Premature Polyadenylation at Multiple Sites within a *Bacillus thuringiensis* Toxin Gene-Coding Region[1]

Scott H. Diehn[2], Wan-Ling Chiu[3], E. Jay De Rocher, and Pamela J. Green*

Michigan State University-Department of Energy Plant Research Laboratory (S.H.D., W.-L.C., E.J.D.R., P.J.G.),
Department of Botany and Plant Pathology (S.H.D.), and Department of Biochemistry (P.J.G.),
Michigan State University, East Lansing, Michigan 48824

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* (*B.t.*), which encode the insecticidal proteins commonly referred to as *B.t.* toxins. As a step toward overcoming such problems most effectively, we sought to elucidate the mechanisms limiting the expression of a typical *B.t.*-toxin gene, *cryIA(c)*, which accumulates very little mRNA in tobacco (*Nicotiana tabacum*) cells. Most cell lines transformed with the *cryIA(c)* *B.t.*-toxin gene accumulate short, polyadenylated transcripts. The abundance of these transcripts can be increased by treating the cells with cycloheximide, a translation inhibitor that can stabilize many unstable transcripts. Using a series of hybridizations, reverse-transcriptase polymerase chain reactions, and RNase-H-digestion experiments, poly(A$^+$) addition sites were identified in the *B.t.*-toxin-coding region corresponding to the short transcripts. A fourth polyadenylation site was identified using a chimeric gene. These results demonstrate for the first time to our knowledge 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 basic as well as applied significance.

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 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 (*Nicotiana tabacum*), 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 *K. pneumoniae* cyclodextrin glycosyltransferase genes accumulate very low levels of the 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 nucleotides accumulate.

The genes best known for their low expression in plants are the *Bacillus thuringiensis* (*B.t.*)-toxin genes (for review, see Diehn et al., 1996). This family of genes from the gram-positive soil bacterium *B.t.* encodes potent insecticidal proteins that target specific orders of insects (Höfte and Whiteley, 1989; Aronson, 1993). Initial efforts to express *B.t.*-toxin genes in plants using standard approaches yielded transgenic plants that produced little or no *B.t.*-toxin mRNA or protein (Barton et al., 1987; Vaeck et al.,

[2] Present address: Pioneer Hi-Bred International, Inc., Traits and Technology Development, 7300 NW 62nd Avenue, P.O. Box 1004, Johnston, IA 50131–1004.

[3] Present address: Department of Biology, University of Richmond, Richmond, VA 23173.

* Corresponding author; e-mail green@pilot.msu.edu; fax 1–517–355–9298.

Abbreviations: 2X35S, doubly enhanced cauliflower mosaic virus 35S promoter; BY-2, Bright Yellow 2; CaMV, cauliflower mosaic virus; CHX, cycloheximide; FUE, far-upstream element; GFP, green fluorescent protein; NUE, near-upstream element; Py, pyrimidine; RT, reverse transcriptase; UTR, untranslated region.

1987). The few plants generated that did express *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, because the plants that accumulated detectable *B.t.*-toxin mRNA were also 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 of *B.t.*-toxin genes, considerable effort has been made to increase the expression of these genes in plants. These efforts include expressing only the region of the gene encoding the insecticidal domain, modifying the 5′ and 3′ UTRs, generating protein fusions, and using a variety of strong promoters (for review, see Diehn et al., 1996). The mRNA and protein levels were eventually increased by resynthesizing the genes to be more "plant like." In most cases, this included changing the codon usage to a plant-preferred codon bias (for review, see Diehn et al., 1996), which also has the effect of raising the G/C content of the wild-type gene. Many plant RNA-processing signals, in particular those for polyadenylation, mRNA decay, and splicing, are A/T rich. Wild-type *B.t.*-toxin genes have an A/T content of approximately 65%. Therefore, increasing the G/C content of the genes may eliminate potential RNA-processing signals.

Why the transcripts of some poorly expressed foreign genes fail to accumulate in plants remains unclear in most cases. The problem can occur at one or more steps in gene expression. mRNA 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 posttranscriptionally (Fischhoff et al., 1987; Vaeck et al., 1987; Oakes et al., 1991) as a result of aberrant splicing and/or degradation of the transcript. Recently, the transcripts of GFP and a *cryIA(b) B.t.*-toxin gene were found to contain one and three cryptic introns, respectively (Haseloff and Amos, 1995; Van Aarssen et al., 1995; Haseloff et al., 1997). The splicing of these transcripts was shown to be partly responsible for the low expression of the GFP and *cryIA(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 *cryIA(c) B.t.*-toxin gene is polyadenylated in the coding region. Multiple sequence elements in the *B.t.*-toxin-coding region are recognized by plant cells as polyadenylation signals. Recognition of these signals appears to be one factor contributing to the low accumulation of the full-length *B.t.*-toxin transcript in plant cells. Another limitation of *cryIA(c) B.t.*-toxin transcript accumulation in plant cells, posed at the level of mRNA stability, is described in the accompanying paper (De Rocher et al., 1998). Elucidating the mechanisms responsible for the low accumulation of *B.t.*-toxin mRNA in plants may make it easier in the future to engineer novel foreign genes and other *B.t.*-toxin genes for high expres-

sion. It may also provide insight into natural gene-expression mechanisms in plants.

MATERIALS AND METHODS

Plant Materials and Treatments

Tobacco (*Nicotiana tabacum* cv 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 tumefaciens*-mediated transformation, also described by Newman et al. (1993), using *A. tumefaciens* strain LBA4404 harboring the appropriate plasmids. Kanamycin-resistant BY-2 calli transformed with the *cryIA(c) Bacillus thuringiensis* (*B.t.*)-toxin gene (described below) were transferred to fresh plates and then screened for GUS-reporter-gene expression by histochemical staining (Jefferson et al., 1986). Positive cell lines were treated with 50 μg mL^{-1} CHX for 2 h in liquid culture after 5 to 7 d of growth. Treated and untreated cells were pelleted at 1000 rpm for 5 min in a centrifuge (model RT-6000D with a model H-1000B rotor, Sorvall) and then frozen in liquid nitrogen. 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 portion of the *cryIA(c) B.t.*-toxin-coding region encoding the insecticidal domain from *B.t.* subsp. *kurstaki* HD-73 (amino acids 9–613) (Schnepf and Whiteley, 1985) was kindly provided by Dr. A.I. Aronson of Purdue University (West Lafