processes may be important factors limiting *B.t.* toxin gene expression in plants. It remains to be shown, however, that the partial modifications described actually affect polyadenylation or mRNA stability.

Another approach used successfully in the modification of *B.t.* toxin sequences has been to alter the codon usage throughout the coding region to conform to that found in plant genes without changing the encoded amino acid sequence (46-49). Since plant codon usage is biased toward codons with G or C in the third position (50), the change from bacterial to plant codon bias automatically raises the GC content of the gene. The resulting increase in GC content eliminates by default most or all AT-rich sequences, thereby removing possible splicing, polyadenylation, and instability signals. The conversion to plant preferred codons has the added potential benefit of improved translational efficiency. In most cases, the codon bias was altered to conform to dicot

codon usage. For three genes, modifications were designed to give a monocot codon

bias, specifically like that found in maize (46,49) and rice genes (48). However, a

monocot- or dicot-like codon bias does not appear to limit the expression of modified *B.t.*

toxin genes to one or the other class of plants. A modified *crylIIA* gene with a dicot

codon bias was expressed well in maize cells (47) and a modified *crylA(c)* gene with a

monocot codon bias was expressed well in *Arabidopsis* and tobacco cells (49). Taken

together, these studies show that extensive modification of AT-rich sequences and codon

usage can be an effective way to engineer high *B.t.* toxin gene expression in a variety of

plant species. The exact molecular basis for the improved expression has been unclear

due to the breadth of the sequence changes that were made in the modified genes.

However, when studies on the modified *B.t.* toxin genes are interpreted together with

expression analyses of their unmodified counterparts, important mechanistic insights are

beginning to emerge, as discussed in the next section.

Further improvements with promoters and other 5' flanking sequences. Many insect

larvae feed in locations that are difficult to treat using conventional insecticides. Such is

the case in controlling the European corn borer. Part of this insect's life cycle is spent

consuming pollen in the leaf axils of maize plants and tunneling through the stalk,

locations that provide protection from most chemical pesticides. Through the use of

tissue-specific and pollen-specific promoters, greater control of this pest has been

achieved. Transgenic maize expressing a modified *crylA(b)* gene under the control of a

maize pollen-specific promoter and a phosphoenolpyruvate carboxylase promoter, a

promoter which is active in mesophyll cells, displayed high levels of *B.t.* toxin production

and minimal damage to the stalk as a result of tunneling. Insect mortality on leaf pieces was 85 to 100%, which was significantly better than most control plants transformed with a 35S-driven gene (46).

It has been observed that a combination of the 5' UTR and transit peptide from the *Arabidopsis thaliana* rubisco small subunit *ats1A* gene can increase the expression of a modified *cryIA(c)* gene 10- to 20-fold in tobacco (51). The increase is not a result of targeting the protein to the chloroplast and appears to be independent of the promoter, because the same gene driven by either the *ats1A* promoter in tobacco plants or an enhanced CaMV 35S in tobacco protoplasts showed a similar increase in expression. However, the effect the *ats1A* 5' UTR and transit peptide have on enhancing expression may depend on the coding region to which these sequences are attached since they increased the expression of a β-glucuronidase (GUS) reporter gene by no more than 6-fold (51).

ACCUMULATION OF *B.t.* TOXIN mRNA IN TRANSGENIC PLANTS AND FACTORS THAT AFFECT ITS ABUNDANCE

In all cases where mRNA levels have been examined, poor expression of unmodified *B.t.* toxin genes is always associated with low mRNA accumulation (15,18,23,24). Conversely, high expression of modified *B.t.* toxin genes always results in higher mRNA levels (see Table 1-1). In additional experiments examining the induction of *B.t.* toxin that occurred upon anthesis in a series of transgenic tobacco plants, the relative differences in the accumulation of *B.t.* toxin mRNA among the lines reflected the

differences in protein levels (28). Each of these observations underscores the strong

correlation between *B.t.* toxin protein and the level of the corresponding mRNA that is

generally observed in transgenic plants. They also emphasize the importance of the

factors that control mRNA abundance in establishing the efficacy of a given *B.t.* toxin

gene in transgenic plants.

What mechanisms, then, limit the accumulation of unmodified *B.t.* toxin transcripts in

plants? The initial observation that poor transcript accumulation was characteristic of

unmodified genes regardless of the promoter used, argued that the problem was post-

transcriptional. However, it was still possible that unmodified *B.t.* toxin coding regions

contained a repressor sequence capable of inhibiting transcription initiation or elongation

from all promoters. Recent nuclear-run-on transcription experiments demonstrate that

there is no significant difference in the transcriptional activities of a poorly expressed,

unmodified *cryIA(b)* gene and a highly expressed control (52) or an unmodified *cryIA(c)*

and a highly expressed, fully modified *cryIA(c)* gene (49). These results were obtained in

experiments carried out with probes from either the 5' (52) or the 3' (49) ends of the

genes, indicating that the expression problem is post-transcriptional. This prompts an

examination of other aspects of *B.t.* toxin mRNA metabolism such as splicing, mRNA

stability, and polyadenylation for their role in limiting gene expression. These topics, as

well as the role of codon usage and translation, are discussed in the sections below.

Splicing

Northern blot analyses of plants expressing different *B.t.* toxin genes often reveal the presence of *B.t.* toxin transcripts of a shorter length than expected. Tobacco plants transformed with a *cry IA(c)* protoxin gene produced a truncated, polyadenylated transcript of 1.7 kb (17), whereas plants expressing a *cryIA(b)* protoxin gene accumulated two short polyadenylated transcripts of approximately 1.6 and 0.9 kb (21). Two transcripts also of approximately 1.6 and 0.9 kb were observed in the poly(A)-enriched fractions of tobacco plants expressing a truncated *cryIA(b)* gene (21). Tobacco plants expressing another *cryIA(b)* gene generated short transcripts, although these transcripts were also generated in an *in vitro* transcription system (18). Two short polyadenylated transcripts of 0.9 kb and 0.6 kb were also identified in tobacco cells expressing a *cryIA(c)* gene (30). Some of these short transcripts have been shown to arise from polyadenylation in the *B.t.* toxin coding region, as discussed in the next section. However, others may be generated through the aberrant splicing of *B.t.* toxin messages.

Pre-mRNA splicing in plants is known to be facilitated by the presence of AU-rich elements within the intron (reviewed in (53) and (54)). Unmodified *B.t.* toxin transcripts have a high AU content; thus, it is possible that these sequences trigger aberrant splicing thereby limiting the accumulation of intact *B.t.* toxin transcripts in plants. This possibility was recently investigated in tobacco using a reverse transcriptase-polymerase chain reaction (RT-PCR) assay (52). Plants expressing a truncated *cryIA(b)* gene were found to excise three regions of the *B.t.* toxin transcript. The sequences surrounding the boundaries of the excised regions conform to consensus splice junctions. Mutations were

made to one or two of the 5' splice junctions to assess the contribution splicing makes to the low accumulation of *B.t.* toxin mRNA in plants. The inhibition of splicing was found to increase expression on the order of 4- to 20-fold depending on the experiment. Unfortunately, these experiments do not address the magnitude of the contribution aberrant splicing makes to the overall poor accumulation of unmodified *B.t.* toxin transcripts because the mutations were evaluated within the context of a partially modified *B.t.* toxin gene. However, they do argue that splicing occurs and has a negative influence.

Polyadenylation

Another explanation for the existence of short, polyadenylated *B.t.* toxin messages in transformed plants is that they are generated through the aberrant polyadenylation of newly synthesized transcripts (17,30). This has been most clearly demonstrated in the case of two short transcripts produced by an unmodified *cryIA(c) B.t.* toxin gene in stably transformed tobacco cells. RT-PCR analysis identified two poly(A) addition sites within the *B.t.* toxin coding region that have positions corresponding to the sites where polyadenylation was predicted to occur based on the size (~600 and ~900 bases) of these short transcripts (30). Oligonucleotide-directed RNase H cleavage and northern blot hybridization with probes from different regions of the *cryIA(c)* gene provided further confirmation that the two short transcripts resulted from poly(A) addition at the mapped

sites. The sequences upstream of both these sites are typical of plant polyadenylation signals (30,55).

Due to their low abundance under normal growth conditions, the two short transcripts can be detected in poly(A)+ but not in total RNA of stably transformed tobacco cells (30). However, if cells are treated with cycloheximide, a protein synthesis inhibitor known to stabilize many unstable transcripts (56) (e.g., by blocking the production of a labile mRNA degradation factor), the abundance of the two short transcripts increases significantly. Under these conditions, the transcripts are routinely observed in total RNA from cells expressing the unmodified *cryIA(c)* gene (30). Therefore, it seems likely that a significant proportion of the transcripts produced by unmodified *cryIA(c)* genes are polyadenlyated at the mapped sites but fail to accumulate because they are unstable. These polyadenylation sites were eliminated in a highly expressed, modified version of the same *cryIA(c)* gene, as verified by the absence of short polyadenylated transcripts in both untreated and CHX-treated tobacco cells (49). Thus, the data indicate that limited accumulation of intact *B.t.* toxin transcripts in plants is in part due to aberrant polyadenylation in the coding region.

mRNA Stability

The failure of unmodified *B.t.* toxin transcripts to accumulate is often attributed to mRNA instability because little or no *B.t.* toxin mRNA can be detected, even when transcription is driven by a strong plant promoter. Further support for the idea that the

transcripts are unstable derives from the observation that unmodified *B.t.* toxin genes support normal levels of transcriptional activity in nuclear run-on assays, as mentioned earlier (49,52). Moreover, the AT richness of *B.t.* toxin genes resembles well-known mRNA instability sequences present in the 3' UTRs of labile mammalian proto-oncogene and cytokine mRNAs such as those encoding c-*fos* and GM-CSF (see 57 for a review). Multiple ATTTA motifs are a common feature of these 3' UTRs and can be essential for them to function as signals for rapid decay (58). Unmodified *B.t.* toxin genes also contain multiple ATTTA sequences and these have been major targets of mutagenesis during the construction of modified genes (26,42-45). ATTTA repeats have recently been shown to trigger rapid decay of reporter transcripts in tobacco (41) so it is likely that some ATTTA sequences in *B.t.* toxin function in a similar manner. However, it should be noted that not all ATTTA sequences act to decrease mRNA stability (57,59), in part because the minimal domain appears to be longer in some cases (60,61).

In principle, it should be possible to test whether unmodified *B.t.* toxin transcripts are unstable and to examine the role of ATTTA sequences by measuring mRNA half-lives by established methods (e.g., blocking transcription with an inhibitor such as actinomycin D and then monitoring the disappearance of the mRNA over time (41,62,63)). Unfortunately, the scarcity of unmodified *B.t.* toxin mRNA in plant cells has presented a technical barrier to this type of analysis. To overcome this obstacle and evaluate the stability of unmodified *B.t.* toxin transcripts, three different approaches have been taken, as outlined below.

In the first study, accumulation of *B.t.* toxin mRNA was investigated in carrot protoplasts after introduction of plasmids containing unmodified *cryIA(b)* or *cryIIIA*

genes under the control of the 35S promoter (21). Abundance of *B.t.* toxin mRNAs was monitored by northern analysis of poly(A)+ RNA isolated from protoplasts harvested at 1, 2, 4, 8 and 18 h after electroporation. A second plasmid carrying an octopine synthase (OCS) gene, also driven by a 35S promoter, was co-electroporated to serve as a control. A basic assumption of this experiment was that transcription of the plasmid-borne genes would stop at some time after electroporation, allowing decay of the transcripts to be observed during the time course. The disappearance pattern of both *B.t.* toxin transcripts differed from that of the OCS transcripts. Intact *cry* mRNA present at 8 hours was replaced by a more intense smear of lower molecular weight transcripts at 18 hours, whereas OCS mRNA remained intact throughout the time course. That the intact *B.t.* toxin transcripts disappeared at 18 h compared to the OCS transcripts, indicated that the former were relatively unstable. In addition, the smear of smaller *cry* transcripts may have been degradation intermediates. These were proposed to arise from a 5' to 3' exonucleolytic degradation mechanism because they were present in the poly(A)+ fraction and multiple 5' ends were mapped by nuclease protection (21). A series of constructs in which increasingly large portions of the *cryIA(b)* coding region were deleted from the 3' end were also electroporated into carrot protoplasts. None of the deletions prevented the differential transcript disappearance with respect to OCS mRNA (21). Similar results were also reported for a *cryIA(c)* deletion series (unpublished data in 21). This suggests that sequences sufficient to trigger rapid decay of unmodified *B.t.* toxin mRNA in protoplasts are present in the 5' portion of the coding region.

A second investigation addressing the stability of unmodified *B.t.* toxin transcripts was carried out in tobacco protoplasts using a *cryIA(b)* gene (52) similar to that in the

first study. However, in the second study, different methodology was used and different results were obtained. *In vitro* synthesized *crylA(b)* and *bar* (bialaphos resistance) RNAs that were 5' capped and polyadenylated were co-electroporated into tobacco protoplasts. The decay of the introduced transcripts was then followed over a 5-h time course. It was assumed that decay of *in vitro* synthesized transcripts introduced into protoplasts by electroporation in these experiments accurately reflected the decay of endogenously generated transcripts. The half-lives of the *crylA(b)* and *bar* transcripts after electroporation were determined to be 7.8 ± 3.0 and 4.9 ± 1.3 hours, respectively. Since the *in vitro* synthesized *crylA(b)* and *bar* transcripts decayed in protoplasts with half-lives that were not statistically different, it was concluded that the low accumulation of *crylA(b)* mRNA relative to *bar* mRNA in transgenic plants was not due to the instability of *crylA(b)* transcripts. It is interesting to note that a *bar* mRNA half-life of 2.3 h was calculated from a time course experiment following actinomycin D treatment of protoplasts from a plant transgenic for *bar* (64). It appears that when similar experiments were carried out using protoplasts from a *crylA(b)* transgenic plant, the *B.t.* toxin mRNA half-life was more reflective of the electroporation results (5.3 h; unpublished results in 52) than in the case of *bar*. Therefore, there was no evidence for rapid decay of unmodified *B.t.* toxin transcripts in this study. In addition, differences in transcriptional activity, measured in run-on assays could not account for the poor accumulation of unmodified *crylA(b)* mRNA (52). Because the run-on data were obtained with a probe against the 5' region of *crylA(c)*, the simplest explanation remaining is that B.t. toxin

mRNA accumulation is limited at some step between transcription initiation and mRNA degradation.

The third study compared the stabilities of transcripts from unmodified and extensively modified *crylA(c)* genes in stably transformed tobacco cells (49). The approach was to measure the half-lives of the transcripts to determine whether highly modified *B.t.* toxin mRNA levels were due to an increase in mRNA stability caused by the sequence modifications. A transient treatment of tobacco cells expressing the unmodified gene with the translational inhibitor cycloheximide was used to induce *B.t.* toxin mRNA to levels that would allow half-life measurements to be made. The ability to induce mRNA accumulation with cycloheximide was itself an indication that the unmodified *B.t.* toxin transcripts might be unstable, since cycloheximide induction is characteristic of other unstable mRNAs (56). Accumulation of the modified *B.t.* toxin mRNA was not increased by cycloheximide. After removal of cycloheximide from the tobacco cells, transcription was blocked with actinomycin D and transcript decay monitored over a 2-h time course. The half-life of unmodified *B.t.* toxin mRNA was determined to be several-fold less than for the modified *B.t.* toxin mRNA. The actual difference may be even greater if the presence or effect of the cycloheximide was not completely reversed before the time course began. In any event, this study showed that unmodified *crylA(c)* mRNA is unstable in plants and that sequence modifications made to generate the modified *B.t.* toxin gene improved mRNA accumulation, at least in part, by increasing mRNA stability. The specific sequence modifications responsible for the increased stability remain to be identified.

At present it is unclear why the results of the second study are inconsistent with those of the other two which found the *B.t.* toxin mRNA to be unstable. It is tantalizing to propose that unmodified *B.t.* toxin transcripts are unstable only if they are transported to the cytoplasm via normal pathways from the nucleus as in the first and third approaches, rather than by electroporation as in the second study. However, an experiment cited in the latter, in which the unmodified *B.t.* toxin mRNA appears to be stable in protoplasts from a transgenic leaves (unpublished results in 52), seems to argue against this possibility. Therefore, alternative explanations, such as differences among the transcripts examined or the plant systems used, would appear to be more viable. In any event, the data indicate that some unmodified *B.t.* toxin transcripts are rapidly degraded in plant cells, which undoubtedly contributes to their poor accumulation.

Codon Usage

The difference in codon usage between plants and unmodified *B.t.* toxin genes is frequently pointed to as an explanation for poor *B.t.* toxin expression in plants. This is because the codon bias of *B.t.* toxin genes also introduces codons that are rare in plant genes (21). In actuality, codon usage covers several possible mechanisms one or more of which could limit expression. The chance of introducing sequences that cause splicing, polyadenylation, or mRNA decay is heightened by the AT-rich sequences that correspond to rare codons, as discussed in previous sections. If pool sizes for the tRNAs specified by rare codons are small in plants, the presence of these codons could cause ribosomes to

stall during translation of *B.t.* toxin transcripts. Stalling of ribosomes at rare codons could simply reduce the rate of protein synthesis. Alternatively, stalled ribosomes could cause mRNA destabilization by leaving *B.t.* toxin transcripts exposed to components of the RNA turnover machinery. Little is known about tRNA pool sizes in plants but an especially restrictive codon bias is found in some highly expressed plant genes (50).

To date there have been no studies directly addressing whether ribosome pausing at rare codons in unmodified *B.t.* toxin transcripts reduces protein synthesis. Some insight could be gained from a careful comparison of the ratios of *B.t.* toxin mRNA and protein produced by unmodified and modified *B.t.* toxin genes. However, the products of unmodified genes are generally extremely low and hard to quantify. In one case, increases in *B.t.* toxin protein and mRNA in tomato plants expressing modified *crylA(b)* genes were compared with the *B.t.* toxin protein and mRNA levels in a small number of plants expressing the unmodified gene (unpublished data in 43). Partial modification of the *crylA(b)* gene was reported to give a 2.5-fold increase in mRNA but a 10-fold increase in protein over wild-type levels and full modification a 5-fold increase in mRNA but a 50-fold increase in protein over wild type. The method used for RNA quantification (nuclease protection using a probe covering a 5' segment of the *B.t.* toxin genes) cannot differentiate between intact and aberrantly processed transcripts, although the latter were not evident on northern blots of total RNA. Assuming they are absent, the data may suggest that translational problems figure more prominently than mRNA accumulation problems for limiting the expression of the unmodified *crylA(b)* gene (43).

The effect, if any, of rare codons in the *B.t.* toxin transcript on mRNA stability is unknown. It is difficult to evaluate the effect of rare codons on *B.t.* toxin transcript

stability separately from the effect of the RNA sequences that encode them since conversion of rare codons to common codons unavoidably alters the mRNA sequence. One approach used to address this problem was to determine the effect on mRNA half-life of introducing a *B.t.* toxin sequence containing rare codons into a gene encoding a stable transcript. A 26-codon segment of a *cryIA(c)* gene, rich in rare codons, was inserted in frame into a phytohemagglutinin (PHA) gene (65). This *B.t.* toxin segment was chosen in part because it could be easily frameshifted to common codons to control for RNA sequence effects that were independent of the rare codons. However, this control proved to be unnecessary because the PHA transcripts were not destabilized by either a single or double insertion of the *cryIA(c)* sequence. While these results suggest that rare codons do not have a strong effect on mRNA stability, they do not rule out the possibility that an even greater number or a different set of rare codons would accelerate mRNA decay.

CHLOROPLAST TRANSFORMATION

Poor expression of wild-type *B.t.* toxin genes in plants appears to result from incompatibility of AT-rich *B.t.* toxin sequences with mechanisms that can affect mRNA accumulation and possibly translational efficiency. An alternative to modifying the sequences to be compatible with nuclear and cytoplasmic mechanisms is to express the gene from the chloroplast genome instead of the nuclear genome. Taking advantage of the bacterial ancestry of the chloroplast circumvents the problem of expressing an AT-rich bacterial gene in the context of GC-rich nuclear genes. Recently, an unmodified *cryIA(c)* protoxin gene was introduced into tobacco chloroplasts by particle bombardment

followed by homologous recombination (66). Expression of the protoxin in regenerated plants with transformed chloroplasts was at least three times greater on a percent total protein basis (>30mg/mg protein) than that reported for the best modified *B.t.* toxin gene transcribed in the nucleus. Interestingly, no obvious effects on plant growth, or fertility were observed which suggests the chloroplast is insensitive to the ill effects the protoxin has been suggested to have in the cytoplasm (18). A high level of *B.t.* toxin mRNA was observed in total RNA preparations from these plants, most of which appears to be in a dicistronic form. The significant accumulation of *B.t.* toxin transcripts in plants with transformed chloroplasts indicates that the mechanisms that limit RNA accumulation in this organelle are likely to differ from those acting on nuclear encoded transcripts. Therefore, this approach could be a powerful solution to the problem of expressing *B.t.* toxin genes in plants if chloroplast transformation can be effectively applied to a wide range of species. It may also prove useful for expression of other problematic genes that can fulfill their intended function when expressed in the chloroplast. The maternal inheritance of transgenes located in the chloroplast would be beneficial in that dispersal of transgenes by pollen would be prevented. However, hybrid lines would have to be generated from female plants containing the transgene. This would be a disadvantage for those breeding programs which utilize paternal inbred gene donors as well as maternal inbred gene donors to efficiently introduce transgenes into multiple hybrid lines.

CONCLUSIONS AND FUTURE PROSPECTS

Now that the hurdles limiting the expression of *B.t.* toxin genes in plants have been largely overcome, it is clear that a great deal has been learned in the process. It can no longer be assumed that attaching a strong promoter to a gene will insure high expression in plants. As illustrated by the work on *B.t.* toxins, expression can be hindered at the mRNA level through mechanisms such as aberrant splicing, aberrant polyadenlyation, and rapid mRNA degradation. Avoiding these problems as well as those affecting translation and post-translational processes has become a popular consideration in plant biotechnology, in large part because of work on *B.t.* toxin genes.

The favored way to achieve high expression of *B.t.* toxin genes has been to create fully modified genes removing all seemingly deleterious sequences. But is this strategy the best means by which to increase the expression of new *B.t.* toxin genes and other problematic genes based on current knowledge? There is no universal answer to this question because different genes have different characteristics and constraints. In some cases it may be possible to circumvent the complications presented by nuclear and cytoplasmic processes through chloroplast transformation. The initial results with *B.t.* toxin expression in chloroplasts have been very exciting and the utility of this approach will become even greater if the technique can be adapted to a variety of crop plants. For gene products that cannot achieve their desired function in chloroplasts or in plants not amenable to chloroplast transformation, gene modification may be the best option. Although resynthesis of *B.t.* toxin genes to be more plant-like without prior knowledge of the underlying problems has been fairly successful, this process is still rather labor

intensive and costly. In addition, there is no guarantee that a fully modified gene will work as expected in transgenic plants. Therefore, placing some emphasis on elucidating the molecular basis of the problem would be advantageous. Determining whether the poor expression of a given foreign gene is a result of low mRNA levels and/or low protein levels is relatively simple and might indicate whether the situation is similar to that encountered with unmodified *B.t.* toxin genes.

Considerable progress has been made toward delineating the sequences responsible for poor expression of unmodified *B.t.* toxin genes, including the identification of anomalous polyadenylation and splicing sites. As this type of analysis progresses, it should be possible to address whether elimination of a limited number of sequences *known* to be deleterious will be sufficient to elevate gene expression to near maximal levels. The significant improvements that have been observed with partially modified *B.t.* toxin genes are very encouraging in this regard.

The knowledge gained from investigating *B.t.* toxin genes in plants has basic as well as applied significance. The mechanisms that limit the expression of unmodified B.t. toxin genes are vital plant mechanisms that naturally affect endogenous RNA and protein levels. There is considerable potential to further our understanding of these basic processes through the study of foreign genes in plants, particularly for those processes where knowledge is most limited. For example, only a few sequences that trigger rapid mRNA decay have been identified in plants so far (67). If those in *B.t.* toxin mRNAs can be identified, our basic knowledge of *cis*-acting sequences controlling mRNA stability in plants will increase appreciably. This information may also help identify the natural targets of the corresponding decay pathways. Moreover, with respect to biotechnology,

the identification of problematic sequences in *B.t.* toxin genes will make it easier to

identify and remove similar sequences from other genes engineered for high expression in

plants. Overall, it is hoped that the efforts devoted to *B.t.* toxin genes in plants will result

in maximum benefits for crop improvement and basic knowledge, and a minimum of

litigation.

ACKNOWLEGDEMENTS

REFERENCES

1 Fray, R. G. and Grierson, D. (1993) Trends Genet. 9, 438-443.

2 Beachy, R. N., Loesch-Fries, S. and Tumer, N. E. (1990) Annu. Rev. Phytopathol. 28, 451-474.

3 Botterman, J. and Leemans, J. (1988) Trends Genet. 4, 219-222.

4 Mariani, C., De Beuckeleer, M., Truettner, J., Leemans, J. and Goldberg, R. B. (1990) Nature 347, 737-741.

5 Mariani, C., Gossele, V., De Beuckeleer, M., De Block, M., Goldberg, R. B., De Greef, W. and Leemans, J. (1992) Nature 357, 384-387.

6 Voelker, T. A., Worrell, A. C., Anderson, L., Bleibaum, J., Fan, C., Hawkins, D. J., Radke, S. E. and Davies, H. M. (1992) Science 257, 72-74.

7 Poirier, Y., Dennis, D. E., Klomparens, K. and Somerville, C. (1992) Science 256, 520-523.

8 Nawrath, C., Poirier, Y. and Somerville, C. (1994) Proc. Natl. Acad. Sci. USA 91, 12760-12764.

9 Tailor, R., Tippett, J., Gibb, G., Pells, S., Pike, D., Jordan, L. and Ely, S. (1992) Mol. Microbiol. 6, 1211-1217.

10 Höfte, H. and Whiteley, H. R. (1989) Microbiological Reviews 53, 242-255.

11 Feitelson, J. S., Payne, J. and Kim, L. (1992) Bio/Technology 10, 271-275.

12 Aronson, A. I. (1993) Mol. Microbiol. 7, 489-496.

13 Burges, H. D. (1982) Parasitology 84, 79-117.

14 McGaughey, W. H. and Whalon, M. E. (1992) Science 258, 1451-1455.

15 Vaeck, M., Reynaerts, A., Höfte, H., Jansens, S., De Beuckeleer, M., Dean, C., Zabeau, M., Van Montagu, M. and Leemans, J. (1987) Nature 328, 33-37.

16 Fischhoff, D. A., Bowdish, K. S., Perlak, F. J., Marrone, P. G., McCormick, S. M., Niedermeyer, J. G., Dean, D. A., Kusano-Kretzmer, K., Mayer, E. J., Rochester, D. E., Rogers, S. G. and Fraley, R. T. (1987) Bio/Technology 5, 807-813.

17 Adang, M. J., Firoozabady, E., Klein, J., DeBoer, D., Sekar, V., Kemp, J. D., Murray, E., Rocheleau, T. A., Rashka, K., Staffeld, G., Stock, C., Sutton, D. and Merlo, D. J. (1987) in Expression of a *Bacillus thuringiensis* Insecticidal Crystal Protein Gene in Tobacco Plants. (Arntzen, C.J. and Ryan, C., eds.) 46. pp. 345-353, Alan R. Liss,Inc.

18 Barton, K. A., Whiteley, H. R. and Yang, N.-S. (1987) Plant Physiol. 85, 1103-1109.

19 MacIntosh, S. C., Stone, T. B., Sims, S. R., Hunst, P. L., Greenplate, J. T., Marrone, P. G., Perlak, F. J., Fischhoff, D. A. and Fuchs, R. L. (1990) Journal of Invertebrate Pathology 56, 258-266.

20 Delannay, X., LaVallee, B. J., Proksch, R. K., Fuchs, R. L., Sims, S. R., Greenplate, J. T., Marrone, P. G., Dodson, R. B., Augustine, J. J., Layton, J. G. and Fischhoff, D. A. (1989) Bio/Technology 7, 1265-1269.

21 Murray, E. E., Rocheleau, T., Eberle, M., Stock, C., Sekar, V. and Adang, M. (1991) Plant Mol. Biol. 16, 1035-1050.

22 Hoffmann, M. P., Zalom, F. G., Wilson, L. T., Smilanick, J. M., Malyj, L. D., Kiser, J., Hilder, V. A. and Barnes, W. M. (1992) J. Econ. Entomol. 85, 2516-2522.

23 Cheng, J., Bolyard, M. G., Saxena, R. C. and Sticklen, M. B. (1992) Plant Sci. 81, 83-91.

24 Dandekar, A. M., McGranahan, G. H., Vail, P. V., Uratsu, S. L., Leslie, C. and Tebbets, J. S. (1994) Plant Sci. 96, 151-162.

25 Parrott, W. A., All, J. N., Adang, M. J., Bailey, M. A., Boerma, H. R. and Stewart Jr, C. N. (1994) In Vitro Cell. Dev. Biol. 30, 144-149.

26 van der Salm, T., Bosch, D., Honee, G., Feng, L., Munsterman, E., Bakker, P., Stiekema, W. J. and Visser, B. (1994) Plant Mol. Biol. 26, 51-59.

27 Shin, D. I., Podila, G. K., Huang, Y. and Karnosky, D. F. (1994) Can. J. For. Res. 24, 2059-2067.

28 Carozzi, N. B., Warren, G. W., Desai, N., Jayne, S. M., Lotstein, R., Rice, D. A., Evola, S. and Koziel, M. G. (1992) Plant Mol. Biol. 20, 539-548.

29 Kay, R., Chan, A., Daly, M. and McPherson, J. (1987) Science 236, 1299-1302.

30. Diehn, S., De Rocher, E. J., Chiu, W. and Green, P. (1996) in preparation.

31 Gallie, D. R. (1993) Annu. Rev. Plant Physiol. Plant Mol. Biol. 44, 77-105.

32   Adang, M., DeBoer, D., Endres, J., Firoozabady, E., Klein, J., Merlo, A., Merlo, D., Murray, E., Rashka, K. and Stock, C. (1988) in Manipulation of *Bacillus thuringiensis* Genes for Pest Insect Control. (Roberts, D.W. and Granados, R.R., eds.) pp. 31-37, Cornell University.

33   Barnes, W. M. (1990) Proc. Natl. Acad. Sci. USA 87, 9183-9187.

34   Obukowicz, M. G., Perlak, F. J., Kusano-Kretzmer, K., Mayer, E. J. and Watrud, L. S. (1986) Gene 45, 327-331.

35   Nambiar, P. T. C., Ma, S. W. and Iyer, V. N. (1990) Appl. Environ. Microbiol. 56, 2866-2869.

36   Skøt, L., Timms, E. and Mytton, L. R. (1994) Plant and Soil 163, 141-150.

37   Bezdicek, D. F., Quinn, M. A., Forse, L., Heron, D. and Kahn, M. L. (1994) Soil-biol-biochem. 26, 1637-1646.

38   Sinibaldi, R. M. and Mettler, I. J. (1992) Prog. Nucleic Acid Res. Mol. Biol. 42, 229-257.

39   Dean, C., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H. and Bedbrook, J. (1986) Nucleic Acids Res. 14, 2229-2240.

40   Shaw, G. and Kamen, R. (1986) Cell 46, 659-667.

41   Ohme-Takagi, M., Taylor, C. B., Newman, T. C. and Green, P. J. (1993) Proc. Natl. Acad. Sci. USA 90, 11811-11815.

42   Perlak, F. J., Deaton, R. W., Armstrong, T. A., Fuchs, R. L., Sims, S. R., Greenplate, J. T. and Fischhoff, D. A. (1990) Bio/Technology 8, 939-943.

43   Perlak, F. J., Fuchs, R. L., Dean, D. A., McPherson, S. L. and Fischhoff, D. A. (1991) Proc. Natl. Acad. Sci. USA 88, 3324-3328.

44   Sutton, D. W., Havstad, P. K. and Kemp, J. D. (1992) Transgenic Research 1, 228-236.

45   Perlak, F. J., Stone, T. B., Muskopf, Y. M., Petersen, L. J., Parker, G. B., McPherson, S. A., Wyman, J., Love, S., Reed, G., Biever, D. and Fischhoff, D. A. (1993) Plant Mol. Biol. 22, 313-321.

46   Koziel, M. G., Beland, G. L., Bowman, C., Carozzi, N. B., Crenshaw, R., Crossland, L., Dawson, J., Desai, N., Hill, M., Kadwell, S., Launis, K., Lewis, K., Maddox, D.,

McPherson, K., Meghji, M. R., Merlin, E., Rhodes, R., Warren, G. W., Wright, M. and Evola, S. V. (1993) Bio/Technology 11, 194-200.

47   Adang, M. J., Brody, M. S., Cardineau, G., Eagan, N., Roush, R. T., Shewmaker, C. K., Jones, A., Oakes, J. V. and McBride, K. E. (1993) Plant Mol. Biol. 21, 1131-1145.

48   Fujimoto, H., Itoh, K., Yamamoto, M., Kyozuka, J. and Shimamoto, K. (1993) Bio/Technology 11, 1151-1155.

49. De Rocher, E. J., Diehn, S. and Green, P. (1996) in preparation.

50   Murray, E. E., Lotzer, J. and Eberle, M. (1989) Nucleic Acids Res. 17, 477-498.

51   Wong, E. Y., Hironaka, C. M. and Fischhoff, D. A. (1992) Plant Mol. Biol. 20, 81-93.

52   Van Aarssen, R., Soetaert, P., Stam, M., Dockx, J., Gosselé, V., Seurinck, J., Reynaerts, A. and Cornelissen, M. (1995) Plant Mol. Biol. 28, 513-524.

53   Luehrsen, K. R., Taha, S. and Walbot, V. (1994) Prog. Nucleic Acid Res. Mol. Biol. 47, 149-193.

54   Filipowicz, W., Gniadkowski, M., Klahre, U. and Liu, H. (1995) in Pre-mRNA splicing in plants. (Lamond, A.I. ed.)pp. 65-77, R.G. Landes Company.

55   Hunt, A. G. (1994) Annu. Rev. Plant Physiol. Plant Mol. Biol. 45, 47-60.

56   Ross, J. (1995) in mRNA stability in mammalian cells. (Joklic, W.K. ed.) ASM Press.

57   Greenberg, M. E. and Belasco, J. G. (1993) in Control of the Decay of Labile Protooncogene and Cytokine mRNAs. (Belasco, J.G. and Brawerman, G., eds.) pp. 199-218, Academic Press.

58   Chen, C.-Y. A. and Shyu, A.-B. (1994) Mol. Cell. Biol. 14, 8471-8482.

59. Walker, E. L., Weeden, N. F., Taylor, C. B., Green, P. J. and Coruzzi, G. M. (1995) Plant Mol. Biol. in press.

60   Lagnado, C. A., Brown, C. Y. and Goodall, G. J. (1994) Mol. Cell. Biol. 14, 7984-7995.

61   Zubiaga, A. M., Belasco, J. G. and Greenberg, M. E. (1995) Mol. Cell. Biol. 15, 2219-2230.

62 Newman, T. C., Ohme-Takagi, M., Taylor, C. B. and Green, P. J. (1993) Plant Cell 5, 701-714.

63 Taylor, C. B. and Green, P. J. (1995) Plant Mol. Biol. 28, 27-38.

64 Cornelissen, M. (1989) Nucleic Acids Res. 17, 7203-7209.

65. van Hoof, A. and Green, P.J. (1996) Plant J. 10, 415-424.

66 McBride, K. E., Svab, Z., Schaaf, D. J., Hogan, P. S., Stalker, D. M. and Maliga, P. (1995) Bio/Technology 13, 362-365.

67 Sullivan, M. L. and Green, P. J. (1993) Plant Mol. Biol. 23, 1091-1104.

68 McCown, B. H., McCabe, D. E., Russell, D. R., Robison, D. J., Barton, K. A. and Raffa, K. F. (1991) Plant Cell Reports 9, 590-594.

69 Serres, R., Stang, E., McCabe, D., Russell, D., Mahr, D. and McCown, B. (1992) J. Am. Soc. Hortic. Sci. 117, 174-180.

DISSERTATION TOPIC AND THESIS OVERVIEW

B.t. toxins continue to be the focus of many biotechnology companies today, just as it did over 10 years ago. The success of several products engineered with *B.t.* toxin genes has ensured the continued search for novel *B.t.* toxins which have different insecticidal activities or broader insecticidal activities. This group of genes has also stimulated the search for other insecticidal activities in organisms other than *Bacillus thuringiensis.* Thus, understanding the mechanisms which limited the accumulation of the wild-type *B.t.* toxin genes in plants should help simplify the engineering of new *B.t.* toxin genes and genes encoding proteins with novel insecticidal activities for high expression in plants by allowing specific mechanisms which limit expression to be addressed.

My thesis project has involved understanding the mechanisms which limit the expression of a typical *B.t.* toxin gene, the *cryIA(c)* gene, in tobacco at the level of mRNA accumulation. Chapter 1 details my work which has demonstrated the *cryIA(c)* coding region contains multiple sequences which are recognized by tobacco as polyadenylation signals. It was while attempting to localize the position of these signals in the coding region that it was observed the recognition of some of these polyadenylation signals was not as efficient in tobacco as in maize. This led to Chapter 3 of my thesis which investigated the generality of poly(A) signal recognition in monocots, specifically gramineous plants, and dicots. This topic was not covered in the introduction of this thesis, but a general overview of plant polyadenylation in comparison to polyadenylation

in mammals and the limited knowledge of poly(A) signal recognition differences between

monocots and dicots is provided at the beginning of the chapter.

Chapter 2

Premature Polyadenylation at Multiple Sites within the Coding Region Contributes to
Poor *B.t.* toxin Gene Expression in Tobacco

ABSTRACT

Some foreign genes introduced into plants are poorly expressed even when transcription is controlled by a strong promoter. Perhaps the best examples of this problem are the *cry* genes of *Bacillus thuringiensis,* which encode the insecticidal proteins commonly referred to as *B.t.* toxins. As a step towards overcoming such problems most effectively, we sought to elucidate the mechanisms limiting the expression of a typical *B.t.* toxin gene, *crylA(c)*, that accumulates very little mRNA in tobacco cells. Most cell lines transformed with a *crylA(c)* gene accumulate two short, polyadenylated transcripts. Interestingly, the abundance of these transcripts can be increased by treating the cells with cycloheximide (CHX), a translation inhibitor which can stabilize many unstable transcripts. Using a series of hybridization, RT-PCR, and RNase H digestion experiments, two poly(A) addition sites were identified in the *crylA(c)* coding region that correspond to the two short transcripts. A third polyadenylation site was identified using a chimeric gene. These results demonstrate for the first time that premature polyadenylation can limit the expression of a foreign gene in plants. Moreover, this work emphasizes that further study of the fundamental principles governing polyadenylation in plants will have not only basic but also applied significance.

43

INTRODUCTION

The ability to express foreign genes in plants has been an invaluable tool in understanding normal plant growth and development. Many molecular and biochemical questions concerning plant metabolism, physiology, development and responses to environmental cues have been addressed in this fashion. With respect to plant biotechnology, the introduction of foreign genes has led to a variety of significant advancements in crop improvement. As a result, the literature contains numerous reports demonstrating the successful expression of foreign genes in a variety of plants. However, this is not always the case. There is a growing list of foreign genes that are poorly expressed in plants.

The gene encoding the green fluorescent protein (GFP) from *Aequorea victoria*, which has been used as a reporter gene in many different biological systems, is not expressed well in some plant species. Little or no GFP-related fluorescence can be detected in Arabidopsis, tobacco or barley protoplasts or in Arabidopsis and tobacco plants transformed with the GFP gene, even when transcription is directed by the CaMV 35S promoter (Haseloff and Amos, 1995; Reichel et al., 1996; Haseloff et al., 1997; Rouwendal et al., 1997). Similarly, the genes encoding T4 lysozyme, *Klebsiella pneumoniae* cyclodextrin glycosyltransferase, and bacterial mercuric ion reductase are expressed at very low levels in plants despite the use of strong or tissue specific promoters to direct their transcription. Potato plants expressing the T4 lysozyme or *Klebsiella pneumoniae* cyclodextrin glycosyltransferase genes accumulate very low levels

of corresponding mRNA and protein (Oakes et al., 1991; Düring et al., 1993). Tobacco plants expressing the T4 lysozyme gene under the control of the mannopine synthase promoter, instead of the CaMV 35S promoter used in the transgenic potato plants, also accumulate very low levels of lysozyme protein (Düring, 1988). The full-length transcript of the bacterial mercuric ion reductase gene is not detectable in transgenic Petunia plants (Thompson, 1990). Instead, two short transcripts of approximately 800 nt accumulate.

The genes best known for their low expression in plants are the *B.t.* toxin genes (reviewed in Diehn et al., 1996). This family of genes from the gram-positive soil bacterium, *Bacillus thuringiensis*, encodes potent insecticidal proteins which target specific Orders of insects (Höfte and Whiteley, 1989; Aronson, 1993). Initial efforts to express *B.t.* toxins genes in plants using standard approaches yielded transgenic plants that produced little or no *B.t.* toxin mRNA and protein (Barton et al., 1987; Vaeck et al., 1987). The few plants generated expressing *B.t.* toxin were only resistant to those insect species that were the most susceptible to the toxin (Fischhoff et al., 1987; Delanney et al., 1989). The problem appeared to be at the level of mRNA accumulation as the plants which accumulated detectable *B.t.* toxin mRNA were the most insect resistant (Barton et al., 1987; Vaeck et al., 1987; Cheng et al., 1992; Dandekar et al., 1994).

Because of the potential agronomic importance, considerable effort has been made to increase the expression of *B.t.* toxin genes in plants. These efforts include expressing only the region of the gene encoding the insecticidal domain, modifying the 5' and 3' untranslated regions, generating protein fusions, and using a variety of strong promoters (reviewed in Diehn et al., 1996). The problem was eventually overcome by resynthesizing

the genes to be more "plant-like". In most cases, this included changing the codon usage to a plant preferred codon bias (reviewed in Diehn et al., 1996), which also has the effect of raising the GC content of the wild type gene. Many plant RNA processing signals such as those for polyadenylation, mRNA decay, and splicing are AT-rich. Wild type *B.t.* toxin genes have an AT content of approximately 65%. Therefore, increasing the GC content of the genes may eliminate potential RNA processing signals.

Why the transcripts of some poorly expressed foreign genes fail accumulate in plants remains unclear in most cases. The problem can occur at one or more steps in gene expression. Messenger RNA accumulation may be limited at the level of transcription by sequences within the coding region that adversely affect transcription initiation or elongation (Adang et al., 1987; Oakes et al., 1991). Alternatively, the problem may occur post-transcriptionally (Fischhoff et al., 1987; Vaeck et al., 1987; Oakes et al., 1991) due to aberrant splicing and/or degradation of the transcript. Recently, the transcripts of GFP and a *crylA(b) B.t.* toxin gene were found to contain 1 and 3 cryptic introns, respectively (Haseloff and Amos, 1995; Van Aarssen et al., 1995; Haseloff et al., 1997). The splicing of these transcripts is partly responsible for the low expression of the GFP and *crylA(b)* genes in plants.

It has been proposed that *B.t.* toxin coding regions contain plant polyadenylation signals (Adang et al., 1987; Perlak et al., 1991). However, no reports documenting polyadenylation in the coding region of *B.t.* toxin transcripts have been published. In this report, we provide direct evidence demonstrating that the transcript of a *B.t.* toxin gene is polyadenylated in the coding region. Multiple sequence elements in the *crylA(c) B.t.* toxin coding region are recognized by plant cells as polyadenylation signals. Recognition

of these signals is partly responsible for the low accumulation of the *crylA(c)* transcript in plant cells. Elucidating the mechanisms responsible for the low accumulation of *B.t.* toxin mRNA in plants may make it easier to engineer novel foreign genes and other *B.t.* toxin genes for high expression in the future. In addition, it may provide insight into natural gene expression mechanisms in plants.

Another limitation of *crylA(c)* transcript accumulation in plant cells, posed at the level of mRNA stability, is described in another paper I co-authored with Jay De Rocher. In this report, we compared the expression of the AU-rich wild-type *crylA(c) B.t.* toxin gene with the expression of a GC-rich synthetic *crylA(c) B.t.* toxin gene. The transcriptional activities of the genes which have identical 5' and 3' flanking regions were equal in nuclear run-on transcription assays. The mRNA half-life measurements, however, directly demonstrated that the wild-type transcript was less stable than the transcript encoded by the synthetic gene. This work has been accepted for publication, but will not be discussed further in this thesis.

MATERIALS AND METHODS

Plant Materials and Treatment

Nicotiana tabacum cv Bright Yellow 2 [BY-2 (also called NT-1)] cells (An, 1985, Nagata et al., 1992) were cultured as described previously by Newman et al. (1993). Stably transformed cell lines were generated by *Agrobacterium*-mediated transformation also described by Newman et al. (1993) using *Agrobacterium tumefaciens* strain LBA4404 harboring the appropriate plasmids. Kanamycin-resistant BY-2 calli transformed with the *cryIA(c)* gene were transferred to fresh plates and then screened for b-glucuronidase (GUS) reporter gene expression by histochemical staining (Jefferson et al., 1986). Positive cell lines were treated with 50 mg mL^{-1} CHX for 2 hours in liquid culture after five to seven days growth. Treated and untreated cells were frozen in liquid nitrogen after the cells were pelleted at 1000rpm for 5 minutes in a Sorvall RT 6000D centrifuge with an H-1000B rotor. Kanamycin-resistant calli transformed with the globin-*B.t.* chimeric gene were collected in pools of 100 and immediately frozen in liquid nitrogen.

Plasmid Construction

The *cryIA(c)* coding region from *Bacillus thuringiensis* subsp. *kurstaki* HD-73 was kindly provided by Dr. A.I. Aronson of Purdue University. The sequence encoding the N-terminal nine amino acids of *LacZ* was fused to the 5' portion of the coding region encoding the insecticidal domain (amino acids 9-613) (Schnepf and Whiteley, 1985).

The *crylA(c)* coding region was further modified in our laboratory after it was introduced into a modified pT7/T3a19 vector (Gibco BRL, Gaithersburg, MD) containing the pSP64(polyA) multiple cloning site (Promega, Madison, WI). The translation initiation site was altered to conform to the plant consensus sequence (Lütcke et al., 1987; Joshi, 1987) and two proline codons were added to the 3' end of the coding region to protect the carboxyl terminus of the toxin from proteolytic activity (Bigelow and Channon, 1982). The resulting plasmid was named p995. Nucleotide position 415 in the *crylA(c)* sequence file, m11068, of the EMBL/GenBank/DDBJ databases corresponds to nucleotide position 31 in our gene which is the first base of the codon for amino acid 9 of the *crylA(c)* coding region. The *crylA(c)* derivative replaced the b-globin coding region of p1185 (described below) to generate plasmid p1204 after both p995 and p1185 were digested with *BglII* and *Bam*HI. For integration into the genome of BY-2 cells, the gene cassette from p1204 was inserted into the *Hin*dIII site of the binary vector pBI121 (accession number X77672) to make p1205.

p1185 is a pUC 8 plasmid that was generated from a plasmid, pMF6, kindly provided by Michael Fromm at Monsanto Corp. In addition to containing the globin coding region, p1185 also contains a doubly enhanced CaMV 35S promoter and the pea *rbcS-E9* 3' untranslated region (3'UTR). The doubly enhanced promoter was constructed in pMF6 by introducing a second copy of the CaMV 35S enhancer, contained on a *Hin*cII-*Eco*RV fragment, into the *Eco*RV site of a second pMF6 CaMV 35S promoter. The NOS polyadenylation sequence of the modified pMF6 plasmid, now called p1079, was removed by digestion with *Kpn*I and *Sca*I. After blunting the *Kpn*I site of p1079, a blunted *BglII-Sca*I fragment from a second p1079 plasmid was ligated to the vector to

generate the plasmid p1138. The ADH1 intron 1 from the original pMF6 plasmid was

removed from p1138 by digestion with *Eco*RV and *Bam*HI. To replace the region of the

doubly enhanced 35S promoter that was excised with the intron, the *Eco*RV-*Bam*HI

fragment from p1163 was ligated into p1138 to form the plasmid p1166. Finally, the

globin coding region and E9 polyadenylation sequence on a *Bgl*II-*Cla*I fragment from

p977 (described in De Rocher et al., 1997) was inserted into p1166 to generate p1185.

This plasmid, p1185, was digested with *Hind*III to release the gene cassette for insertion

into the *Hind*III site of the binary vector pBI121 to make p1190/p1528.

To construct the globin-*B.t.* chimeric gene, the *Acc*I-*Bam*HI fragment (segment 4)

was excised from the *cryIA(c)* coding region of plasmid p995. After blunting the ends

with T4 DNA polymerase, the fragment was introduced into the unique *Eco*RV site of

p948, a Bluescript II SK(-) vector in which the region of the polylinker between the *Sac*I

and *Cla*I sites was replaced with a polylinker containing the *Bgl*II, *Xba*I, *Eco*RV and

*Bam*HI restriction sites. After digestion with *Bgl*II and *Bam*HI, the DNA fragment was

inserted into the unique *Bam*HI site of p1185 between the globin coding region and the

E9 polyadenylation sequence to give p1188. The gene cassette was then introduced into

the *Hind*III site of the binary vector, pBI121, to make plasmid p1194 for standard

Agrobacterium mediated transformation of tobacco cells.

RNA Methods

RNA was isolated from BY-2 cells as described previously (Newman et al., 1993), except

a phenol/chloroform extraction followed by a chloroform extraction was performed after

solubilization of the lithium chloride pellet. 20μg of total RNA or 2μg of poly(A)$^+$ RNA

was denatured and separated on 2% (v/v) formaldehyde/1% (w/v) agarose gels in 1x

MOPS buffer (20mM 3-[N-morphilino]propanesulfonic acid, 5mM sodium acetate, 1mM

EDTA, 1mg mL^{-1} ethidium bromide) before capillary transfer to BioTrace HP (Gelman

Sciences, Ann Arbor, MI) membrane. Blots were prehybridized and hybridized as

described by De Rocher et al. (1997), except prehybridization was overnight.

Radiolabeled DNA probes were synthesized by the random primed method described in

Feinberg and Vogelstein (1983) using restriction fragments separated on low melting

agarose gels. Probes corresponding to segments 1, 2, and 3 of the *cryIA(c)* coding region

(Figure 2-3) were double labeled with [α-^{32}P]dCTP and [α-^{32}P]dATP. The probe

corresponding to the full-length coding region was labeled with [α-^{32}P]dCTP only. All

labeled probes were separated from unincorporated nucleotides using push columns

(Stratagene, La Jolla, CA). Blots were washed for 30 minutes at 65°C in 2X SSC, 0.1%

(w/v) SDS followed by a wash in 1X SSC, 0.1% (w/v) SDS under the same conditions.

Radioactive bands were detected using a PhosphorImager (Molecular Dynamics,

Sunnyvale, CA).

RNA probes corresponding to the entire E9 3' untranslated region or the *AccI/Bam*HI

segment of the *cryIA(c)* coding region (segment 4) were *in vitro* transcribed using the

Riboprobe System (Promega, Madison, WI) from linearized Bluescript II SK(-) plasmids,

p1425 and p1522, which contain only the E9 and segment 4 sequences, respectively. [α⁻

^{32}P]UTP was used in the labeling reaction and the probes purified with push columns

(Stratagene, La Jolla, CA). Prehybridization of the blots was performed as described

above, however, hybridization with the riboprobe was done overnight at 65°C in

hybridization solution, described in De Rocher et al. (1998), which was modified by

increasing the formaldehyde concentration to 50% (v/v) and decreasing the SSC

concentration to 1x. Blots were washed as described above with an additional wash at

0.2x SSC, 0.1% (w/v) SDS

RT-PCR Analysis

RT-PCR was performed using the 3' RACE System (Gibco BRL, Gaithersburg, MD).

Total RNA isolated from stably transformed cell lines was used for first strand cDNA

synthesis. Gene specific primers (Macromolecular Structure Facility, Michigan State

University) PG-177 (5'-CTCTCAATGGGACGCATTTCTTG-3') which hybridizes to

bases 213-235 relative to the *cryIA(c)* translation initiation site and PG-170 (5'-

CTATCAGAAAGTGGTGGCTGGTGTGGCTAATG-3') which anneals 386-417 bases

from the globin translation start site were used to amplify the *cryIA(c)* and globin-*B.t.*

chimeric transcripts, respectively, in conjunction with the reverse primer PG-192 (5'-

GGCCACGCGTCGACTAGTAC-3') which anneals to the adapter region of the oligo

d(T)$_{17}$ primer supplied with the kit. The PCR of total RNA samples without prior cDNA

synthesis or the size of the amplified products in combination with the presence of a

poly(A) tail at the 3' end of the cDNA clone was used to verify genomic DNA was not

amplified. The amplification protocol was 5 min. at $94^{\circ}C$ followed by 30 cycles of 2 min. at $94^{\circ}C$, 2 min. at $55^{\circ}C$ (for PG-177/PG-192) or 2 min. at $65^{\circ}C$ (for PG-170/PG-192) and 3 min. at $72^{\circ}C$. A 15 min. incubation at $72^{\circ}C$ completed the amplification. The crylA(c) PCR products were digested with XbaI and SalI while the globin-B.t. PCR products were digested with BamHI and SalI. All the PCR products were subcloned into p948, the Bluescript II SK(-) vector described above. Positive clones were identified by gel electrophoresis or by Southern blot using a DNA probe consisting of the crylA(c) coding region. Cycle sequence analysis identified the poly(A) addition sites (Plant Biochemistry Facility, Michigan State University).

Oligo-directed RNase H Cleavage Analysis

Approximately $2\mu g$ of the oligonucleotides (Macromolecular Structure Facility, Michigan State University) PG-229 (5'GAGCAACGATATCTAATAC-3') which hybridizes upstream of the segment 3 poly(A) site (+721 to +739) and PG-234 (5'CTGGGTTTGTATAAATTTCTC-3') which hybridizes downstream of the segment 3 poly(A) site (+797 to +817) were annealed to $20\mu g$ of total RNA isolated from CHX treated tobacco cells. Hybridization was performed in a 400mL $65^{\circ}C$ water bath that was allowed to cool to room temperature. Afterwards, the RNA samples were incubated in 4mM Tris-HCl pH 8, 10mM $MgCl_2$, 20mM KCl, 1mM DTT and one unit of RNase H (Gibco BRL, Gaithersburg, MD) for one hour at $37^{\circ}C$. For removal of the poly(A) tail, cleavage experiments were performed using $0.5\mu g$ oligo $d(T)_{12-18}$ and $10\mu g$ total RNA. RNase H cleavage experiments were also performed using $3.5\mu g$ poly(A)$^{+}$ RNA from

untreated tobacco cells annealed to 0.3μg of either PG-229 or PG-234. The hybrids were

incubated in 20mM Tris-HCl pH 7.5, 10mM $MgCl_2$, 100mM KCl, and 5% (w/v) sucrose

with 3 units of RNase H for 60 min. at 37^0C. All samples were alcohol precipitated and

resuspended in loading buffer prior to their electrophoresis through a 2% (v/v)

formaldehyde/1.7% (w/v) agarose gel in 1x MOPS buffer (described above). The gel was

blotted and the membrane probed with the full-length *cryIA(c)* coding region as described

above.

RESULTS

Low Accumulation of *B.t.* toxin Transcripts and the Detection of Short, Polyadenylated

Transcripts in Tobacco Cells

The gene encoding the *crylA(c)* protoxin was among the first *B.t.* toxin genes to be

isolated from *Bacillus thuringiensis* (Adang et al., 1985). Consequently, the gene has

been extensively characterized and, because of its agricultural potential, previously

introduced into plants (Adang et al., 1987; Murray et al., 1991). As observed with the

transcripts of other *B.t.* toxin genes, the *crylA(c)* transcript does not accumulate to

detectable levels in mature tobacco plants (Murray et al. 1991) making the *crylA(c)* gene

a good candidate for further investigation of factors limiting *B.t.* toxin gene expression.

Figure 2-1A shows the derivative of the *crylA(c)* gene that was used in this study. The

coding region consists of only the insecticidal domain which is located in the 5' half of

the protoxin gene (Schnepf and Whiteley, 1985). The 3' portion of the gene is

dispensable (Adang et al., 1985) and most plant transformations with *B.t.* toxin genes

utilize 3' truncations which contain only the insecticidal domain. As shown in Figure 2-

1A, transcription of the gene used in this study is controlled by a modified 35S promoter

containing a duplicated enhancer region (2x35S) and the polyadenylation signal is

provided by the well characterized pea *rbcS-E9* 3' untranslated region (UTR) (Hunt and

MacDonald, 1989; Mogen et al., 1990; Mogen et al., 1992; Li and Hunt, 1995).

Figure 2-1. Structure of the Genes Stably Introduced into Tobacco Cells.

(**A**) The portion of the wild-type *Bacillus thuringiensis* var. *kurstaki crylA(c)* gene
encoding the insecticidal domain (amino acids 9-613; Schnepf et al., 1985) used in this
study. Transcription of the chimeric gene was controlled by the Cauliflower Mosaic
Virus 35S promoter which was modified by duplicating the upstream enhancer region (2
x 35S). The pea *rbcS-E9* 3' UTR (E9) provides the elements necessary for
polyadenylation.

(**B**) Chimeric globin-*B.t.* toxin gene used to identify the polyadenylation site in segment 4
of the *crylA(c)* coding region. A 650 bp AccI-BamH1 restriction fragment of the *crylA(c)*
coding region (segment 4) was inserted between a b-globin reporter gene under the
control of the modified 2 x 35S promoter and the E9 3' UTR. Utilization of any poly(A)
addition sites within the segment 4 insert will result in polyadenylated transcripts that
lack the E9 3' UTR sequences.

A

2 x 35S	*B.t.* toxin	E9

B

Segment 4

2 X 35S	Globin	E9
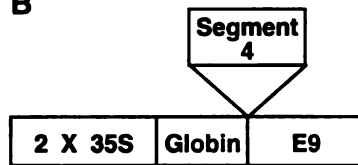
Tobacco cells stably transformed with this *crylA(c)* derivative were analyzed for

expression on RNA gel blots. Figure 2-2A, lane 1 shows the low accumulation of the

full-length *crylA(c)* transcript (i.e. transcripts terminating in the E9 3' UTR) in one cell

line. This line was selected because of its relatively high expression level. The full-

length transcript could not be visualized in most of the stably transformed tobacco cell

lines that were examined, presumably because the transcript is below the level of

detection (data not shown). This is despite the fact that transcription is directed by the

doubly enhanced 35S promoter. Run-on transcription experiments have shown that the

gene is efficiently transcribed in nuclei isolated from stably transformed cells, arguing

against the possibility that the *crylA(c)* coding region contains a repressor sequence

capable of inhibiting transcription initiation or elongation (De Rocher et al., 1997). The

discrepancy between the RNA gel blot and the run-on transcription results suggests the

mechanisms responsible for the low abundance of the *crylA(c)* transcript in plant cells are

post-transcriptional.

Figure 2-2A, lane 2 shows the poly(A)$^+$ RNA fraction from the same cell line

shown in Figure 2-2A, lane 1. The full-length *crylA(c)* transcript can be easily detected in

this fraction. Surprisingly, two short transcripts of 900 nt and 600 nt can also be detected.

The abundance of these two transcripts can be significantly increased by treating the cell

line with the translation inhibitor, cycloheximide (CHX) (compare Figure 2-2A, lanes 1

and 3 as well as lanes 2 and 4). CHX often increases the abundance of unstable

transcripts because many of these transcripts are degraded in a translation-dependent

manner. This suggests that the 900 and 600 nt transcripts are unstable in plant cells. The

full-length *crylA(c)* transcript does accumulate slightly on CHX treatment (compare

Figure 2-2. Poor Accumulation of *cryIA(c)* mRNA and the Detection of Short, Polyadenylated Transcripts in Tobacco Cells.

(A) Twenty micrograms of total RNA (T) from stably transformed tobacco cells was isolated from a pool of kanamycin resistant calli growing in liquid culture and electrophoresed next to poly(A)$^+$ RNA (A) isolated from the same cells. Cell cultures were either treated (+) or not treated (-) with cycloheximide (CHX). The RNA gel blot was probed with a 730bp DNA fragment corresponding to the 5' portion of the *cryIA(c)* coding region. The autoradiograph was overexposed to show the *cryIA(c)* transcripts in tobacco cells not treated with CHX. This panel was contributed by Wan-Ling Chiu.

(B) Detection of poly(A) tails on the short *cryIA(c)* transcripts. Twenty micrograms of total RNA isolated from cycloheximide treated tobacco cells was incubated with oligo d(T)$_{18}$ in the presence (+) or absence (-) of RNase H. Hybridization of the oligonucleotides to the poly(A) tail results in cleavage of the poly(A) tail by RNase H. As a consequence, the transcripts will have an increased mobility in an RNA gel. The RNA gel blot was hybridized with a probe for the *cryIA(c)* coding region. Two cell lines (A and B) which do not accumulate the full-length *cryIA(c)* transcript upon CHX treatment were used in this experiment. This panel was contributed by Wan-Ling Chiu.

A

Lane: 1 2 3 4

◄Full-length

• 900nt

• 600nt

RNA: T A T A
CHX: - - + +

B

Cell Line: A B

• 900nt

• 600nt

RNase H: - + - +
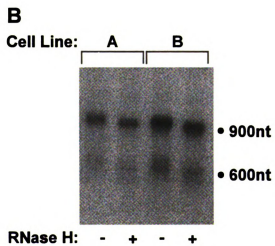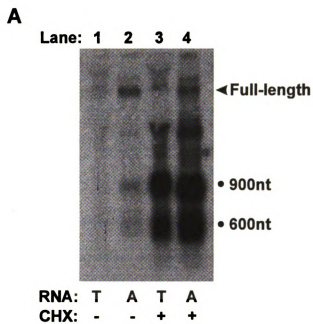
Figure 2-2A, lanes 1 and 3) which is consistent with the notion that the *cryIA(c)* transcript

is rapidly degraded. However, the most prominent effect is on the accumulation of the

900 and 600 nt transcripts. Nearly all of the stably transformed cell lines analyzed

accumulate the two short transcripts, even when the full-length *cryIA(c)* transcript is

below the limit of detection (Figure 2-2B).

Characterization of the Short *B.t.* toxin Transcripts

The results of the CHX induction experiments were consistent with instability of the

cryIA(c) transcripts in tobacco cells. Accordingly, the 900 and 600 nt transcripts could be

degradation intermediates that were stabilized by the CHX treatment. Alternatively, they

could be unstable products of cellular processes such as splicing or polyadenylation. To

distinguish between these possibilities the 900 and 600 nt transcripts were characterized

further. Figure 2-2A, lane 4 shows the presence of the short transcripts as well as the full-

length transcript in the poly(A)$^+$ fraction from CHX treated cells. This indicates the 900

and 600 nt transcripts are polyadenylated. Importantly, the 900 and 600 nt transcripts

also are present in the poly(A)$^+$ RNA fraction from tobacco cells that were not treated

with CHX (Figure 2-2A, lane 2).

The existence of a poly(A) tail on the short transcripts was verified in Figure 2-2B

using oligo-directed RNase H cleavage. RNase H cleaves the RNA strand of an

RNA/DNA hybrid. Therefore, annealing oligo d(T) to the poly(A) tail will result in

cleavage of the tail by RNase H. The removal of the poly(A) tail can be detected by an

increased mobility of the transcripts on an RNA gel blot. As shown in Figure 2-2B, both

the 900 and 600 nt transcripts decrease in size upon treatment with oligo d(T) and RNase

H, consistent with the removal of the poly(A) tail.

The RNA gel blot in Figure 2-2 was probed with the first 730 bases of the *cryIA(c)*

coding region. Detection of the 900 and 600 nt transcripts with this probe and the fact

that these transcripts are polyadenylated argues against them resulting from degradation

of the full-length transcript. Instead, the data support splicing or polyadenylation within

the coding region as the mechanism responsible for the formation of the short transcripts.

To further delineate what sequences are present in the 900 nt and 600 nt transcripts,

the *cryIA(c)* coding region was divided into four segments using convenient restriction

sites as shown in Figure 2-3A. Each segment was then used to generate probes to

hybridize against poly(A)$^+$ RNA isolated from CHX treated cells producing the full-

length and both short transcripts (Figure 2-3B). All of the probes hybridize to the full-

length transcript as expected. However, the 600 nt transcript hybridizes with only the

segment 1 and 2 probes while the 900 nt transcript hybridizes with the segment 1, 2, and

3 probes. Hybridization is detected in the 900 nt region with the segment 4 probe.

However, this is also evident in the untransformed control and therefore, does not

correspond to the 900 nt *cryIA(c)* transcript. Neither the 600 nt nor 900 nt transcripts

hybridize with the probe spanning the E9 3' UTR (last panel of Figure 2-3B). This

indicates the poly(A) tail is attached directly to the *cryIA(c)* coding region. The simplest

explanation of these data is that sequences within the *cryIA(c)* coding region are

recognized as polyadenylation signals. The 600 nt transcript is consistent with

polyadenylation within segment 2 of the *cryIA(c)* coding region and the 900 nt transcript

Figure 2-3. Characterization of the Short *cryIA(c)* Transcripts by Hybridization with Different Segments of the Coding Region.

(A) The *cryIA(c)* coding region was divided into four segments using convenient restriction sites. Segment 1 consists of an SphI-XbaI restriction fragment from position 28 to position 295. Segment 2 is a restriction fragment extending from position 295 to position 670 and is defined by XbaI sites at each end. Segment 3 is an XbaI-AccI restriction fragment from position 670 to position 1201 and segment 4 is the remainder of the coding region from the AccI site at position 1201 to the BamH1 site at position 1851.

(B) An RNA gel blot consisting of alternating lanes of poly(A)$^+$ RNA (2μg/lane) from stably transformed tobacco cells (left lane) or untransformed cells (right lane) was sectioned and hybridized with a probe for the entire coding region or with the individual segments of the *cryIA(c)* coding region (designated below each panel). The arrowhead indicates the full-length transcript while the dark circles indicate the 900 and 600 nt transcripts. The box diagrams indicate the segments of the *cryIA(c)* coding region found in each transcript.
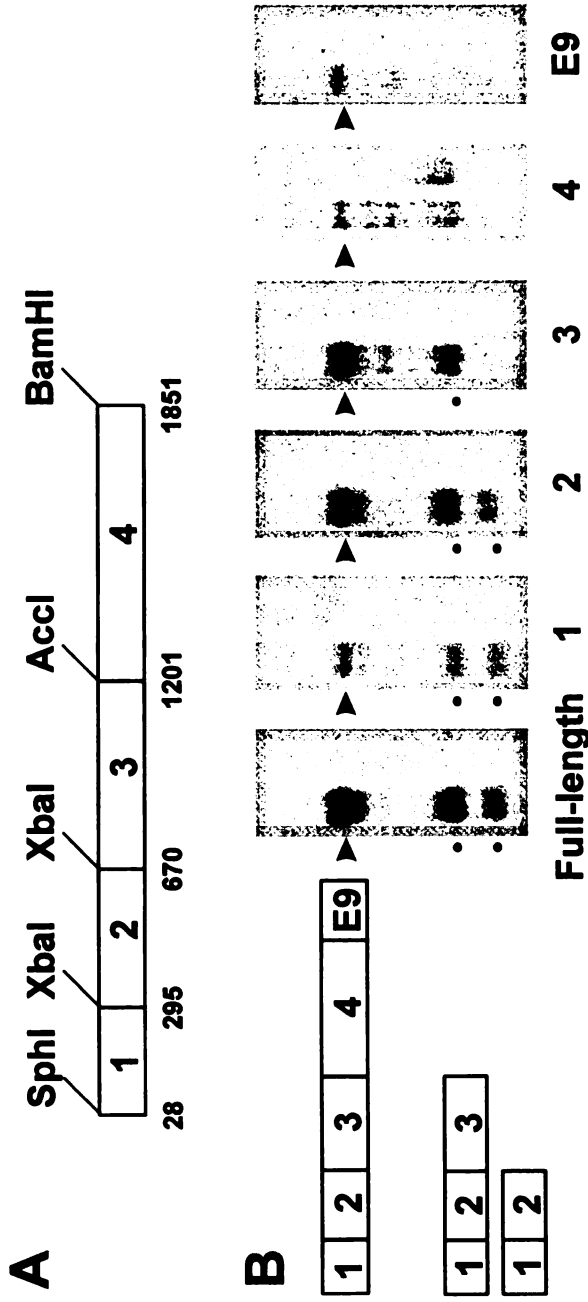
is consistent with polyadenylation in segment 3. Other minor transcripts can be observed hybridizing to some of the *crylA(c)* probes (and in Figure 2-2A), however, these transcripts are not consistently found in the stably transformed cell lines in contrast to the 900 and 600 nt transcripts. Thus, these transcripts were not pursued further.

Identification of Polyadenylation Sites Within the *B.t.* toxin Coding Region

If tobacco utilizes polyadenylation sites within segments 2 and 3, then it should be possible to determine the exact sites where the poly(A) tail is added using RT-PCR. To this end, total RNA from CHX treated cells was reverse transcribed using an oligo d(T)-adapter primer. Aliquots of the cDNA were used as template for PCR with primers which hybridize to the 3' portion of segment 1 and to the adapter region at the 5' end of the oligo d(T) primer. The major products of these reactions were consistent in size with RNAs polyadenylated about 180 bases into segment 2 and about 120 bases into segment 3 (data not shown). These products were subcloned into a Bluescript vector and sequenced. As diagrammed in Figure 2-4, the shorter of the two PCR products mapped to nucleotide position 479 in segment 2 and the longer PCR product mapped to position 787 in segment 3. These sites are consistent with the sizes of the PCR products and the short transcripts in Figure 2-2 when assuming a poly(A) tail of 110-120 bases.

The difference in size between the 900 and 600 nt transcripts compared to the positions of the poly(A) addition sites is likely due to the length of the poly(A) tail. To test this possibility and confirm the segment 3 polyadenylation site identified by RT-PCR

corresponds to the same polyadenylation site used to generate the 900 nt *in vivo* transcript, oligo-directed RNase H cleavage analysis was performed. A DNA oligonucleotide hybridizing to an mRNA upstream of the poly(A) site should direct cleavage of that RNA by RNase H, whereas an oligonucleotide hybridizing downstream should not. As shown in Figure 2-5A, an oligonucleotide that hybridizes starting 65 bases upstream of the segment 3 polyadenylation site directs cleavage of the full-length and 900 nt transcripts in the presence of RNase H. The mobility of the 600 nt transcript, which lacks sequences complementary to the oligonucleotide, is unaltered in the presence of RNase H, as expected.

The 180 base decrease in the size of the 900 nt transcript to approximately 720 nt in the presence of RNase H (Figure 2-5A) indicates the poly(A) tail is about 115 bases long. Sixty-five bases of the 180 base difference are accounted for by the sequences to which the oligonucleotide anneals as well as the distance between the oligonucleotide and the poly(A) addition site. A poly(A) tail length of 115 bases corresponds well with the 110-120 base poly(A) tail predicted from the discrepancy between the position of the segment 3 polyadenylation site and the *in vivo* size of the transcript. The absence of the 3' end of the full-length transcript on the RNA gel blot in Figure 2-5A is most likely a result of its degradation by RNase H due to partial sequence complementarity between the oligonucleotide and this region of the transcript. When RNase H cleavage is less efficient the 3' end can be detected as shown in Figure 2-5C.

Similar RNase H cleavage experiments were carried out using an oligonucleotide which hybridizes starting approximately 10 bases downstream of the polyadenylation site mapped in segment 3 (Figure 2-5B). This oligonucleotide should not anneal to the 900 nt

Figure 2-4. Identification of Two Poly(A) Addition Sites within the *cryIA(c)* Coding Region.

cDNA was synthesized from total RNA isolated from stably transformed tobacco cells and amplified by PCR. The resulting PCR products were cloned and sequenced to determine the polyadenylation sites that are indicated.

$C_{(479)}AA_{(n)}$ $C_{(757)}AA_{(n)}$

1	2	3	4

28   295   670   1201   1851
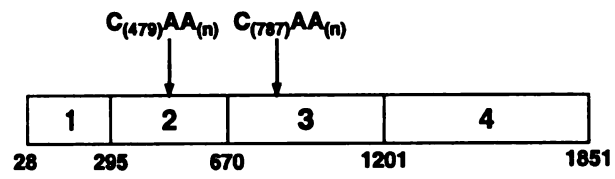
Figure 2-5. The Mapped Poly(A) Addition Site in Segment 3 Corresponds to the Polyadenylation Site of the 900 nt *in vivo* Transcript.

Total RNA from CHX treated tobacco cells expressing the *cryIA(c)* gene was incubated with oligonucleotides which hybridize either upstream (panel A) or downstream (panel B) of the segment 3 polyadenylation site identified by RT-PCR. Poly(A$^+$) RNA from transgenic tobacco cells not treated with CHX also was incubated with the upstream and downstream oligonucleotides (panel C). Incubations were performed in the presence (+) or absence (-) of RNase H. If sequences complementary to the oligonucleotides are present in the 900 nt transcript, RNase H will cleave the RNA strand of the RNA/DNA duplex which will result in a band shift on the RNA gel blots.

A

Upstream
Primer

◀ Full-length

● 900 nt
● 600 nt

RNase H: − +

B

Downstream
Primer

◀ Full-length

● 900 nt
● 600 nt

RNase H: − +

ated
nel B)

n and
e (+)

NA

C

Upstream
Primer
Downstream
Primer

◀ Full-length
] Full-length 3' ends
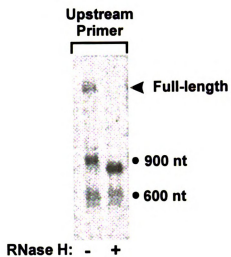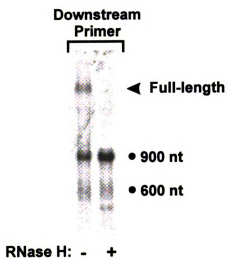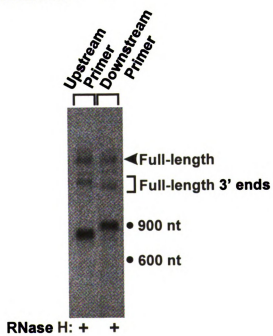
● 900 nt
● 600 nt

RNase H: + +

transcript unless it actually extends beyond the mapped poly(A) site. Figure 2-5B shows

the size of the 900 nt transcript was not altered after incubation with the oligonucleotide

and RNase H, demonstrating sequences immediately downstream of the segment 3

polyadenylation site are not present in the 900 nt transcript. Cleavage experiments using

poly(A)$^+$ RNA from untreated tobacco cells instead of total RNA from CHX treated cells

show the same results with both the upstream and downstream oligonucleotides (Figure

2-5C). These data indicate the segment 3 polyadenylation site identified by RT-PCR is

the same site used to generate the 900 nt transcript. More importantly, they also indicate

the same polyadenylation site is utilized in both CHX treated and untreated transformed

tobacco cells. RNase H cleavage experiments also have been performed to confirm the

segment 2 polyadenylation site. The diffuse nature of the 600 nt transcript makes it

difficult to access the shift in bands precisely. However, an oligonucleotide

complementary to a region upstream of the cleavage site was able to shift the 600 nt

transcript to a smaller size while an oligonucleotide hybridizing downstream of the

cleavage site did not appear to decrease the size of the transcript (data not shown).

Sequences Typical of Plant Polyadenylation Signals Are Present Upstream of the
Identified Poly(A) Sites

The sequences upstream of the poly(A) addition sites were examined for similarities

to known plant polyadenylation signals. Unlike mammalian poly(A) signals, plant

poly(A) signals do not have a strict consensus sequence requirement for AAUAAA

upstream of the cleavage site as shown in Figure 2-6A. In addition, plants do not require

a *cis*-regulatory element downstream of the cleavage site as in animal systems. However,

plant polyadenylation signals do require two *cis*-regulatory elements: the Far Upstream

Element (FUE) and the Near Upstream Element (NUE) (reviewed in Hunt, 1994; Wu et

al., 1995; Rothnie, 1996). There is no known consensus sequence for either element, but

each has key sequence characteristics based on nucleotide composition.

Located approximately 40 to 150 bases upstream of the cleavage site, the FUE is

required for efficient 3' end formation. The most common motif that is evident among

known FUEs is the presence of multiple UG-rich regions (reviewed in Hunt, 1994; Wu et

al., 1995; Rothnie, 1996). Several UG-rich stretches can be found upstream of the

segment 2 and segment 3 poly(A) addition sites in positions that correspond to a putative

FUE (Figure 2-6B).

The NUE is an AU-rich element typically found ten to thirty bases upstream of the

cleavage site. These elements are essential for polyadenylation and control poly(A)

addition at specific cleavage sites. Thus, a plant transcript with multiple polyadenylation

sites will have an NUE corresponding to each site. NUEs can contain the mammalian

canonical AAUAAA sequence, as in the case of the CaMV polyadenylation signal

(Sanfacon et al., 1991; Rothnie et al., 1994), but more often an AAUAAA-like sequence

is present in which 1 or 2 of the bases do not match (reviewed in Hunt, 1994; Wu et al.,

1995; Rothnie, 1996). Upstream of both poly(A) addition sites in the *B.t.* toxin coding

region, an AAUAAA-like sequence typical of plant polyadenylation signals can be

identified. A comparison of these sequences to other known NUEs revealed sequence

similarity (Figure 2-6B).

Figure 2-6. The *cryIA(c)* Coding Region Contains Elements Characteristic of Plant Polyadenylation Signals.

(A) A schematic representation comparing the structure of typical plant and mammalian polyadenylation signals. Sequence motifs characteristic of the plant far-upstream element (FUE) and near-upstream element (NUE) as well as the mammalian downstream element (DSE) are indicated. The poly(A) addition sites are represented by the arrows. Plants can utilize multiple poly(A) addition sites downstream of specific NUEs within a transcript. The cleavage of a plant transcript usually occurs at a pyrimidine/adenosine dinucleotide. Mammalian transcripts are usually cleaved at a single site corresponding to a CA dinucleotide.

(B) Identification of elements characteristic of plant polyadenylation signals upstream of the poly(A) addition sites in the *cryIA(c)* coding region. The putative plant polyadenylation signals in segments 2, 3, and 4 of the *cryIA(c)* coding region were compared to the most commonly used polyadenylation sites in the *rbcS-E9* and octopine synthase (*ocs*) genes. The position of the FUE and NUE relative to the cleavage site (CS) are indicated.

A


```
                                    YA
                                 ↓ ↓ ↓
  ─────┌──────────┐───┌────────┐──────────────── plants
       │   FUE    │   │ NUE(S) │
       └──────────┘   └────────┘
       (UG-rich)      (AAUAAA-like)

                                    CA
                                    ↓
  ──────────────────────┌────────┐──────┌─────┐──── mammals
                         │ AAUAAA │      │ DSE │
                         └────────┘      └─────┘
                                         (U/GU-rich)
```
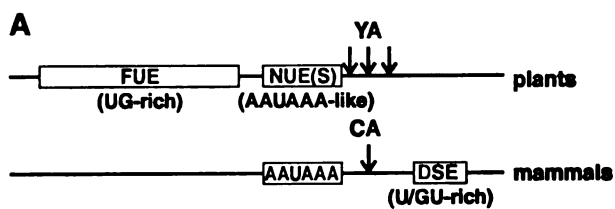
B

	FUE	NUE	CS
rbcS-E9 Site 2	-160 ┌────────────────┐ -40 UG-rich Sequences	-25 AAUGAA	CA
ocs Site 1	ND ┌────────────────┐ -83 UG-rich Sequences	-17 AAUAAU	CA
B.t. toxin Segment 2	-139 ┌────────────────┐ -27 UG-rich Sequences	-23 AAUUAU	CA
B.t. toxin Segment 3	-135 ┌────────────────┐ -45 UG-rich Sequences	-40 AAUUAU	CA
B.t. toxin Segment 4	-132 ┌────────────────┐ -25 UG-rich Sequences	-19 AAUAAU	GA

Cleavage of the *cryIA(c)* transcript in both segments 2 and 3 occurs at a CA

dinucleotide (Figure 2-6B). This is consistent with other known plant poly(A) addition

sites. Again, no strict consensus sequence is known which defines the cleavage site in

plants. However, cleavage typically occurs at a pyrimidine/adenosine dinucleotide in

plant transcripts (reviewed in Hunt, 1994; Wu et al., 1995; Rothnie, 1996).

Further sequence analysis of the *cryIA(c)* gene reveals the presence of other possible

plant polyadenylation signals in the coding region. In particular, segment 4 (Figure 2-3A)

contains sequences that resemble plant polyadenylation signals. Transcripts terminating

in this segment were not detected on RNA gel blots or by RT-PCR using total RNA from

CHX treated cells as template. However, these transcripts may be produced in very small

amounts if a substantial portion of the transcripts is polyadenylated in segments 2 and 3.

To determine whether segment 4 of the *cryIA(c)* coding region contains a functional plant

polyadenylation signal, a chimeric gene was constructed which consists of the segment

inserted between a doubly enhanced 35S driven b-globin reporter gene and the E9 3' UTR

(Figure 2-1B). A polyadenylated transcript terminating in segment 4 would result if a

polyadenylation signal exists in the segment. Otherwise, the transcript would be

polyadenylated at the E9 poly(A) sites. The RNA gel blot in Figure 2-7A shows stably

transformed tobacco cells expressing the globin-*B.t.* gene accumulate only a small

amount of transcript at a position consistent with termination in the E9 region. Most of

the globin-*B.t.* transcripts in these cells accumulate as discrete bands at sizes that are

more consistent with termination in segment 4. Hybridization with the E9 3' untranslated

region shows these abundant transcripts lack the E9 region (Figure 2-7A). Similar

Figure 2-7. Identification of a Third Polyadenylation Site in the *crylA(c)* Coding Region.

(**A**) Truncated transcripts are produced in tobacco cells expressing a chimeric globin-*B.t.* toxin gene containing bases 1201 to 1851 of the *crylA(c)* coding region. An RNA gel blot of total RNA extracted from two pools of 100 tobacco calli stably transformed with the globin-segment 4 chimeric gene (4) was probed with either the globin coding region (left panel) or the E9 3' UTR (right panel). Total RNA from tobacco calli expressing a gene containing the globin coding region and *rbcS-E9* 3' UTR, but lacking the segment 4 insert, was used as a control (0). The arrows indicate the position of the transcripts terminating in the E9 3' UTR while the bracket marks the position of the globin-*B.t.* toxin transcripts lacking E9 3' UTR sequences. The autoradiographs were overexposed to show the globin-segment 4 transcripts terminating in the E9 3' UTR. This panel was contributed by Wan-Ling Chiu.

(**B**) The short globin-*B.t.* toxin chimeric transcripts are polyadenylated. Total RNA isolated from transgenic tobacco cells expressing the control gene (0) or the globin-segment 4 chimeric gene (4) was incubated with oligo d(T)$_{18}$ in the presence (+) or absence (-) of RNase H. After electrophoresis, the RNA was blotted and probed with the globin coding region. The arrows show the position of the transcripts terminating in the E9 3' UTR and the bracket indicates the position of the globin-*B.t.* toxin transcripts that lack E9 3' UTR sequences. This panel was contributed by Wan-Ling Chiu.

A

0   4   0   4

Globin probe   E9 probe

B

0   4
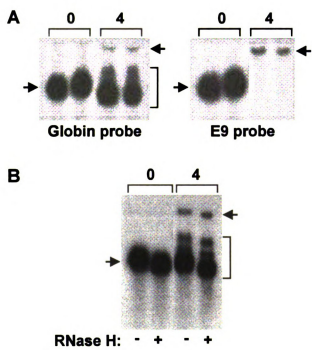
RNase H: −   +   −   +

transcript patterns were reproducibly observed in transgenic tobacco plants and in

protoplasts transiently expressing the gene (data not shown).

The presence of a poly(A) tail on the short transcripts was determined using RNase

H and oligo d(T). As shown in Figure 2-7B, the short transcripts decrease in size in the

presence of RNase H indicating they are polyadenylated. These data are supported by

RT-PCR analysis which identified a poly(A) addition site 201 bases into segment 4. This

site is located 7 bases downstream of a putative NUE in segment 4 (Figure 2-6B). Taken

together, these data show segment 4 of the *cryIA(c)* coding region does contain sequences

which function as polyadenylation signals in plants.

DISCUSSION

The goal of this study and the one that follows was to determine what processes play a role in limiting the accumulation of the *cryIA(c)* *B.t.* toxin transcript in plants. Elucidating the mechanisms responsible for the lack of accumulation of this transcript may make it easier to resynthesize novel *B.t.* toxin genes, but more importantly, it may provide an understanding as to why the transcripts of some foreign genes fail to accumulate in plants. In this report we have demonstrated that the *cryIA(c)* coding region contains multiple sequence elements which are recognized by plant cells as polyadenylation signals. Utilization of these polyadenylation signals is at least partially responsible for the low accumulation of the *cryIA(c)* transcript in plants.

To the best of our knowledge, this study is the first to show that sequences within the coding region of a foreign gene can be recognized as polyadenylation signals by plants. It had been previously suggested that the *cryIA(c)* coding region contains plant polyadenylation signals (Adang et al., 1987). Tobacco plants expressing a *cryIA(c)* protoxin gene or a 3' truncated version produced a polyadenylated 1.7 kb transcript. However, the transcript disappeared as the plants matured. In addition, the transcript was not observed in the progeny of these plants (Murray et al., 1991). Therefore, the mechanism responsible for the production of the 1.7 kb transcript was not elucidated and the lack of full-length transcript could not be explained. In another study, transcripts of 1.6 and 0.9 kb were detected in the poly(A)$^{+}$ RNA fractions of plants expressing a

cryIA(b) B.t. toxin gene. However, these transcripts were believed to be degradation intermediates (Murray et al., 1991).

In this study, nearly every tobacco cell line stably transformed with the *cryIA(c) B.t.* toxin gene accumulate two polyadenylated transcripts of 900 nt and 600 nt in length. Both transcripts hybridize to probes corresponding to the 5' end of the coding region with the 900 nt transcript containing segments 1, 2 and 3 of the *cryIA(c)* coding region and the 600 nt transcript containing just segments 1 and 2. This indicated the two short transcripts result from polyadenylation within the *cryIA(c)* coding region. RT-PCR analysis identified two specific poly(A) addition sites. This corroborated the hybridization data as one poly(A) addition site was mapped to segment 2 and a second poly(A) addition site was mapped to segment 3 of the *cryIA(c)* coding region. Polyadenylation at these sites results in the formation of the 600 nt and 900 nt transcripts, respectively.

It is likely polyadenylation at these sites plays a significant role in limiting the accumulation of the full-length *cryIA(c)* transcript in tobacco cells. Although the 900 and 600 nt transcripts could not be detected in total RNA preparations, the abundance of these transcripts could be increased after treating the cells with CHX to a point where they were the most abundant *cryIA(c)* transcripts in both the total and poly(A)$^+$ RNA fractions. In addition, these two transcripts accumulate to approximately the same level as the full-length transcript in the poly(A)$^+$ RNA fractions of untreated tobacco cells and can be detected by RNA gel blot analysis without the use of more sensitive techniques such as RT-PCR or RNase protection analysis. But, polyadenylation is not the only mechanism limiting the accumulation of the full-length *cryIA(c)* transcript in plant cells. The

increase in abundance of the *cryIA(c)* full-length and short transcripts upon CHX

treatment is consistent with *B.t.* toxin transcript instability, as suggested previously

(Fischhoff et al., 1987; Vaeck et al., 1987), since CHX induced accumulation is

characteristic of many unstable transcripts. For those transcripts that must be translated

to be rapidly degraded, CHX is presumed to block translation in *cis*, thereby blocking

degradation. Alternatively, CHX can inhibit mRNA degradation in *trans* by blocking

translation of a labile mRNA degradation factor. It seems likely that one or both of these

mechanisms explain the enhanced accumulation of *cryIA(c)* transcripts upon CHX

treatment because in De Rocher et al., 1998, direct evidence is presented showing the

cryIA(c) transcripts are inherently unstable in plant cells.

The high AT content of *B.t.* toxin genes raises the possibility that regions of the

cryIA(c) mRNA are recognized as introns in plant cells. The presence of spliced *B.t.*

toxin transcripts was recently reported in tobacco cells expressing a *cryIA(b)* gene (Van

Aarssen et al., 1995). Our results demonstrate the 900 and 600 nt transcripts are not a

result of splicing. The hybridization data, the RT-PCR analysis, and the RNase H

experiments are consistent with the conclusion that these two transcripts are a result of

polyadenylation in the *cryIA(c)* coding region. This is not to suggest splicing of the

cryIA(c) transcript does not occur in tobacco cells. Splicing could account for some of

the minor transcripts hybridizing to the various *cryIA(c)* probes observed on our RNA gel

blots. However, most of these transcripts were not reproducibly detected in our

transformed tobacco cell lines and therefore, were not pursued further.

The presence of plant poly(A) addition sites within the *cryIA(c)* coding region raises

the question of whether other closely related *B.t.* toxin genes might contain plant

polyadenylation signals within their coding regions. The *cryIA(c)* gene belongs to one of

six classes of *B.t.* toxin genes, the *cryI* class (see Höfte and Whiteley, 1989; Feitelson et

al., 1992 for a detailed description of the different classes). These genes share significant

nucleotide sequence identity with each other and encode insecticidal proteins which are

active against the insect Order Lepidoptera. A subclass of the *cryI* genes, the *cryIA*

genes, contains members which are more than 80% identical at the nucleotide level. The

relationships among these genes, designated *cryIA(a)*, *cryIA(b)*, *cryIA(c)*, and *cryIA(d)*,

are shown in the dendrogram in Figure 2-8. Alignment of the *cryIA(c)* and *cryIA(b)*

coding regions shows that a region extending 200bp upstream of the segment 2

polyadenylation site is identical between the two genes (data not shown). This 200bp

region should be of sufficient length to contain the elements necessary for

polyadenylation in plants. A similar alignment with the *cryIA(a)* gene shows the same

segment 2 poly(A) signal is probably common to this gene as well (data not shown). The

cryIA(d) coding region has 95% nucleotide identity over a region spanning the putative

segment 2 polyadenylation signal (data not shown). Approximately the same sequence

identity over the region containing the segment 3 polyadenylation signal can be observed

for the *cryIA(a)*, *cryIA(b)*, and *cryIA(d)* genes. One would expect, therefore, that the

same poly(A) sites are used in the *cryIA(a-d)* coding regions, although this remains to be

proven. The other genes in Figure 2-8 share lower levels of nucleotide identity in the

region of the poly(A) signals (eg. 57-87% for the segment 2 poly(A) signal) which could

affect poly(A) signal recognition, so it is difficult to predict if the same poly(A) addition

sites are used in these genes.

Figure 2-8. *cry* Genes with High Sequence Similarity to the *cryIA(c)* Coding Region.

The nucleotide sequence of several *cry* coding regions were compared to the *cryIA(c)* coding region using the GCG pileup program. Some of the genes most closely related to the *cryIA(c)* coding region are shown. The percent identity of each gene compared to the *cryIA(c)* coding region was calculated using the GCG gap function. The accession numbers are *cryIA(a)*, m11250; *cryIA(b)*, m13898; *cryIA(c)*, m11068; *cryIA(d)*, m73250; *cryIE(a)*, m73252; *cryIF*, m73254; *cry ID*, x54160; and *PrtA*, z22510.

n.

c)
ed to
o the

250;

% identity
to *crylA(c)*

crylA(a)	88
crylA(b)	90
crylA(d)	82
crylA(c)	-
crylE(a)	77
crylF	76
crylD	76
Prt A	73
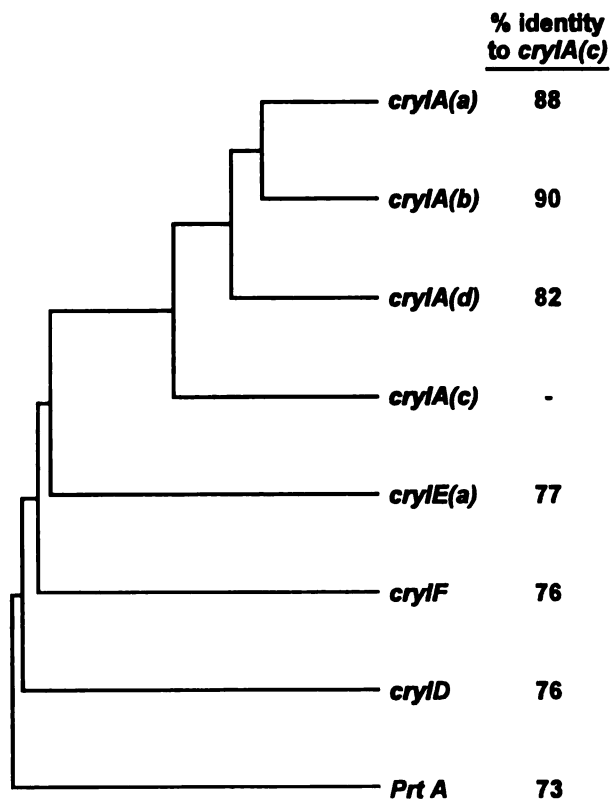
B.t. toxin transcripts are not likely to be the only foreign transcripts that are prematurely polyadenylated in plants. Although there are no other documented cases yet, other examples are expected to arise as the expression of other problematic foreign genes is investigated. This contention is supported by the finding that even a transcript normally produced in one plant species can be differentially polyadenylated when it is transcribed in another plant. Specifically, the maize Activator (*Ac*) transposase transcript is polyadenylated at four sites within a 200 bp region of exon 2 when it is expressed in Arabidopsis plants (Jarvis et al., 1997; Martin et al., 1997). Recognition of these poly(A) addition sites has been suggested to contribute to the low abundance of correctly processed tranposase transcripts and hence the low frequency of transposition in this plant species. The low accumulation of the T4 lysozyme and *Klebsiella pneumoniae* cyclodextrin glycosyltransferase transcripts in potato plants also may be a result of premature poly(A) addition sites. These genes have a high A/U bias like *B.t.* toxin genes. Currently, it is not possible to predict which putative polyadenylation signals will be recognized in plants strictly on the basis of sequence analysis. Nevertheless, as more poly(A) signals are scrutinized and the mechanisms by which they are recognized are elucidated, designing an algorithm to achieve this goal may indeed be feasible.

ACKNOWLEDGMENTS

REFERENCES

Adang MJ, Staver MJ, Rocheleau TA, Leighton J, Barker RF, Thompson DV (1985) Characterized full-length and truncated plasmid clones of the crystal protein of *Bacillus thuringiensis* subsp. *kurstaki* HD-73 and their toxicity to *Manduca sexta*. Gene 36: 289-300

Adang MJ, Firoozabady E, Klein J, DeBoer D, Sekar V, Kemp JD, Murray E, Rocheleau TA, Rashka K, Staffeld G, Stock C, Sutton D, Merlo DJ (1987) Expression of a *Bacillus thuringiensis* insecticidal crystal protein gene in tobacco plants. *In* CJ Arntzen, C Ryan, eds, Molecular Strategies for Crop Protection, Vol 46. Alan R. Liss, Inc. New York City, pp 345-353

An G (1985) High-efficiency transformation of cultured tobacco cells. Plant Physiol. 79: 568-570

Aronson AI (1993) The two faces of *Bacillus thuringiensis*: insecticidal proteins and post-exponential survival. Mol. Microbiol. 7: 489-496

Barton KA, Whiteley HR, Yang N-S (1987) *Bacillus thuringiensis* -endotoxin expressed in transgenic *Nicotiana tabacum* provides resistance to lepidopteran insects. Plant Physiol. 85: 1103-1109

Bigelow CC, Channon M (1982) In G.D. Fasman, ed. CRC Handbook of Biochemistry and Molecular Biology, Ed 3, Proteins Vol 1. CRC Press, Boca Raton, FL, pp 209-243

Burges HD (1982) Control of insects by bacteria. Parasitology 84: 79-117

Cheng J, Bolyard MG, Saxena RC, Sticklen MB (1992) Production of insect resistant potato by genetic transformation with a -endotoxin gene from *Bacillus thuringiensis* var. *kurstaki*. Plant Sci. 81: 83-91

Dandekar AM, McGranahan GH, Vail PV, Uratsu SL, Leslie C, Tebbets JS (1994) Low levels of expression of wild type *Bacillus thuringiensis* var. *kurstaki cryIA(c)* sequences in transgenic walnut somatic embryos. Plant Sci. 96: 151-162

Delannay X, LaVallee BJ, Proksch RK, Fuchs RL, Sims SR, Greenplate JT, Marrone PG, Dodson RB, Augustine JJ, Layton JG, Fischhoff DA (1989) Field performance of transgenic tomato plants expressing the *Bacillus thuringiensis* var.*kurstaki* insect control protein. Bio/Technology 7: 1265-1269

De Rocher EJ, Vargo-Gogola TC, Diehn SH, Green PJ (1997) Direct evidence for rapid degradation of *B.t.* toxin mRNA as a cause of poor expression in plants. (manuscript submitted)

Diehn SH, De Rocher EJ, Green PJ (1996) Problems that can limit the expression of foreign genes in plants: lessons to be learned from *B.t.* toxin genes. *In* JK Setlow, ed, Genetic Engineering: Principles and Methods, Vol 18. Plenum Press, New York, pp 83-99

Düring K (1988) Wundinduzierbare Expression und Sekretion von T4 Lysozym und monoklonalen Antikörpern in *Nicotiana tabacum*. PhD thesis. University of Cologne, Cologne

Düring K, Porsch P, Fladung M, Lörz H (1993) Transgenic potato plants resistant to the phytopathogenic bacterium *Erwinia carotovora*. Plant J. 3: 587-598

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity (addendum). Anal. Biochem. 137: 266

Feitelson JS, Payne J, Kim L (1992) *Bacillus thuringiensis*: insects and beyond. Bio/Technology 10: 271-275

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 (1987) Insect tolerant transgenic tomato plants. Bio/Technology 5: 807-813

Haseloff J, Amos B (1995) GFP in plants. Trends Genet. 11: 328-329

Haseloff J, Siemering KR, Prasher DC, Hodge S (1997) Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. Proc. Natl. Acad. Sci. USA 94: 2122-2127

Höfte H, Whiteley HR (1989) Insecticidal crystal proteins of *Bacillus thuringiensis*. Microbiological Reviews 53: 242-255

Hunt AG (1994) Messenger RNA 3' end formation in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 45: 47-60

Hunt AG, MacDonald MH (1989) Deletion analysis of the polyadenylation signal of a pea ribulose-1,5-bisphosphate carboxylase small-subunit gene. Plant Mol. Biol. 13: 125-138

Jarvis P, Belzile F, Dean C (1997) Inefficient and incorrect processing of the *Ac* transposase transcript in *iae1* and wild-type *Arabidopsis thaliana*. Plant J. 11: 921-931

Jefferson RA, Burgess SM, Hirsh D (1986) b-Glucuronidase from *Escherichia coli* as a gene-fusion marker. Proc. Natl. Acad. Sci. USA 83: 8447-8451

Joshi CP (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res. 15: 6643-6653

Li Q, Hunt AG (1995) A near-upstream element in a plant polyadenylation signal consists of more than six nucleotides. Plant Mol. Biol. 28: 927-934

Lütcke HA, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA (1987) Selection of AUG initiation codons differs in plants and animals. EMBO J. 6: 43-48

Martin DJ, Firek S, Moreau E, Draper J (1997) Alternative processing of the maize *Ac* transcript in *Arabidopsis*. Plant J. 11: 933-943

Mogen BD, MacDonald MH, Graybosch R, Hunt AG (1990) Upstream sequences other than AAUAAA are required for efficient messenger RNA 3'-end formation in plants. Plant Cell 2: 1261-1272

Mogen BD, MacDonald MH, Leggewie G, Hunt AG (1992) Several distinct types of sequence elements are required for efficient mRNA 3' end formation in a pea rbcS gene. Mol. Cell. Biol. 12: 5406-5414

Murray EE, Rocheleau T, Eberle M, Stock C, Sekar V, Adang M (1991) 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
Nagata T, Nemoto Y, Hasezawa S (1992) Tobacco BY-2 cell line as the "HeLa" cell in the cell biology of higher plants. Int. Rev. Cytol. 132: 1-30

Newman TC, Ohme-Takagi M, Taylor CB, Green PJ (1993) DST sequences, highly conserved among plant *SAUR* genes, target reporter transcripts for rapid decay in tobacco. Plant Cell 5: 701-714

Oakes JV, Shewmaker CK, Stalker DM (1991) Production of cyclodextrins, a novel carbohydrate, in the tubers of transgenic potato plants. Bio/Technology 9: 982-986

Perlak FJ, Fuchs RL, Dean DA, McPherson SL, Fischhoff DA (1991) Modification of the coding sequence enhances plant expression of insect control genes. Proc. Natl. Acad. Sci. USA 88: 3324-3328

Reichel C, Mathur J, Eckes P, Langenkemper K, Koncz C, Schell J, Reiss B, Maas C (1996) Enhanced green fluorescence by the expression of an *Aequorea victoria* green fluorescent protein mutant in mono- and dicotyledonous plant cells. Proc. Natl. Acad. Sci. USA 93: 5888-5893

Rothnie HM, Reid J, Hohn T (1994) The contribution of AAUAAA and the upstream element UUUGUA to the efficiency of mRNA 3'-end formation in plants. EMBO J. 13: 2200-2210

Rothnie HM (1996) Plant mRNA 3'-end formation. Plant Mol. Biol. 32: 43-61

Rouwendal GJA, Mendes O, Wolbert EJH, de Boer AD (1997) Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage. Plant Mol. Biol. 33: 989-999

Sanfacon H, Brodmann P, Hohn T (1991) A dissection of the cauliflower mosaic virus polyadenylation signal. Genes & Dev. 5: 141-149

Schnepf HE Whiteley HR (1985) Delineation of a toxin-encoding segment of a *Bacillus thuringiensis* crystal protein gene. J. Biol. Chem. 260: 6273-6280

Thompson DM (1990) Transcriptional and post-transcriptional regulation of the genes encoding the small subunit of ribulose-1, 5- bisphosphate carboxylase. Ph.D. thesis. University of Georgia, Athens

Vaeck M, Reynaerts A, Höfte H, Jansens S, De Beuckeleer M, Dean C, Zabeau M, Van Montagu M, Leemans J (1987) Transgenic plants protected from insect attack. Nature 328: 33-37

Van Aarssen R, Soetaert P, Stam M, Dockx J, Gosselé V, Seurinck J, Reynaerts A, Cornelissen M (1995) *cry* IA(b) transcript formation in tobacco is inefficient. Plant Mol. Biol. 28: 513-524

Wu L, Takashi U, Messing J (1995) The formation of mRNA 3'-ends in plants. Plant J. 8: 323-329

Chapter 3

DIFFERENTIAL UTILIZATION OF POLYADENYLATION SIGNALS IN
GRAMINEOUS AND DICOTYLEDONOUS PLANTS

ABSTRACT

The general structure of poly(A) signals appears to be conserved in plants, but variations in the sequence and spacing of the elements are common. This variability raises the possibility that the recognition of some polyadenylation signals may have diverged between species or groups of plants. It was observed while studying the expression of a *cryIA(c) B.t.* toxin gene that one of the poly(A) addition sites in the coding region of this gene was less efficiently utilized in tobacco protoplasts than in maize protoplasts. This, together with observations from the literature, provided evidence that some polyadenylation signals may be differentially recognized in plants and that maize and tobacco may have different requirements for poly(A) signal utilization. The generality of differential poly(A) signal utilization was investigated by expressing chimeric genes containing the polyadenylation signals from monocot (Gramineae) and dicot plant genes in maize, wheat, tobacco, and Arabidopsis. Maize and wheat cells efficiently utilized each signal regardless of its source. In contrast, tobacco cells did not efficiently utilize some of the monocot poly(A) signals. Similar results were obtained with Arabidopsis seedlings; however, the efficiency of utilization was slightly better than observed in tobacco. These results show that maize and wheat cells are able to utilize a greater array of plant poly(A) signals than tobacco cells and Arabidopsis seedlings. This is consistent with a monocot(Gramineae)/dicot difference in the efficiency of recognition of some plant poly(A) signals. This work also demonstrates that further study of the fundamental principles governing polyadenylation in plants will have both basic and applied significance.

INTRODUCTION

The post-transcriptional processing of pre-mRNAs is an essential requirement for gene expression in eukaryotic cells. The primary transcripts are capped at the 5' end, intervening sequences are removed from the coding region, and poly(A) tails are added to the 3' end of the transcript after endonucleolytic cleavage. Poly(A) tails play a vital role in the function and metabolism of mRNAs. They facilitate the translation of mRNAs (reviewed in Jackson and Standart, 1990) and are important in mRNA degradation (reviewed in Jacobson and Peltz, 1996). The process of 3' end formation is also required for transcription termination (reviewed in Proudfoot, 1989) and has been shown to be necessary for the export of histone mRNAs from the nucleus (Eckner et al. 1991). Since poly(A) tails are found at the end of nearly every eukaryotic transcript, the functional roles of poly(A) tails are likely to be conserved among eukaryotes. However, the structure of the signals for 3' end formation can be different, in particular between mammals and plants.

The process of 3' end formation in mammals consists of endonucleolytic cleavage of the primary transcript, tightly coupled to the addition of adenylate residues to the 3' end of the upstream cleavage product (Moore and Sharp 1985, for review Wickens, 1990, Wahle, 1992, Wahle and Keller, 1996). The cleavage event is directed by two sequence elements in the transcript, the canonical AAUAAA hexanucleotide sequence, located 10-30 nucleotides upstream of the cleavage site, and a GU-rich (or U-rich) element located downstream of the cleavage/polyadenylation site (Wahle, 1992, Wahle and Keller, 1992, Colgan and Manley 1997). In addition to cleavage of the primary transcript, the

AAUAAA sequence is essential for polyadenylation. Ninety percent of all the known polyadenylation signals contain the AAUAAA sequence. Most of the remaining 10% of transcripts contain a variant sequence with AUUAAA as the most common variant (Wickens, 1990, Colgan and Manley, 1997). This is because mutations are not well tolerated in the AAUAAA sequence (Sheets et al. 1990). Saturation mutagenesis of the sequence resulted in cleavage and polyadenylation efficiencies at an average of less than 30% of the wild-type sequence. The only exception was the AUUAAA sequence which had a frequency of cleavage and polyadenylation of approximately 66% and 77%, respectively, compared to the AAUAAA sequence (Sheets et al. 1990).

The GU-rich (or U-rich element) is located within 50 bases downstream of the cleavage site and is required for cleavage of the primary transcript. There is no strict consensus sequence for this element, but it typically contains single or multiple stretches of up to five consecutive U resides, often interrupted by single G residues (MacDonald et al. 1994; Takagaki and Manley 1997; reviewed in Colgan and Manley, 1997). Considerably more variation in sequence composition is tolerated in this element. Point mutations introduced into this element do not have the same severe effect on 3' end formation as mutation of the AAUAAA sequence (McDevitt et al., 1986). However, specific G to U conversions in the downstream element of the SV40 early polyadenylation signal can improve the efficiency of 3' end formation three-fold or decrease the efficiency by four-fold which argues that the downstream element of this polyadenylation signal is sequence specific and not just characterized by its nucleotide content (McDevitt et al., 1986).

The polyadenylation signals in plants are much more diffuse, redundant and complex than their mammalian counterparts. Comparative studies of the regions surrounding plant poly(A) sites show only about 1/3 of plant genes contain the AAUAAA sequence. Another 50% have 1 to 2 base substitutions in the hexanucleotide sequence. Fifteen percent have no AAUAAA-like motif (Hunt, 1994). In addition to differences in sequence requirements, mammalian polyadenylation signals do not function properly in plants (Hunt et al. 1987). Thus, plant poly(A) signals appear to differ from mammalian poly(A) signals.

Plant polyadenylation signals consist of two cis-regulatory elements: the Far Upstream Element (FUE) and the Near Upstream Element (NUE). Unlike mammalian polyadenylation signals, both of these elements are upstream of the site of polyadenylation. There is no consensus sequence for either element, but each has key characteristics based on nucleotide composition.

The FUE, located approximately 40 to 150 bases upstream of the polyadenylation site, functions to enhance overall polyadenylation efficiency (for review Hunt, 1994; Rothnie, 1996; Li and Hunt, 1997). The FUE may span a region of 100nts (Li and Hunt 1997), but the relative position is variable and difficult to determine because the only distinguishing feature of the FUE is the presence of multiple UG-rich motifs scattered throughout the element. The FUE is known to be a functionally redundant element that can tolerate many mutations. Twenty base linker scanning substitutions and small deletions through the pea rubisco E9 FUE have subtle effects on the function of the element (Mogen et al., 1992). Only large deletions have a dramatic effect on normal 3' formation (Hunt and MacDonald, 1989; Mogen et al., 1990; reviewed in Li and Hunt,

1997). Similar results were obtained with the Cauliflower Mosaic Virus (CaMV) and

Figwort Mosaic Virus polyadenylation signals (Sanfacon et al., 1991; Sanfacon et al.,

1994). Polyadenylation was reduced to 8% of wild-type when most of the CaMV FUE is

deleted (Sanfacon et al., 1991).

The NUE is an AU-rich region located approximately 10 to 40 bases upstream of the

cleavage site (reviewed in Hunt, 1994; Wu et al., 1995; Rothnie, 1996, Li and Hunt,

1997). The position of the NUE and the fact that they contain AAUAAA or AAUAAA-

like sequences suggest they may be functionally similar to the AAUAAA sequence of

mammalian poly(A) signals (Wu et al., 1995). Unlike the mammalian sequence,

however, a large number AAUAAA variants are able to participate in polyadenylation.

Saturation mutagenesis of the AAUAAA sequence of the CaMV NUE showed that all

single base mutations were recognized with processing efficiencies upwards of 60% of

the wild-type sequence. Only the replacement of AAUAAA with a GC-rich sequence

reduced the efficiency to a level of a deletion mutant (Rothnie et al., 1994). The NUE is

essential for polyadenylation and controls poly(A) addition at specific sites. Thus, a plant

transcript with multiple polyadenylation sites will have an NUE corresponding to each

site. The existence of multiple poly(A) addition sites within a plant transcript is quite

common and represents yet another difference with mammalian transcripts which are

polyadenylated only at a single site. Poly(A) addition usually occurs at

pyrimidine/adenosine dinucleotides in plants and at CA dinucleotides in animal

transcripts.

Despite our understanding of the structure of plant polyadenylation signals and how

the structure is different from mammalian polyadenylation signals, it remains unclear

whether plant polyadenylation signals function similarly in different plant species. For

instance, are the polyadenylation signals from monocot plant species efficiently

recognized in dicot plant species and visa versa? The general structure of plant

polyadenylation signals appears to be conserved in monocot and dicot species which

might suggest the basic polyadenylation machinery is conserved in all plants and that

there is no difference in poly(A) signal recognition between monocots and dicots. The

proper utilization of the maize 27kD zein and the wheat histone H3 polyadenylation

signals in dicot cells support this hypothesis (Tabata et al., 1987; Wu et al., 1994).

However, variations in the sequence and spacing of the FUE and NUE as well as the

diffuse and redundant nature of these elements suggests that plant poly(A) signals may be

more efficiently recognized by some plant species than by others. It has been reported

that the polyadenylation signal of the wheat rbcS gene is not efficiently utilized in

tobacco plants (Keith and Chua, 1986). Only 50% of the total wheat transcripts were

polyadenylated at the site normally used in wheat plants. The remaining transcripts were

polyadenylated at novel sites further downstream. These conflicting observations

demonstrate the need for more analysis of poly(A) signal recognition in plants.

In this report, we investigated the generality of differential poly(A) signal utilization

in plants. We initially became interested in this topic while studying the poor expression

of a *cryIA(c) B.t.* toxin gene in tobacco (Diehn et al., 1998; De Rocher et al., 1998). The

transcript of the *cryIA(c)* gene, which encodes an insecticidal protein specific for the

order Lepidoptera, accumulates to barely detectable levels in transformed tobacco cells.

We found that the low accumulation was due to the inherent instability of the *cryIA(c)*

transcript and the presence of poly(A) addition sites in the coding region which results in

the production of short transcripts. A set of chimeric genes were constructed to analyze the poly(A) addition sites further. Surprisingly, it was observed that tobacco cells utilized one of the poly(A) addition sites less efficiently than maize cells. These results, which are presented in this report, and other published observations prompted us to address whether the polyadenylation signals from plant genes are also differentially utilized in maize and tobacco. We also wanted to determine whether the differential utilization of the polyadenylation signals is specific to maize and tobacco or a more general phenomenon divided between monocots and dicots.

To this end, a second set of chimeric genes were constructed which contained the polyadenylation signals from several plant genes. Expression of the genes in maize, wheat, tobacco, and Arabidopsis showed that the polyadenylation signals from dicot genes were efficiently utilized in each system. However, the monocot (Gramineae) polyadenylation signals were not always efficiently utilized in tobacco and Arabidopsis. These results demonstrated that the polyadenylation signals from some plant genes are differentially utilized in plants and that this difference is divided between monocots (Gramineae) and dicots. In addition, these results suggest that monocots (Gramineae) may have a less stringent requirement for plant poly(A) signal utilization than dicots. This may have important implications in biotechnology where improper poly(A) signal selection could have a negative impact on the expression of foreign genes in plants. These results also may advance our understanding of the differences in gene expression mechanisms between monocot and dicot plant species.

METHODS AND MATERIALS

PCR and Plasmid Construction

The plasmid, p1185, served as the basis for all the chimeric genes used in this report.

This plasmid consists of a modified 35S promoter followed by the human β-globin

coding region, and the 3'UTR of the pea rubisco small subunit E9 gene in pUC 8 (details

of the construction are described in Diehn et al., 1998).

Plasmids p1187 and p1188 have chimeric genes that contain segment 2 and segment

4, respectively, from the *cryIA(c)* coding region. The *cryIA(c)* coding region from

Bacillus thuringiensis subsp. *kurstaki* HD-73 was kindly provided by Dr. A.I. Aronson at

Purdue University. The sequences were modified at the 5' and 3' ends as described in

Diehn et al. (1998) to form plasmid p995. Unique restriction sites were selected to divide

the coding region into the segments which contain the polyadenylation signals. Segment

2 is a 375bp *Xba*I-*Xba*I restriction fragment (nucleotides 679-1054). To obtain the

appropriate ends for ligation into p1185, segment 2 was inserted into the *Eco*RV site of

p948 (described in Diehn et al., 1998) to form p1147. Digestion of this plasmid with

*Bam*HI and *Bgl*II released segment 2 which was then inserted into the *Bam*H1 site

upstream of an E9 3'UTR in p977 (described in De Rocher et al., 1998) to make p1151.

The *Bam*H1-*Cla*I fragment from p1151 which contained segment 2 and the E9 3'UTR

was introduced into p1138 (described in Diehn et al., 1998) to form p1156. Plasmid

p1187 was eventually generated by replacing the *Bam*H1-*Sca*I fragment of p1185 which

has the E9 3'UTR and part of the ampicillin resistance gene with the *Bam*H1-*Sca*I

fragment of p1156 containing segment 2, E9 and the missing portion of the ampicillin

resistance gene. Construction of plasmid p1188 which has the chimeric gene with

segment 4 of the *cryIA(c)* coding region (nucleotides 1585-2225, plus 10bp from the

modifications) is described in Diehn et al. (1998). Segment 4 is defined by the restriction

sites *Acc*I and *Bam*H1 where the *Bam*H1 site is part of the multiple cloning site of p995.

The polyadenylation signals from the plant genes were obtained by PCR of the

appropriate genomic clone using sense oligonucleotides which hybridize to the 5' end of

the 3'UTR and antisense oligonucleotides which hybridize 160 bp downstream of the

cleavage site. The sequence of the oligonucleotides for each 3'UTR is as follows:

CHI2, sense 5'GCGATATCGGATCCAGGAAGCTCCTAAGTTTA3',

    antisense 5'GCGATATCAGATCTTTTAATTAGAAAAAAATTTC3';

LTP, sense 5'GCGATATCGGATCCTCAACCACTTGCTGCCTATAG3',

    antisense 5'GCGATATCAGATCTTGTTGTCAAACAGTTTTAGGG3';

MT-L, sense 5'GCGATATCGGATCCTCACATCGATCGACGACC3',

    antisense 5'GCGATATCAGATCTTTGAACAAGTGGATTTTAG3';

PEPC, sense 5'GCGATATCGGATCCGCGGCTTCTCTTCACTCAC3',

    antisense 5'GCGATATCAGATCTCTAGGAGCGAGTAACAAG3';

PR-1a, sense 5'GCGATATCGGATCCTTGAAACGACCTACGTCC3',

    antisense 5'GCGATATCAGATCTTTTTAAGACTAACGGTCAG3';

TrpA, sense 5'GCGATATCGGATCCGTCCATGACAAAGTAAAACG3',

    antisense 5'GCGATATCAGATCTCTTGGTATCTTATACCTC3'.

The 5' end of each sense oligonucleotides contains *Eco*RV and *Bam*H1 restriction sites and the antisense oligonucleotides have *Eco*RV and *Bgl*II restriction sites to facilitate the cloning of the PCR products.

The amplification of the regions containing polyadenylation signals was done in 30 cycles of 1 min. at 95°C, 2min. at 55°C and 1 min. at 72°C, after an initial 5 min. denaturization step at 95°C. The final step was a 10 min. extension at 72°C. Afterwards, the reactions were chloroform treated to remove the mineral oil and electrophoresed through a 1% low melting agarose gel (FMC BioProducts, Rockland, ME) in 1X TAE. The PCR products were extracted from gel pieces using a hot phenol technique which consisted of incubating the gel pieces in TE/0.3M NaCl and 50ng μl^{-1} glycogen at 65°C for 5 min. After vortexing, an equal volume of phenol equilibrated to 65°C was added and the mixture was vortexed for 1 min. After phenol treatment, the PCR products were chloroform extracted and precipitated. The resuspended pellets were digested with *Bam*H1 and *Bgl*II, gel purified as above and ligated into the unique *Bam*H1 site of p1185 between the globin coding region and the E9 3'UTR. Clones with proper orientation as determined by restriction digest analysis were sequenced in both directions across the amplified region by automated cycle sequencing (Novartis Corporation, Research Triangle, NC) to ensure no mutations were introduced during PCR. Plasmids selected to be expressed in plant cells were numbered accordingly: p1719 (CHI2); p1720 (MT-L); p1721 (PEPC); p1722 (PR-1a); p1723 (TrpA); p1741/p1742 (LTP 4.2). For integration into the Arabidopsis genome, the chimeric gene cassettes were ligated into the *Hind*III site of pBI121: p1743 (CHI2); p1744 (MT-L); p1775/p1776 (PEPC); p1789/p1790 (LTP 4.2). The gene cassette with the TrpA poly(A) signal was integrated into pBI121

differently because of the presence of a *Hind*III site in the amplified DNA fragment. The

TrpA gene cassette was removed from p1741 by *Not*I digestion. pBI121 digested with

*Hind*III and the gene cassette were blunt ended with Klenow enzyme. pBI121 was further

treated with shrimp alkaline phosphatase. The TrpA gene cassette and pBI121 were both

electrophoresed through a 1% agarose gel in 1xTBE and then extracted using DEAE

paper (Sambrook et al., 1989). After ligation into pBI121, correct orientation of the gene

cassette was determined by restriction digest analysis and named (p1809 (TrpA)).

RT-PCR of the segment 2 globin-*B.t.* transcript from maize was performed as

described in Diehn et al, (1998).

Plant Material and Transformation

Nicotiana tabacum cv Bright Yellow 2 (BY-2; also called NT-1) cells (Nagata et al.,

1992) were cultured as previously described by Newman et al. (1993). Plasmids p1185,

p1187 and p1188 were introduced into BY-2 protoplasts by electroporation as described

by van Hoof and Green (1996). Plasmids p1743 (CHI2), p1744 (MT-L), p1775 (PEPC),

and p1789 (LTP) were introduced into BY-2 protoplasts by PEG transformation. Control

experiments using p1185, p1187 and p1188 showed there was no difference in the

expression pattern whether electroporation or PEG transformation was used. Protoplasts

for PEG transformation were prepared by incubating the cells from 3-4 day old liquid

cultures in 2% cellulase RS and 1% macerozyme R10 (Kanematzu-Goshu, Los Angles,

CA), in KMC 700, pH 5.7 (8.65g KCl, 16.47g $MgCl_2.6H_2O$, 12.5g $CaCl_2.2H_2O$, 5g MES

in 900ml ddH_2O, pH 6 with KOH; osmolarity 700mOsm) or 2% cellulysin (Calbiochem,

La Jolla, CA), 1% cytolase (Genencor International, Rolling Meadows, IL), and 0.2%

pectolyase, (Karlan, Santa Rosa, CA) in KMC 700, pH 5.7 for 3-5 hrs. at 28°C. The

protoplasts were sieved through a 100μm mesh and washed in KMC 650 (KMC 700

except with an osmolarity adjusted to 650mOsm) followed by a wash in KMC 600 (KMC

700 except with an osmolarity adjusted to 600 mOsm). After counting, the protoplasts

were resuspended in resuspension buffer (500mM mannitol, 15mM $CaCl_2$. $2H_2O$, 0.1%

(w/v) MES (2[N-morpholino] ethanesulfonic acid), pH 5.6 with KOH) to a concentration

of 8×10^6 protoplasts ml^{-1}. One half milliliter of protoplasts was incubated with 100μg of

affinity purified (Qiagen, Chatsworth, CA) or cesium chloride-ethidium bromide purified

(Sambrook et al., 1989) plasmid DNA (in a vol. of 50 μl or 100μl) in the presence of

0.5ml 40% PEG (40% PEG 8000 (w/v), 0.4M mannitol, 0.1M $Ca(NO_3)_2.4H_2O$, pH 9.0).

After 30 min. at room temperature, W6 solution (0.38g KCl, 9g $CaCl_2.2H_2O$, 9g NaCl, 9g

glucose, 1g MES (2[N-morpholino]ethanesulfonic acid) in 600ml ddH_2O, pH 6.0,

osmolarity 550-600mOsm) was added stepwise in 1ml, 2ml, and 5ml volumes at 5 min.

intervals. After gentle centrifugation, protoplasts were resuspended in FW media (4.3g l^{-1}

MS salts (Gibco BRL, Gaithersburg, MD), 5ml l^{-1} 200X B5 vitamins, 30g l^{-1} sucrose,

1.5g l^{-1} proline, 54g l^{-1} mannitol, 3mg l^{-1} 2,4-D, pH 5.7 with KOH) and incubated at 28°C

in petri plates in the dark for approximately 12 hrs. before they were harvested for RNA.

Each plasmid was transformed in replicate and then pooled at the time of harvest to be

counted as a single experiment. Experiments were repeated atleast once for each plasmid.

Transformation of maize (Zea mays) Black Mexican Sweet (BMS) cells was by the

methods described above. p1187 and p1188 were introduced using electroporation and

the chimeric genes containing the poly(A) signals from the plant genes were introduced

using PEG. Control experiments showed there was difference in expression pattern using the two different methods.

Wheat cell suspension cultures (cv Mustang, Chang et al., 1991) were maintained by weekly transfer of 5ml of culture (approximately 1ml packed cell volume) to 60ml of 2MS media (4.3g l^{-1} MS basal salts (Gibco BRL, Gaithersburg, MD), 10ml l^{-1} 100x MS vitamins (Gibco BRL, Gaithersburg, MD), 2ml l^{-1} 2,4-D at 1mgml^{-1}, 100mg l^{-1} myo-inositol, 300mg l^{-1} glutamine, 150mg l^{-1}, asparagine, 500mg l^{-1} proline, 30g l^{-1} sucrose, pH 5.8) in a 250ml baffle flask (Bellco Biotech, Vineland, NJ). Cells were grown in dark at 28°C on a New Brunswick G10 gyratory shaker at 120rpm.

The day before bombardment, 10ml of wheat cells were transferred to a 250ml baffle flask containing 50ml fresh 2MS media. On the day of bombing, the cells were divided into three parts and distributed evenly on Millipore AP10 filters (approximately 2.3cm in diameter). Cells were bombarded with the gene constructs using a DuPont PDS-1000 helium-driven particle delivery system. Fifty microliters of 1μm gold particle suspension (60 mg ml^{-1} in sterile glycerol) was mixed with 5μl of DNA at 1μg μl^{-1}. The DNA was precipitated onto the particles by adding 50μl CaCl$_2$ (2.5M) and 20μl spermidine (0.1M), mixing well on a vortex for 3 min. The particles were washed with 250μl of 100% ethanol and resuspended in 70μl of 100% ethanol. Ten microliters of the resuspended particles were pipetted onto the microcarrier sheet. Plated suspension cells were bombarded under partial vacuum (28mmHg) at a distance of 6cm using 1100psi rupture discs. Each target was bombed twice and immediately resuspended in a 60ml flask with 10ml fresh 2MS media. Cells were shaken until harvested for RNA at approximately 12 hrs after bombardardment.

Stably transformed BY-2 cells expressing the globin-*B.t.* chimeric genes were obtained as described in Diehn et al. (1998). Transgenic tobacco plants were generated as described in Newman et al. (1993).

Transformation of Arabidopsis plants was done according to van Hoof and Green (1996). The only exceptions were the final concentrations of the antibiotics in the YEP medium (50μg ml^{-1} rifampicin, 25 μg ml^{-1} gentamycin, and 100 μg ml^{-1} kanamyin) and the concentration of kanamycin in the seed selection plates (50 μg ml^{-1} kanamycin).

RNA Methods

RNA was extracted from wheat cells, cultured BY-2 cells, and transgenic plant tissue by the method of Puissant and Houdebine (1990) with the modifications described in Newman et al. (1993). RNA was isolated from BY-2 and BMS protoplasts using the same methods except that the protoplasts were not ground under liquid N_2, but thawed directly in 2.5ml (one half the vol. described in Newman et al., 1993) of guanidinium thiocyanate solution while vortexing. The volumes of the subsequent solutions were also reduced by half, up to the first lithium chloride step. A phenol/chloroform extraction was performed after solubilization of the second lithium chloride pellet for all RNA preparations. Aliquots of total RNA were treated on occasion with DNaseI (RQ1; Promega, Madison, WI) for 15 min at 37°C in DNase assay buffer (40mM Tris-HCl pH 7.9, 10mM NaCl, and 6mM $MgCl_2$, supplement with 2mM DTT and 40 units RNasin (Promega, Madison, WI) final conc.) to remove residual DNA. Samples were phenol/cholorform extracted and RNA pellets were resuspended in sterile RNase-free

water. 20μg of total RNA was denatured and separated on 2% (v/v) formaldehyde/1%

(w/v) agarose gels in 1X MOPS buffer (20mM 3-[N-morphilino]propanesulfonic acid,

5mM sodium acetate, 1mM EDTA, 1mg ml⁻¹ ethidium bromide) before capillary transfer

to BioTrace HP (Gelman Sciences, Ann Arbor, MI) for the globin-*B.t.* chimeric

transcripts or Nytran Plus (Schleicher and Schuell, Keene, NH) for the chimeric

transcripts containing the poly(A) signals from the plant genes. Prehybridization and

hybridization of the blots with the globin-*B.t.* chimeric transcripts was as described by De

Rocher et al. (1998) except prehybridization was overnight. Blots with the chimeric

transcripts containing the plant poly(A) signals were prehybridized overnight at 52°C in

25ml of gel elution and hybridization buffer described in Church and Gilbert, (1984).

The prehybridization buffer was replaced with fresh buffer containing radiolabeled probes

to the globin coding region at 1×10^6 cpm ml⁻¹. The probes were labeled with [α-

^{32}P]dCTP by the random primed method described in Feinberg and Vogelstein (1983)

using DNA restriction fragments separated on low melting agarose gels. Unincorporated

nucleotides were removed by using push columns (Stratagene, La Jolla, CA) or by using

NICK columns (Pharmacia Biotech, Piscataway, NJ). Blots were washed for 20-30

minutes at 65°C in 2X SSC, 0.1% SDS (w/v) followed by a wash in 1x SSC, 0.1% SDS

(w/v) under the same conditions.

[α-^{32}P]UTP radiolabeled RNA probes corresponding to the entire E9 3'UTR were *in*

vitro transcribed using the Riboprobe System (Promega, Madison, WI) from linearized

Bluescript II SK(-) plasmid, p1425, which contains the E9 3'UTR. The globin probes

were removed from blots by two successive washes in near boiling DEPC-treated water

which was allowed to cool to room temperature. Then, blots were prehybridized

overnight in prehybridization buffer described in De Rocher et al. (1998) at 52°C. The

blots were hybridized with 1×10^6 cpm ml^{-1} E9 riboprobes overnight at 65°C in

hybridization solution, also described in De Rocher et al. (1998), which was modified by

increasing the formaldehyde concentration to 50% (v/v) and decreasing the SSC

concentration to 1X. Blots were washed as described above with an additional wash at

0.2x SSC, 0.1% (w/v) SDS.

Transcripts were visulalized by autoradiography and quantified using a

Phosphorimager (Molecular Dynamics, Sunnyvale, CA). Poly(A) site selection was

determined by comparing the actual size of each chimeric transcript to the predicted size

based on utilization of either the introduced or E9 sites. The length of the poly(A) tail

was determined by subtracting the actual size of the globin/E9 control transcript from the

predicted size (approximately 785nt). The chimeric transcripts were assumed to be

polyadenylated to a similar extent as the control transcript and any gel artifacts were also

assumed to be distributed throughout the gel.

Oligo-directed RNaseH Cleavage Analysis

0.5μg oligo d(T)$_{12\text{-}18}$ was annealed to 10 μg total RNA in a 400ml 65°C water bath that

was allowed to cool to room temperature. Afterwards, the hybrids were incubated in

4mM Tris-HCL pH 8, 10mM MgCl$_2$, 20mM KCl, 1mM DTT, and 1 unit of RNaseH

(Gibco BRL, Gaithersburg, MD) for one hour at 37°C. The samples were alcohol

precipitated and resuspended in loading buffer prior to their electrophoresis through a 2%

(v/v) formaldehyde/1.7% (w/v) agarose gel in 1xMOPS buffer (described above). The

gels were blotted to BioTrace HP membrane and probed as described above.

RESULTS

Differential Utilization of the Polyadenylation Signals from the *crylA(c) B.t.* toxin Coding

Region.

Poly(A) addition sites were recently identified in the coding region of a *crylA(c) B.t.*

toxin gene (Diehn et al., 1998). Two of these poly(A) addition sites were further

analyzed in chimeric genes which consisted of segments of the *crylA(c)* coding region

containing the polyadenylation sites inserted between a globin coding region under the

control of a modified 35S promoter and the well characterized pea rubisco small subunit

E9 3'UTR as shown in Figure 3-1C. The polyadenylation signal in the E9 3'UTR is an

efficiently recognized signal that acts as a "trap" for transcripts that are not efficiently

polyadenylated at the upstream poly(A) site or "test" site. Efficient utilization of the

"test" site will result in transcripts on an RNA gel blot which are short compared to the

size expected for polyadenylation in the E9 3'UTR. In addition, these short transcripts

will lack E9 sequences. If the "test" sites are inefficiently utilized, larger transcripts will

be observed on an RNA gel blot which contain the E9 3'UTR. Chimeric genes

containing segments 2 and 4 of the *crylA(c)* coding region (Figure 3-1C) as well as a

control gene which lacks an insert and consists of only the modified 35S promoter, globin

coding region, and E9 3'UTR were expressed in tobacco and maize protoplasts.

An unexpected result was obtained expressing the segment 2 chimeric gene in

tobacco and maize protoplasts. As shown in Figure 3-2A, lanes 3 and 4, a chimeric

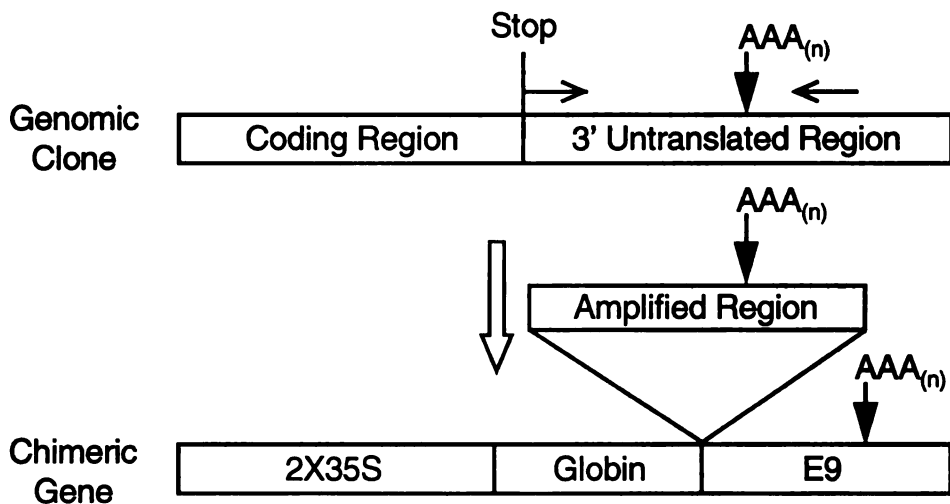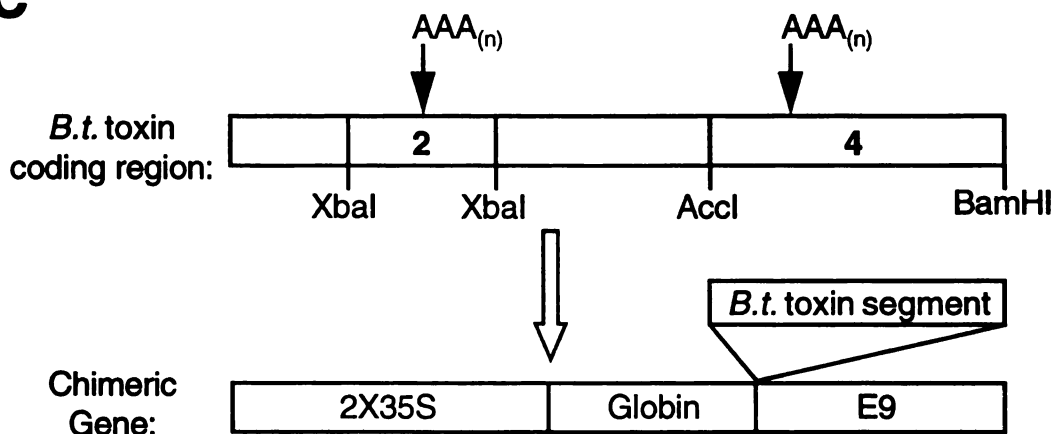**Figure 3-1.** Structure of the Chimeric Genes Expressed in Maize, Wheat, Tobacco, and Arabidopsis

A) Genes used to provide the 3' untranslated regions (3'UTR) for studying poly(A) signal utilization in gramineous and dicotyledonous plants. The genomic clones from stress-induced and housekeeping genes were selected from dicot and monocot plant species for PCR amplification. The polyadenylation signals from the *cryIA(c) B.t.* toxin coding region also were used to study their utilization in each plant system. The accession numbers are: *CHI2*, m84214; *PR-1a*, x12737; *rbcS-E9*, m21375; *MT-L*, s57628; *PEPC*, x15642; *TrpA*, x76713; *LTP*, z66529; *cryIA(c)*, m11068.

B) Structure of the chimeric genes containing the polyadenylation signal from each plant gene. Oligonucleotides (horizontal arrows) hybridizing to the 5' end of the 3'UTR and to a region 160bp downstream of the most 3' poly(A) addition site (closed arrows) were synthesized to amplify the region containing the polyadenylation signal from the genomic clone of each gene listed in Figure 3-1A. The amplified regions were subsequently inserted between a β–globin coding region (globin) and the pea *rbcS-E9* (E9) 3' UTR. The E9 region provides the elements necessary for polyadenylation and serves as a "trap" for the transcripts that are not polyadenylated at the site(s) provided by the introduced polyadenylation signal. Transcription of the chimeric genes was controlled by the Cauliflower Mosaic Virus 35S promoter which was modified by duplicating the upstream enhancer region (2x35S).

C) Structure of the chimeric globin-*B.t.* genes used to study the utilization of the polyadenylation signals from the *cryIA(c)* coding region. Segments of the wild-type *Bacillus thuringiensis* var. *kurstaki cryIA(c)* coding region (amino acids 9-613) containing the polyadenylation signals identified in Diehn et al., 1998 (segments 2 and 4) were excised using the appropriate restriction endonucleases. The restriction fragments were cloned into an intermediary plasmid and then subcloned between the β–globin coding region (globin) under the control of the modified 35S promoter (2x35S) and the E9 3'UTR (E9).

A

Dicot

Pathogenesis-related Protein	(*PR-1a*)	Tobacco
Chitinase (ClassIII)	(*CHI2*)	Cucumber
Rubisco	(*rbcS-E9*)	Pea

Monocot (Gramineae)

Metallothionein-like Gene	(*MT-L*)	Maize
Tryptophan Synthase A	(*TrpA*)	Maize
Phosphoenolpyruvate Carboxylase	(*PEPC*)	Maize
Lipid Transfer Protein 4.2	(*LTP*)	Barley

Non-plant

B.t. toxin coding region	(Segment 2)	*Bacillus thuringiensis*
B.t. toxin coding region	(Segment 4)	*Bacillus thuringiensis*

B

Stop   AAA$_{(n)}$

Genomic Clone

Coding Region	3' Untranslated Region

AAA$_{(n)}$

Amplified Region

AAA$_{(n)}$

Chimeric Gene

2X35S	Globin	E9

C

AAA$_{(n)}$ AAA$_{(n)}$

B.t. toxin coding region:

	2		4

XbaI   XbaI   AccI   BamHI

B.t. toxin segment

Chimeric Gene:

2X35S	Globin	E9

transcript accumulates in tobacco protoplasts that is larger than in maize protoplasts. The size of this abundant transcript is consistent with polyadenylation in the E9 3'UTR (denoted by the arrow). Rehybridization of the blot with the E9 3'UTR shows this transcript does contain E9 sequences, as expected (Figure 3-2B, lane 3). The shorter transcript produced in maize cells does not hybridize to the E9 3'UTR probe (Figure 3-2B, lane 4) and is consistent in size with polyadenylation in segment 2. The presence of a poly(A) tail on this transcript was verified in Figure 3-3, lanes 3 and 4 by oligo-directed RNaseH cleavage analysis. Annealing oligo d(T) to the poly(A) tail in the presence of RNaseH will result in cleavage of the poly(A) tail, causing an increased mobility of the transcript in an RNA gel. The increased mobility of the maize segment 2 transcript is consistent with the removal of the poly(A) tail. RT-PCR analysis confirmed maize utilized the poly(A) addition site mapped previously in segment 2 (data not shown). These results indicate that maize is able to recognize the segment 2 polyadenylation signal more efficiently than tobacco and suggests a maize/tobacco difference in the recognition of polyadenylation signals.

It is unlikely that the context in which the segment 2 polyadenylation signal was inserted prevents utilization of the signal in tobacco. Figures 3-2A and 3-2B, lanes 5 and 6 show tobacco and maize protoplasts expressing the chimeric gene containing segment 4 of the *cryIA(c)* coding region accumulate transcripts that are similar in size. These transcripts do not hybridize with the E9 3'UTR (Figure 3-2B, lanes 5 and 6) and are consistent in size with polyadenylation in the segment 4 DNA fragment. Polyadenylation in the E9 3'UTR would produce transcripts of the size indicated by the arrow in Figure 3-2A. Some hybridization which is consistent with polyadenylation in the E9 3'UTR as

Figure 3-2. Tobacco and Maize Protoplasts Do Not Efficiently Utilize the Same Polyadenylation Signals from the *crylA(c)* Coding Region.

A gel blot of total RNA isolated from tobacco (T) or maize (M) protoplasts transiently expressing either a control gene (No Insert) which consists of the modified 35S promoter, globin coding region and E9 3'UTR or the chimeric genes which contain either segment 2 (2) or segment 4 (4) of the *crylA(c)* coding region. The blot was hybridized with the globin coding region in A, and rehybridized with the E9 3'UTR in B. The closed box in each panel indicates the expected position for the control transcript and the arrowheads indicate the position of the segment and segment 4 transcripts that are polyadenylated in the E9 3'UTR. The lanes are numbered across the bottom of the blot for reference to the text.
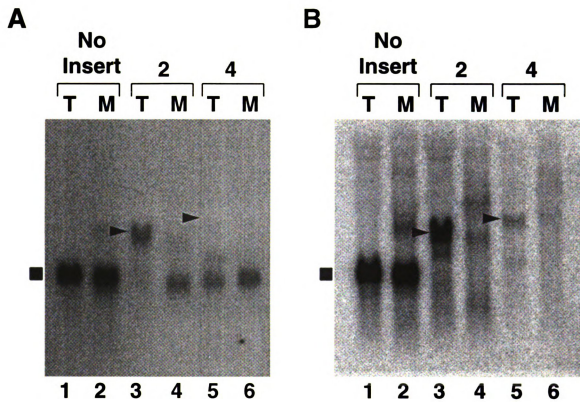
Figure 3-3. The Globin-*B.t.* Chimeric Transcripts Are Polyadenylated in Maize Protoplasts.

Total RNA was isolated from maize protoplasts expressing the control gene (No Insert) or the chimeric genes containing segment 2 (2) or segment 4 (4) of the *crylA(c)* coding region. Each sample was incubated with oligo $d(T)_{15-18}$ in the presence (+) or absence (-) of RNaseH, blotted and probed with the globin coding region. The closed box indicates the position of the polyadenylated globin/E9 control transcript. Each lane is numbered across the bottom of the blot.
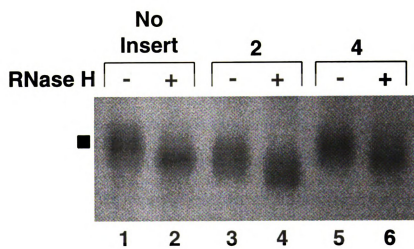
shown in Figure 3-2B, lane 5 and Figure 3-4, lanes 5, 6, and 7 (denoted by the arrow) can be detected in tobacco cells, but this transcript constitutes a small proportion of the total amount of chimeric transcripts produced in these cells. Short transcripts similar to those produced in protoplasts also accumulate in transgenic plants and transformed cultured cells expressing the segment 4 and chimeric gene (compare Figure 3-4, lanes 5, 6, and 7). RNaseH cleavage experiments shown in Figure 3-5, lanes 5 and 6 demonstrate the short transcripts produced in transformed cultured cells as well as the transcript terminating in the E9 3'UTR are polyadenylated. RNaseH cleavage experiments have also shown the maize segment 4 transcripts are polyadenylated (Figure 3-3, lanes 5 and 6). These results show insertion of polyadenylation signals between the globin coding region and the E9 3'UTR does not prevent their recognition in tobacco protoplasts.

The difference in poly(A) signal recognition also is not an artifact of transient expression in protoplasts. The transcripts encoded by the chimeric genes containing the *B.t.* toxin segments were compared in transgenic plants, cultured cells, and protoplasts. Figure 3-4, lanes 4 and 5 show plants and cultured cells expressing the segment 2 chimeric gene accumulate the large transcript as in protoplasts, but also a smaller transcript. The smaller transcript migrates to a position consistent with polyadenylation at the mapped segment 2 poly(A) site. Separate RNA gel blots probed with the E9 3'UTR showed these short globin/*B.t.* transcripts lack E9 sequences, as expected (data not shown). Figure 3-5, lanes 3 and 4 show both the large and short transcripts have an increased mobility in the presence of RNaseH in oligo-directed RNaseH cleavage experiments consistent with the removal of the poly(A) tail. These results indicate the segment 2 poly(A) addition site is recognized by tobacco plants and cultured tobacco

Figure 3-4. The *crylA(c)* Segment 2 Polyadenylation Signal is Inefficiently Utilized in Transgenic Tobacco Plants and Stably Transformed Cultured Cells.

Leaves of transgenic plants (L), stably transformed cells (C), and transiently transformed protoplasts (P) of tobacco expressing the control gene (No Insert) and the indicated globin-*B.t.* chimeric genes. Total RNA was analyzed by gel blot, hybridized with the globin coding region. Lane 7 (P') is an overexposure of lane 6 (P) for the globin-*B.t.* chimeric gene containing segment 2 of the *crylA(c)* coding region. The closed box and arrowheads denote the transcripts polyadenylated in the E9 3'UTR for the control and chimeric transcripts, respectively. The lanes are numbered across the bottom of the blot.
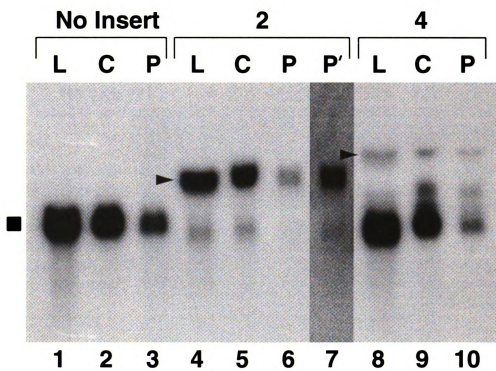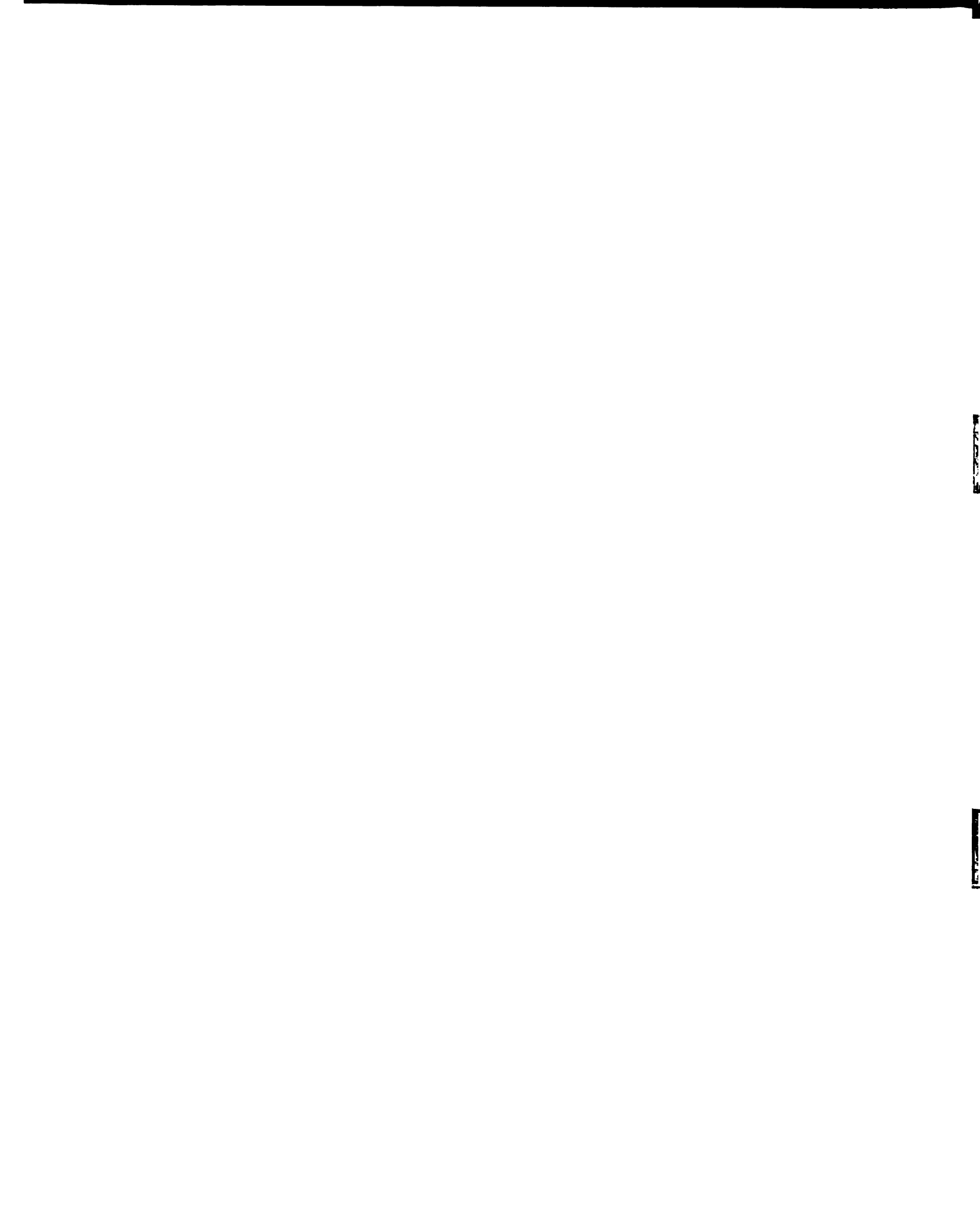
Figure 3-5. The Globin-*B.t.* Chimeric Transcripts Are Polyadenylated in Stably Transformed Tobacco Cells.

A gel blot of total RNA isolated from transformed cultured cells expressing the control gene (No Insert) or the chimeric genes containing segment 2 (2) or segment 4 (4) of the *crylA(c)* coding region. Samples were incubated with oligo $d(T)_{15-18}$ in the presence (+) or absence(-) of RNaseH, blotted and probed with the globin coding region. The closed box denotes the globin/E9 transcript polyadenylated in the E9 3'UTR and the arrowheads indicate globin/*B.t.* transcripts also polyadenylated in the E9 3'UTR. Each lane is numbered across the bottom of the blot for reference to the text. Figure courtesy of Dr. Wan-Ling Chiu.
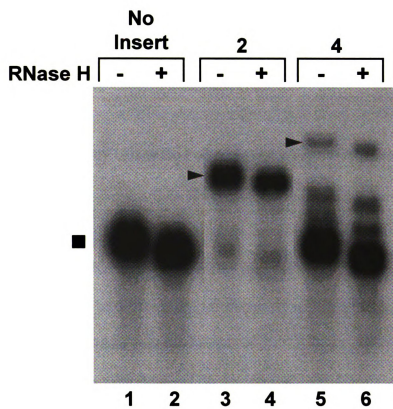
cells. Overexposure of lane 6 in Figure 3-4 shows a short transcript corresponding in size

to the short transcript produced in plants and cultured cells accumulates in protoplasts

(Figure 3-4, lane 7). This indicates that tobacco protoplasts are able to recognize the

segment 2 poly(A) signal similar to tobacco plants and cultured cells and that the

differential utilization of the segment 2 poly(A) signal in tobacco and maize cells is not

an artifact of transient expression in protoplasts. Therefore, the difference in the

utilization of this signal reflects a difference in poly(A) signal recognition between these

two species.

Maize Protoplasts Efficiently Utilize the Polyadenylation Signals from both Gramineae
and Dicot Plant Species

The results using the *B.t.* toxin poly(A) sites suggest that polyadenylation signals

from plant genes may also be differentially utilized in tobacco and maize. To this end,

several monocot and dicot plant genes were selected, as listed in Figure 3-1A, from which

the polyadenylation signals were obtained for introduction into tobacco and maize

protoplasts. Because of the importance of the grasses in agriculture, the polyadenylation

signals representing the monocots are from the family, Gramineae. Thus, the comparison

of poly(A) signal utilization is actually between gramineous and dicotyledonous plants.

The selection of the seven genes was based on several criteria. First, only genes were

chosen in which an *in vivo* poly(A) addition site had been identified. This was to

facilitate the identification of chimeric transcripts which are polyadenylated at the correct

site. Second, each gene should not produce a complex transcript pattern *in planta*. The presence of multiple transcripts could imply tandem or overlapping poly(A) signals which would complicate the analysis in this study. Third, the transcript of the selected gene should accumulate to high levels. Finally, there should be an absence of restriction sites which would interfere with the cloning of the poly(A) signal. The 3'UTRs of the selected genes were compared to ensure the poly(A) signals differed in sequence and that the putative FUEs and NUEs had differing characteristics.

A PCR approach, as shown in Figure 3-1B, was used to obtain a DNA fragment containing the polyadenylation signal from the genomic clone of each gene listed in Figure 3-1A. Since the exact location of the polyadenylation signal for each of these genes had not been previously mapped, the region from the stop codon to 160 bp downstream of the most 3' poly(A) addition site was amplified. There is no known downstream element in plant polyadenylation signals, however, the 160bp downstream region was included because inclusion of the region has been shown to enhance processing at the proper poly(A) sites in the octopine synthase and E9 polyadenylation signals (Hunt and MacDonald, 1989; MacDonald et al., 1991). Sequence composition and/or potential secondary structure in the downstream region might influence processing efficiency. The DNA fragments containing the plant polyadenylation signals were inserted upstream of the E9 polyadenylation signal, similar to the globin-*B.t.* chimeric genes. The E9 polyadenylation signal acts as a "trap" for transcripts not polyadenylated at the "test" site.
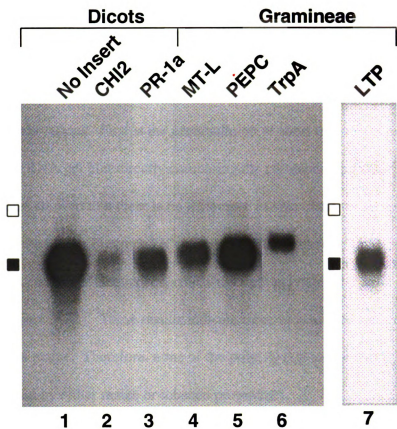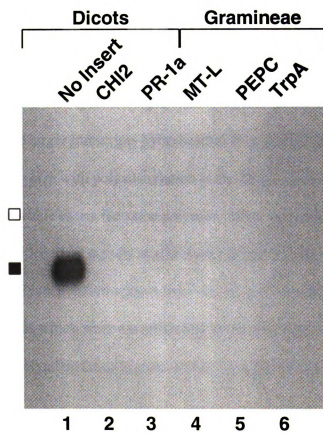
The chimeric genes were transiently expressed in maize protoplasts. Each chimeric gene, except the gene containing the cucumber chitinase poly(A) signal, produces a

single transcript in maize protoplasts as shown in Figure 3-6A. Utilization of the

downstream E9 polyadenylation signal in the chimeric genes would produce transcripts

between 1.3 and 1.6 kb in size (CHI2$\cong$1.35kb, PR-1a$\cong$1.42kb, MT-L$\cong$1.45kb,

PEPC$\cong$1.41kb, TrpA$\cong$1.55kb, and LTP$\cong$1.43kb). Instead, transcripts ranging in size from

approximately 0.9 to 1.2 kb accumulate (No Insert$\cong$0.985kb, CHI2$\cong$0.940kb, PR-

1a$\cong$0.990kb, MT-L$\cong$1.040kb, PEPC$\cong$0.987kb, TrpA$\cong$1.137kb, and LTP$\cong$1.021kb). These

sizes are consistent with polyadenylation in the introduced region. Rehybridization of the

blot in Figure 3-6A with the E9 3'UTR detected only the transcript from the control gene

(Figure 3-6B). The lack of E9 hybridization to the chimeric transcripts is expected with

polyadenylation in the introduced regions. The correspondence in size of the chimeric

transcripts with the known polyadenylation sites within the inserts and the absence of E9

sequences argues against a role for internal splicing in the production of the chimeric

transcripts. Therefore, the short transcripts which accumulate in maize protoplasts

expressing the chimeric genes are a result of polyadenylation at the introduced sites.

The nature of the smaller transcript produced in maize protoplasts expressing the

CHI2 chimeric gene is unknown, but was consistently observed in repeated experiments.

The transcript also accumulates in wheat and Arabidopsis as shown in Figure 3-8A and B,

lane 2). Probing the maize and wheat gel blots with the E9 3'UTR showed no

hybridization to the smaller transcript (Figure 3-6B, lane 2 and data not shown) This

indicates polyadenylation at a cryptic upstream site may be responsible for the presence of

the transcript.

Figure 3-6. Maize Protoplasts Efficiently Utilize the Introduced Polyadenylation Signals

Total RNA was isolated from maize protoplasts transiently expressing the chimeric genes containing the cucumber chitinase (CHI2), tobacco pathogenesis-related protein (PR-1a), maize metallothionein-like (MT-L), maize phosphoenolpyruvate carboxylase (PEPC), maize tryptophan synthase A (TrpA), and barley lipid transfer protein (LTP) polyadenylation signals. The lanes are also numbered across the bottom. RNA gel blots were probed with the globin coding region (A) and then rehybridized with the E9 3'UTR (B). The LTP sample was blotted from a separate gel. Closed boxes indicate the position of the globin/E9 control transcript (No Insert) on each blot. Open boxes indicate the position expected for chimeric transcripts polyadenylated in the E9 3'UTR. The nature of the smaller transcript in the CHI2 sample is presently unknown, but was consistently observed in repeated experiments. The smaller transcript in the LTP sample was inconsistently observed.

Tobacco Protoplasts Do Not Efficiently Utilize Some Gramineae Polyadenylation Signals
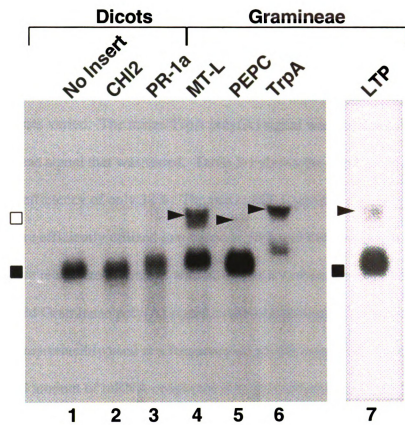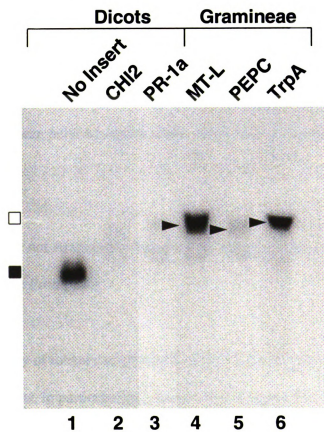

Transient expression of the same chimeric genes in tobacco protoplasts resulted in

two important observations. First is the accumulation of short transcripts as shown in

Figure 3-7A. An RNA gel blot directly comparing the corresponding short transcripts in

tobacco and maize showed that there is no difference in size (data not shown).

Hybridization of the blot in Figure 3-7A with the E9 3'UTR showed these tobacco

transcripts do not contain E9 sequences, consistent with polyadenylation at the introduced

poly(A) sites (Figure 3-7B). These results indicate tobacco is able to recognize the same

poly(A) signals as maize. Therefore, none of the poly(A) signals used in this study are

exclusively utilized by either maize or tobacco protoplasts.

The second important observation made with tobacco protoplasts expressing the

chimeric genes is the accumulation of larger transcripts not detected in maize. The

additional, larger transcripts were detected in protoplasts expressing the genes with the

maize MT-L, maize PEPC, maize TrpA and barley LTP poly(A) signals (Figure 3-7A,

lanes 4-7). These larger transcripts hybridized to E9 probes (Figure 3-7B, lanes 4-6) and

were consistent in size with polyadenylation at the E9 poly(A) sites. Therefore, even

though tobacco is able to use the same polyadenylation signals as maize, tobacco is not

always able to utilize these signals as effectively as maize, supporting the observations

made with the polyadenylation signals from the *B.t.* toxin coding region. Interestingly,

the poly(A) signals which were not efficiently utilized by tobacco, the MT-L, PEPC,

TrpA, and LTP polyadenylation signals, are all from gramineous plants. The dicot

Figure 3-7. Tobacco Protoplasts Do Not Efficiently Utilize the Same Polyadenylation Signals As Maize Protoplasts.

The chimeric genes containing the chitinase (CHI2), pathogenesis-related protein (PR-1a), metallothionein-like (MT-L), phosphoenolpyruvate carboxylase (PEPC), tryptophan synthase A (TrpA), and lipid transfer protein (LTP) polyadenylation signals were transiently expressed in tobacco protoplasts. Total RNA was blotted (the LTP sample from a separate gel) and probed with the globin coding region (A) and then reprobed with the E9 3'UTR (B). The position of the control transcript (No Insert) on each blot is indicated by the closed boxes. The open boxes correspond to the expected position of the chimeric transcripts polyadenylated in the 3'UTR and the arrowheads indicate the actual position. The lane numbers are indicated across the bottom of the blot.

A

Dicots | Grameneae

No Insert · CHI2 · PR-1a · MT-L · PEPC · TrpA · LTP

1 2 3 4 5 6 7

B

Dicots | Grameneae

No Insert · CHI2 · PR-1a · MT-L · PEPC · TrpA

1 2 3 4 5 6

polyadenylation signals were efficiently recognized. This supports a maize/tobacco difference in poly(A) signal utilization.

The degree with which the gramineous polyadenylation signals were utilized in tobacco protoplasts varied. The maize TrpA poly(A) signal was consistently the least utilized Gramineae signal that was tested. Table 3-1 shows the TrpA poly(A) site was utilized with an efficiency of only 38%. The maize MT-L and barley LTP poly(A) signals were more efficiently utilized in tobacco at 58% and 66%, respectively. These same signals were recognized in maize with an efficiency of greater than 95%. The most efficiently utilized Gramineae poly(A) signal in tobacco protoplasts was the PEPC signal. This signal was reproducibly used at a frequency of greater than 95%. Unlike maize, however, a small amount of mRNA consistent with polyadenylation in the E9 3'UTR can be detected (Figure 3-7B, lane 5, denoted by the arrow), indicating utilization of the PEPC poly(A) signal is not as efficient as in maize protoplasts. These data indicate tobacco utilize different Gramineae poly(A) signals with different efficiencies. Therefore, not all poly(A) signals are the same in Gramineae. There must be differences between Gramineae poly(A) signals which affect their utilization in tobacco.

Wheat Cells, but not Arabidopsis Seedlings, Efficiently Utilize the Same Polyadenylation Signals as Maize Protoplasts

The inability of tobacco to efficiently utilize some of the poly(A) signals of gramineous plants, in particular the barley LTP polyadenylation signal, suggests the

differential utilization of poly(A) signals may extend beyond a maize/tobacco difference.

The difference may be more general and divided between gramineous and dicotyledonous

plants. To test this hypothesis, the chimeric genes were transiently expressed in wheat

cells and stably expressed in transgenic Arabidopsis seedlings. The transcript pattern

produced in wheat cells is essentially identical to that produced in maize protoplasts, as

shown in Figure 3-8A. An RNA gel blot directly comparing the sizes of each transcript

generated in wheat and in maize cells expressing the chimeric genes with the plant

poly(A) signals showed no difference (data not shown). The lack of any detectable larger

transcripts suggests each poly(A) signal is utilized with approximately the same

efficiency in wheat whether the signal is from gramineous or dicotyledonous plant

species. Rehybridization of the left panel in Figure 3-8A with the E9 3'UTR confirmed

the absence of E9 sequences in these chimeric transcripts (data not shown). These results

indicate maize and wheat are able to utilize the same polyadenylation signals. The

efficiency with which each poly(A) signal is utilized also is the same between maize and

wheat (Table 3-1). This is consistent with a Gramineae/dicot difference in poly(A) signal

utilization.

The transcript pattern of Arabidopsis is more similar to the transcript pattern of

tobacco. Transcripts consistent in size with polyadenylation at the appropriate sites in the

maize MT-L and TrpA inserts can be detected in lanes 4 and 6 of Figure 3-8B. Larger

transcripts can also be detected which correspond to polyadenylation at the E9 poly(A)

sites. Similar results were obtained expressing the chimeric gene with the segment 2

polyadenylation signal from the *crylA(c)* coding region (Figure 3-8B, lane 7). Short and

large transcripts which correspond to polyadenylation at the segment 2 and E9 poly(A)
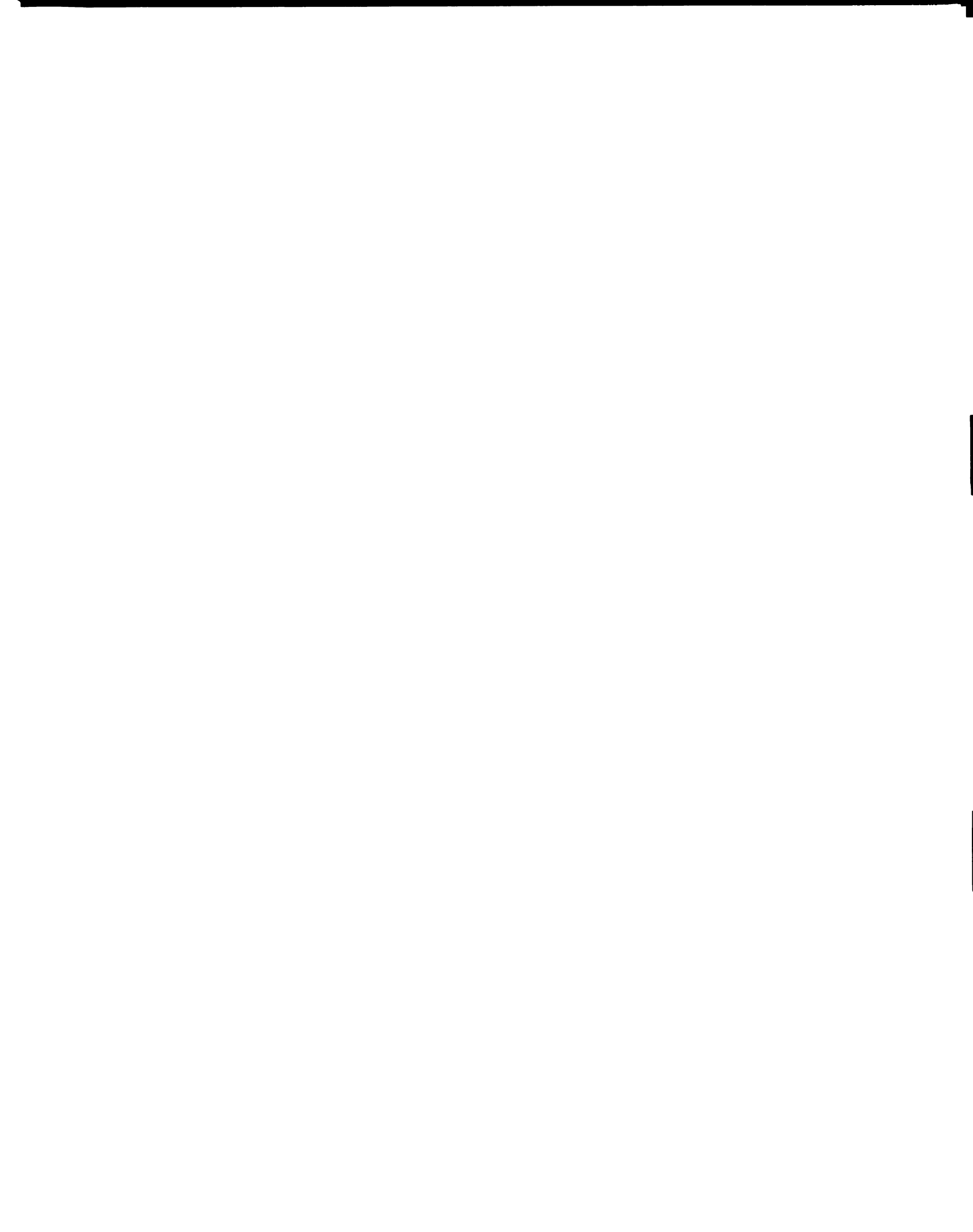
Figure 3-8. Wheat Cells and Arabidopsis Seedlings Differ in the Efficiency of Poly(A) Signal Utilization.

Gel blots of total RNA isolated from wheat cells (A) or Arabidopsis seedlings (B) expressing the indicated chimeric genes were hybridized with the globin coding region. The closed boxes indicate the position of the control transcript on each blot and the arrowheads indicate the chimeric transcripts which are the size expected for polyadenylation in the E9 3'UTR. The lanes are numbered across the bottom of the blots in each panel. Wheat and Arabidopsis accumulate the smaller CHI2 transcript observed in maize protoplasts. The nature of this transcript is unknown.

A

Dicots   Gramineae   *B.t.* toxin

No Insert   CHI2   PR-1a   MT-L   PEPC   TrpA   2   4

1   2   3   4   5   6   7   8

B

Dicots   Gramineae   *B.t.* toxin

No Insert   CHI2   LTP   MT-L   PEPC   TrpA   2   4

1   2   3   4   5   6   7   8
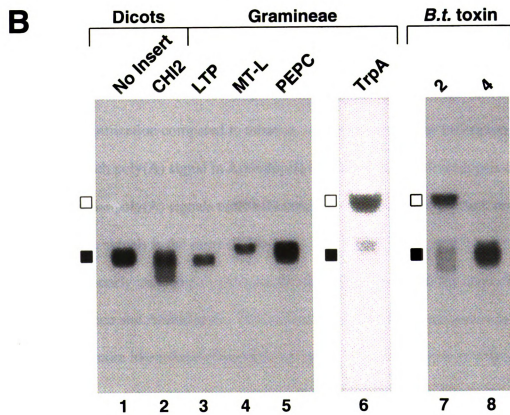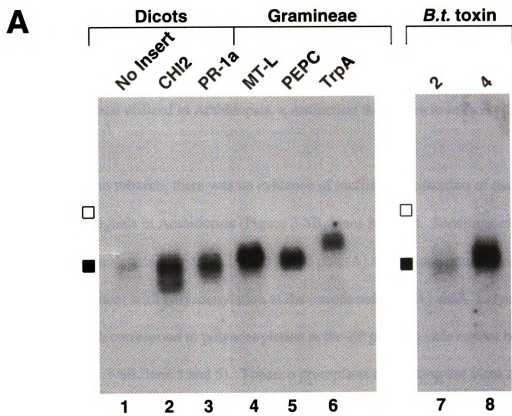
sites, respectively, accumulate in Arabidopsis seedlings. This suggests the MT-L, TrpA, and the segment 2 polyadenylation signals are not efficiently recognized in Arabidopsis. Based on the apparent ratios of the large transcripts to the short transcripts these poly(A) signals are the least utilized in Arabidopsis, a distinction they share in tobacco (Table 3-1).

In contrast to tobacco, there was no evidence of inefficient utilization of the LTP and PEPC poly(A) signals in Arabidopsis (Figure 3-8B, lanes 3 and 5). Seedlings expressing the chimeric genes containing the LTP and PEPC poly(A) signals accumulate transcripts which are consistent with polyadenylation at the introduced poly(A) sites. Larger transcripts which correspond to polyadenylation at the E9 poly(A) sites cannot be detected (Figure 3-8B, lane 3 and 5). Tobacco protoplasts expressing the same chimeric genes accumulate these larger transcripts. However, the relative level of the large PEPC and LTP transcripts in tobacco was less than observed for MT-L, TrpA, and segment 2 (compare Figure 3-8B, lanes 3 and 5 to Figure 3-7A, lanes 4 and 6). Therefore, the lack of the transcripts in Arabidopsis seedlings could be a reflection of a slight difference in poly(A) signal utilization compared to tobacco. A comparison of the utilization efficiency of each poly(A) signal in Arabidopsis to tobacco shows Arabidopsis does utilize each of the poly(A) signals more efficiently than tobacco, but the rank order of the polyadenylation signals is the same (Table 3-1). The TrpA and segment 2 poly(A) sites were the most poorly utilized poly(A) sites and the PEPC signal was the most efficiently utilized in tobacco and Arabidopsis. This indicates poly(A) signal utilization in Arabidopsis is more like tobacco than maize or wheat. The similarities in poly(A) signal utilization between maize and wheat compared to tobacco and Arabidopsis suggest

134

Table 3-1. The Efficiency of Poly(A) Signal Utilization in Maize, Wheat, Tobacco, and Arabidopsis

The utilization of each introduced polyadenylation signal in each plant system was calculated by determining the ratio of the amount of transcript polyadenylated at the introduced poly(A) sites versus the total amount of transcript produced, expressed as a percentage. An average of at least two experiments for each chimeric gene in each plant system is represented in the table. The polyadenylation signals utilized with an efficiency of greater than 95% are indicated by a "+". The accumulation of two transcripts in plant cells expressing the chimeric gene containing the cucumber chitinase polyadenylation signal is indicated by the asterisk.

3' UTR	Source	Maize	Wheat	Tobacco	Arabidopsis
Dicot					
E9	Pea	+	+	+	+
PR1a	Tobacco	+	+	+	N.D.
Chitinase	Cucumber	+*	+*	+	+*
Gramineae					
LTP	Barley	+	N.D.	66%	+
PEPC	Maize	+	+	+	+
MT	Maize	+	+	58%	88%
TrpA	Maize	+	+	38%	43%
Bacteria					
Segment 2	B.t.	+	+	23%	46%
Segment 4	B.t.	+	+	+	+

Gramineae plant species have a less stringent requirement for poly(A) signal utilization than dicot plant species.

The difference in the efficiency of poly(A) signal utilization between tobacco protoplasts and Arabidopsis seedlings is not likely to be due to transient expression of the genes versus stable integration into the genome. The relative abundance of the transcripts polyadenylated at the segment 2 poly(A) site in transgenic plants and transformed cell cultures did not show a substantial difference to the relative abundance of the same transcript in protoplasts (16%-vs-23%, respectively). This may indicate tobacco has a more stringent requirement for poly(A) signal recognition than Arabidopsis, but does not alter our observations that gramineous plants have a less stringent requirement for poly(A) signal utilization than dicotyledonous plants.

DISCUSSION

In this study we demonstrate that polyadenylation signals are utilized differently in

gramineous and dicotyledonous plant species. The polyadenylation signals from

gramineous plant genes and a polyadenylation signal from the *crylA(c) B.t.* toxin coding

region are efficiently utilized in maize and wheat, but not in tobacco and Arabidopsis.

This indicates gramineous plants have a less stringent requirement for poly(A) signal

utilization than dicots. Elucidating the underlying differences of poly(A) signal

utilization could be important in advancing our understanding of the differences in gene

expression mechanisms between monocots and dicots. This understanding may also have

important implications in biotechnology where poly(A) signal utilization could have a

major impact on foreign gene expression in plants.

The utilization of polyadenylation signals in monocots versus dicots has been an

open issue. There are reports which suggest that there is no difference in poly(A) signal

recognition between monocots and dicots. For example, the maize 27kD zein transcript

and the wheat histone H3 transcript are polyadenylated at the expected sites in tobacco

protoplasts and sunflowers cells, respectively (Tabata et al., 1987; Wu et al., 1994). In

contrast, the transcript of a wheat *rbcS* gene was improperly polyadenylated in transgenic

tobacco plants. Only 50% of the wheat transcripts were polyadenylated at the site

normally utilized by wheat. The rest were polyadenylated at novel sites located further

downstream (Keith and Chua, 1986). A wheat histone H4 transcript in sunflower cells

did not map to the expected sites, however, this was reasoned to be due to the lack of all

the necessary sequences for proper 3' end formation (Tabata et al., 1987).

The overall structure of plant polyadenylation signals has not provided any insight to resolve this issue. Both monocot and dicot polyadenylation signals share a common FUE/NUE architecture. In addition, the FUEs of different plant polyadenylation signals are interchangeable. The FUE of the Cauliflower Mosaic Virus can substitute for the FUE of the maize 27kD zein gene, the Figwort Mosaic Virus, and the pea *rbcS-E9* polyadenylation signals (Mogen et al., 1992; Sanfacon, 1994; Wu et al., 1994). Also, the FUE of the maize 27kD zein gene can substitute for the FUE of the Cauliflower Mosaic Virus (Wu et al., 1994). This implies that there may not be any distinctive feature between the elements of polyadenylation signals from monocots and dicots and supports a functional conservation of poly(A) signal utilization in both groups of plants.

The present report, to the best of our knowledge, represents the most extensive effort to date aimed at addressing poly(A) signal utilization in monocots (Gramineae) versus dicots. The study was initiated while elucidating the mechanisms limiting the expression of a *crylA(c) B.t.* toxin gene in tobacco. The transcript for this gene fails to accumulate due to mRNA instability and the presence of premature polyadenylation sites in the coding region. In an attempt to further analyze the poly(A) signals using the chimeric genes described in Figure 3-1C, it was observed that one poly(A) addition site, the segment 2 poly(A) site, was more efficiently utilized in maize than in tobacco (Figure 3-2). The inability of tobacco to efficiently polyadenylate at the segment 2 poly(A) addition site used in maize suggested a maize/tobacco difference or possibly a monocot (Gramineae)/dicot difference in poly(A) signal utilization.

Prior to our analysis of the segment 2 poly(A) addition site, the only other polyadenylation signal to our knowledge that had been reported to be efficiently utilized

in one plant system and inefficiently in another was the wheat *rbcS* polyadenylation signal (Keith and Chua, 1986). This meant that other poly(A) signals needed to be identified to support our hypothesis that poly(A) signals are differentially utilized in maize and tobacco. As a first approach, the polyadenylation signals from dicot and monocot plant genes, listed in Figure 3-1A, were chosen based on the criteria previously described in the Results section. Expression of chimeric genes containing the polyadenylation signals from these plant genes in maize and tobacco protoplasts showed some of the polyadenylation signals were not efficiently utilized in tobacco. Each of these signals was from gramineous plant species which supported a monocot (Gramineae)/dicot difference in poly(A) signal utilization.

Often differences between monocots and dicots are based on conclusions drawn from studies using only one monocot species and one dicot species. This approach has the risk of only distinguishing differences between the two species and may not actually represent differences between monocots and dicots in general. Therefore, the chimeric genes containing the introduced polyadenylation signals were expressed in wheat cells and Arabidopsis seedlings in addition to maize and tobacco in this study. The goal was to determine whether the difference in poly(A) signal utilization was a maize/tobacco difference or actually a Gramineae/dicot difference. Maize and wheat produced a transcript pattern that was essentially identical. The size of the transcripts and the lack of E9 sequences argue that polyadenylation occurred at the appropriate sites in these transcripts, regardless of the source of the polyadenylation signal. Tobacco and Arabidopsis produced a transcript pattern that was different from maize and wheat. Some of the gramineous polyadenylation signals were not efficiently utilized in these two dicot

systems, lending support to a monocot/dicot difference in poly(A) signal utilization in which monocots have a less stringent requirement for poly(A) signal utilization than dicots.

Interestingly, tobacco was less efficient than Arabidopsis in utilizing most of the Gramineae polyadenylation signals. However, the rank order of signal utilization between tobacco and Arabidopsis was similar. This suggested Arabidopsis is more like tobacco than maize or wheat in poly(A) signal utilization. The difference in the degree of selectivity could be due to a difference between transient and stable expression of the genes. However, this is unlikely since the relative abundance of the transcripts polyadenylated at the segment 2 site in transgenic tobacco plants and transformed cultured cells showed no substantial difference compared to the relative abundance in protoplasts. Therefore, the degree of selectivity could be simply species to species variation.

Our observation that gramineous plants have a less stringent requirement for poly(A) signal recognition than dicots parallels with other observed differences in RNA processing between monocots and dicots. Specifically, the splicing of nuclear pre-mRNA differs between these two classes of plants. Dicot introns are generally spliced efficiently in monocots, however, monocot introns are often less efficiently spliced in dicots (reviewed in Luehrsen et al., 1994; Filipowicz et al., 1995). The difference in splicing efficiency has been suggested to be partly a result of the difference in UA composition of monocot and dicot introns. Several surveys of plant introns have shown dicot introns are approximately 70-75% UA-rich on average and maize introns are generally 60-65% UA-rich (Goodall and Filipowicz, 1991; White et al., 1992; Luehrsen et al., 1994). In

addition, a synthetic intron with a high AU content is efficiently excised in tobacco.

However, as the GC content of the intron was increased, the efficiency of excision

decreased (Goodall and Filipowicz, 1991). The negative effect of GC content on splicing

in tobacco was more severe than in maize cells.

Since the nucleotide composition of plant introns provided insight to the differences

in splicing efficiency between monocots and dicots, it was possible that a similar analysis

of efficiently and inefficiently utilized poly(A) signals would reveal particular

characteristics that could explain the results presented in this report. Each of the

polyadenylation signals tested in this study differed in primary sequence and had putative

FUEs and NUEs that were similar or dissimilar to FUEs and NUEs from characterized

poly(A) signals. After extensive analysis of the nucleotide composition of the FUE,

NUE, and surrounding sequences no direct correlation could be made between the

sequences and the utilization of the poly(A) signals in gramineous and dicotyledonous

plants. However, it was interesting that the putative FUEs of the polyadenylation signals

which were inefficiently utilized in tobacco had a G content that was approximately 3-8%

higher than the surrounding sequences (from the cleavage site to the stop codon)

compared to the putative FUEs of polyadenylation signals inefficiently utilized in

tobacco. In addition, a comparison of the sequences upstream of the cleavage site (stop

codon to cleavage site) to the sequences downstream of the cleavage site (cleavage site to

3' end of amplified region) for each polyadenylation signal inefficiently utilized in

tobacco showed generally a 6-8% decrease in the G content in the downstream sequences.

There was essentially no change in the G content between the sequences upstream and

downstream of the cleavage site for the polyadenylation signals efficiently utilized in

tobacco. Whether these small difference in G content are significant is not known.

Spacing between elements, in particular between the NUE and the cleavage site, had

been proposed as a possible reason for improper poly(A) signal utilization (Hunt, 1994).

This hypothesis was based on a comparison of the maize 27kD zein polyadenylation

signal to three other dicot polyadenylation signals. However, the 27kD zein

polyadenylation signal has been shown to be efficiently utilized in transformed tobacco

protoplasts (Wu et al., 1994). Also, the wheat histone H3 polyadenylation signal which

has a similar NUE-cleavage site spacing is also efficiently utilized in transformed

sunflower cells (Tabata et al., 1987). A comparison of the spacing between the FUEs and

NUEs from other published Gramineae and dicot polyadenylation signals does not

provide any further insight which could explain why a poly(A) signal might be utilized in

Gramineae and not in dicots. Interestingly, the insertion of unrelated DNA sequences as

short as 17bp between the FUE and NUEs or between the two NUEs of the maize 27kD

zein gene drastically reduced polyadenylation efficiency in maize cells (Wu et al., 1994),

indicating spacing between these elements can affect poly(A) signal utilization, but these

derivatives were not tested in dicots.

The differences between efficiently and inefficiently utilized poly(A) signals do not

appear to be distinguishable from sequence analysis alone. This suggests a mechanistic

difference in the recognition of poly(A) signals between Gramineae and dicots. The

conserved FUE/NUE architecture of plant polyadenylation signals in monocots and dicots

could imply the basal polyadenylation factors in both groups of plants are also conserved.

Thus, the difference in poly(A) signal utilization may lie with the presence of auxiliary

factors. Such factors may allow the monocot polyadenylation machinery to be more flexible and able to recognize a wider range of sequences as polyadenylation signals. Dicots may lack these auxiliary factors or have homologues which do not function efficiently with some monocot polyadenylation signals. Alternatively, the basal polyadenylation factors actually may not be conserved between monocots and dicots. Differences may exist which allow the basal factors in monocots to have more flexibility to recognize signals with subtle variations in structure, whereas in dicots these variations are less tolerated.

To fully understand how polyadenylation signals are differentially utilized in plants, the components of the polyadenylation machinery will have to be identified. With the availability of several EST databases for a variety of plant species and the cloning of many of the genes encoding the cleavage and polyadenylation factors from mammals and yeast, the plant homologues should soon be identified. To date there are two ESTs that correspond to known polyadenylation factors. One corresponds to a subunit of the mammalian cleavage and polyadenylation specificity factor (CPSF), the 100kD subunit, and the other is a rice EST that has significant similarity to the yeast and mammalian poly(A) polymerase. Also, several plant poly(A) binding proteins (PABs) have been cloned. Although the *in vivo* function of these PABs is largely unknown, PABs in mammals and yeast, PABII and Pab1p, respectively, have been implicated in controlling poly(A) tail length (Amrani et al, 1997; Minvielle-Sebastia et al., 1997). No homologs of PABII have been found in plants, but a wheat PAB and an Arabidopsis PAB, PAB5, partially complement the *pab1* mutant in yeast (Belostotsky and Meagher, 1996; Le et al., 1997). Although few factors involved in polyadenylation have been identified in plants,

more components will be identified as the genomes of plants continue to be sequenced. This will be important since several questions concerning the basic mechanism of 3' end formation in plants and its similarity to other systems such as mammals and yeast need to be addressed. The answers to these questions will advance our understanding of the differences in poly(A) signal utilization between Gramineae and dicots and will be necessary for the elucidation of the mechanism responsible for differential poly(A) signal utilization.

REFERENCES

Amrani, N., Minet, M., Le Gougar, M., Lacroute, F., and Wyers, F. (1997). Yeast Pab1 Interacts with Rna15 and Participates in the Control of the Poly(A) Tail Length In Vitro. Mol. Cell. Biol. 17, 3694-3701.

Chang, Y.-F., Wang, W.C., Warfield, C.Y., Nguyen, H.T., and Wong, J.R. (1991). Plant Regeneration from Protoplasts Isolated from Long-Term Cell Cultures of Wheat (*Triticum aestivum* L.). Plant Cell Rep 9, 611-614.

Colgan, D. and Manley, J.L. (1997). Mechanism and Regulation of mRNA Polyadenylation. Genes & Dev. 11, 2755-2766.

de Framond, A.J. (1991). A Metallothionein-like Gene From Maize (Zea Mays). FEBS 290, 103-106.

De Rocher, E.J., Vargo-Gogola, T.C., Diehn, S.H., and Green, P.J. (1998). Direct Evidence for Rapid Degradation of *B.t.* toxin mRNA as a Cause of Poor Expression in Plants. (manuscript submitted).

Diehn, S., Chiu, W., DeRocher, E.J., and Green, P.J. (1998). Identification of Multiple Plant Poly(A) Addition Sites Within a B.t. Toxin Coding Region. (manuscript submitted).

Eckner, R., Ellmeier, W., and Birnstiel, M.L. (1991). Mature mRNA 3' End Formation Stimulates RNA Export from the Nucleus. EMBO J 10, 3513-3522.

Hudspeth, R.L. 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.

Hunt, A.G. (1994). Messenger RNA 3' End Formation in Plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 45, 47-60.

Hunt, A.G., Chu, N.M., Odell, J.T., Nagy, F., and Chua, N. (1987). Plant Cells do not Properly Recognize Animal Gene Polyadenylation Signals. Plant Mol Biol 8, 23-35.

Hunt, A.G. and MacDonald, M.H. (1989). Deletion Analysis of the Polyadenylation Signal of a Pea Ribulose-1,5-Bisphosphate Carboxylase Small-subunit Gene. Plant Mol Biol 13, 125-138.

Jackson, R.J. and Standart, N. (1990). Do the Poly(A) Tail and 3' Untranslated Region Control mRNA Translation? Cell 62, 15-24.

Jacobson, A. (1996). Poly(A) Metabolism and Translation: The Closed-Loop Model. Translational Control 451-479.

Jacobson, A. and Peltz, S.W. (1996). Interrelationships of the Pathways of mRNA Decay and Translation in Eukaryotic Cells. Annual Review of Biochemistry 65,

Keith, B. and Chua, N. (1986). Monocot and Dicot Pre-mRNAs Are Processed With Different Efficiencies in Transgenic Tobacco. EMBO J 5, 2419-2425.

Kramer, V.C. and Koziel, M.G. (1995). Structure of a Maize Tryptophan Synthase Alpha Subunit Gene With Pith Enhanced Expression. Plant Mol Biol 27, 1183-1188.

Lawton, K.A., Beck, J., Potter, S., Ward, E., and Ryals, J. (1994). Regulation of Cucumber Class III Chitinase Gene Expression. Molecular Plant-Microbe Interactions 7, 48-57.

Li, Q. and Hunt, A.G. (1997). The Polyadenylation of RNA in Plants. Plant Physiol. 115, 321-325.

Luehrsen, K.R., Taha, S., and Walbot, V. (1994). Nuclear Pre-mRNA Processing in Higher Plants. Progress in Nucleic Acid Research and Molecular Biology 47, 149-193.

MacDonald, M.H., Mogen, B.D., and Hunt, A.G. (1991). Characterization of the Polyadenylation Signal from the T-DNA-Encoded Octopine Synthase Gene. Nucleic Acids Research 19, 5575-5581.

McDevitt, M.A., Hart, R.P., Wong, W.W., and Nevins, J.R. (1986). Sequences Capable of Restoring Poly(A) Site Function Define Two Distinct Downstream Elements. EMBO J 5, 2907-2913.

Minvielle-Sebastia, L., Preker, P., Wiederkehr, T., Strahm, Y., and Keller, W. (1997). The Major Yeast Poly(A)-Binding Protein is Associated with Cleavage Factor IA and Functions in Premessenger RNA 3'-end Formation. Proc. Natl. Acad. Sci. USA 94: 7897-7902.

Mogen, B.D., MacDonald, M.H., Graybosch, R., and Hunt, A.G. (1990). Upstream Sequences Other than AAUAAA Are Requested for Efficient Messenger RNA 3'-End Formation in Plants. Plant Cell 2, 1261-1272.

Mogen, B.D., MacDonald, M.H., Leggewie, G., and Hunt, A.G. (1992). Several Distinct Types of Sequence Elements Are Required For Efficient mRNA 3' End Formation in a Pea rbcS Gene. Mol. Cell. Biol. 12, 5406-5414.

Newman, T.C., Ohme-Takagi, M., Taylor, C.B., and Green, P.J. (1993). DST Sequences, Highly Conserved among Plant SAUR Genes, Target Reporter Transcripts for Rapid Decay in Tobacco. Plant Cell 5, 701-714.

Payne, G., Parks, T.D., Burkhart, W., Dincher, S., Ahl, P., and Metraux, J.P. (1988). Isolation of the Genomic Clone for Pathogenesis-related Protein 1a From Nicotiana tabacum cv. Xanthi-nc. Plant Mol Biol 11, 89-94.

Proudfoot, N. (1991). Poly(A) Signals. Cell 64, 671-674.

Proudfoot, N.J. (1989). How RNA Polymerase II Terminates Transcription in Higher Eukaryotes. TIBS 14, 105-110.

Rothnie, H.M. (1996). Plant mRNA 3'-End Formation. Plant Mol Biol 32, 43-61.

Rothnie, H.M., Reid, J., and Hohn, T. (1994). The Contribution of AAUAAA and the Upstream Element UUUGUA to the Efficiency of mRNA 3'-End Formation in Plants. EMBO J. 13, 2200-2210.

Sambrook, J., Fritsch, E.J., and Maniatis, T. (1989). Molecular Cloning; a laboratory manual. 2nd ed. Cold Spring Harbor Press, Cold Spring Harbor, NY.

Sanfacon, H. (1994). Analysis of Figwort Mosaic Virus (Plant Pararetrovirus) Polyadenylation Signal. Virol. 198, 39-49.

Sanfacon, H., Brodmann, P., and Hohn, T. (1991). A Dissection of the Cauliflower Mosaic Virus Polyadenylation Signal. Genes & Dev. 5, 141-149.

Sheets, M.D., Ogg, S.C., and Wickens, M.P. (1990). Point Mutations in AAUAAA and the Poly (A) Addition Site: Effects on the Accuracy and Efficiency of Cleavage and Polyadenylation in Vitro. Nucleic Acids Research 18, 5799-5805.

van Hoof, A., and Green, P.J. (1996). Premature Nonsense Codons Decrease the Stability of Phytohemagglutinin mRNA in a Position-Dependent Manner. Plant J. 10, 415-424.

Wahle, E. (1992). The End of the Message: 3'-End Processing Leading to Polyadenylated Messenger RNA. BioEssays 14, 113-118.

Wahle, E. and Keller, W. (1992). The Biochemistry of 3'-End Cleavage and Polyadenylation of Messenger RNA Precursors. Annu. Rev. Biochem. 61, 419-440.

Wahle, E. and Keller, W. (1996). The Biochemistry of Polyadenylation. TIBS 21, 247-250.

White, A.J., Dunn, M.A., Brown, K., and Hughes, M.A. (1994). Comparative Analysis of Genomic Sequence and Expression of a Lipid Transfer Protein Gene Family in Winter Barley. Journal of Experimental Botany 45, 1885-1892.

Wickens, M. (1990). How the Messenger Got Its Tail: Addition of Poly(A) in the Nucleus. TIBS 15, 277-281.

Wu, L., Ueda, T., and Messing, J. (1995). The Formation of mRNA 3'-Ends in Plants. Plant J. 8, 323-329.

Chapter 4

SUMMARY AND CONCLUSIONS

At the onset of the research described in this thesis, little was known about what post-transcriptional mechanisms could limit foreign gene expression at the level of mRNA accumulation in plants. There were few published reports describing low mRNA accumulation and hence, low expression of foreign genes in plants. However, there was one group of genes well-known for their poor expression in plants, the *B.t.* toxin genes from *Bacillus thuringiensis*. The development of plant transformation techniques and the agronomic potential of these insecticidal proteins expedited the cloning of *B.t.* toxin genes. It was apparent after the first transformation events, however, that these genes would not be expressed well in plants because the transcripts failed to accumulate. The problem was eventually overcome by resynthesizing the genes and changing the codon bias to make the genes more "plant-like". Unfortunately, these gross changes did not make it possible to elucidate the mechanisms responsible for the low transcript accumulation.

The goal of the research described in chapter two of this thesis was to understand the mechanisms limiting *B.t.* toxin transcript accumulation in plants. This was to be a first step in understanding the mechanisms which limit foreign gene expression in plants as the same mechanism could likely limit the expression of other foreign genes. Studying *crylA(c) B.t.* toxin gene expression in tobacco provided the first evidence that *B.t.* toxin genes contain sequences that can be recognized by plants as polyadenylation signals. More importantly, studying this gene showed that foreign gene expression in plants can be limited by premature polyadenylation.

Premature polyadenylation within the *crylA(c)* coding region is likely to have a significant role in limiting the accumulation of the transcript and expression of the gene

since the 900 and 600nt transcripts described in chapter two can be detected by RNA gel blot analysis. More sensitive techniques such as RNase protection or RT-PCR were not needed to visualize the transcripts. However, the contribution of premature polyadenylation to the low accumulation of the full-length *cryIA(c)* transcript needs to be addressed more quantitatively by systematically inactivating of each of the polyadenylation sites and accessing the contribution of each polyadenylation site to the overall low transcript accumulation. These experiments will also provide valuable information relevant to the modification of novel *B.t.* toxin genes and other foreign genes by determining the extent to which the genes need to be modified in order to achieve high expression levels. For instance, can the FUE or NUE be mutagenized to eliminate polyadenylation at a particular site in the coding region or are both elements required to be inactivated to prevent polyadenylation at a cryptic site. The results from these types of experiments may also be important when modifying a foreign gene for expression in different plant systems as the polyadenylation signals of the *cryIA(c)* coding region are known to be differentially recognized in gramineous and dicotyledonous plants based on the results presented in chapter 3. Thus, different mutations may have separate effects in different plant species.

The differential utilization of polyadenylation signals in gramineous and dicotyledonous plants, described in chapter 3 of this thesis, is the most extensive study to the best of my knowledge that addresses this question. Most plant polyadenylation signals prior to this study were thought to work efficiently in both plant groups based on the structure of the signals and a few studies which interchanged elements or entire polyadenylation signals. It had been previously reported that a polyadenylation signal

from wheat was not efficiently recognized in tobacco plants, but, the generality of differential poly(A) signal utilization was not persued. Chapter 3 of this thesis shows the polyadenylation signals from Gramineae genes are not always efficiently recognized in dicots, suggesting gramineous plants have a less stringent requirement for poly(A) signal recognition than dicots.

Ultimately, the basis of differential poly(A) signal recognition will have to be elucidated. Analysis of the sequence elements used in this study only showed weak correlations (discussed in Chapter 3). Plant polyadenylation signals are complex, redundant, and diffuse making analysis of the sequences difficult. Thus, it may be more prudent to address the question of poly(A) signal recognition by identifying the factors of the plant polyadenylation machinery and developing an *in vitro* polyadenylation system. Surprisingly, an *in vitro* cleavage and polyadenylation system has not been developed for plants and the only factor of the polyadenylation machinery which has been cloned is poly(A) polymerase. This contrasts yeast and mammalian systems where *in vitro* systems were developed almost 15 years ago and many factors of the polyadenylation machinery have been cloned over the past few years. Currently, there is much focus in the yeast and mammalian systems aimed at understanding the interactions of these factors, not only with the sequence elements, but with each other. The plant field desperately needs to move beyond the sequence elements of plant polyadenylation signals and begin a more intensive effort, aimed at identifying the components of the polyadenylation machinery.

The opportunity for plants to contribute to the field polyadenylation of is now. With the near completion of the sequencing of the Arabidopsis genome and the availability of other EST databases such as rice, the identification and cloning of the plant homologues

of the yeast and mammalian polyadenylation factors should be rapid. Then, the plant

polyadenylation field should be able to advance more quickly and address specific

questions concerning the differences in 3' end formation between plants and mammals.

More important is the potential to discover differences in 3' end formation between

Gramineae and dicots, discoveries which can explain the biological basis of differential

poly(A) signal recognition.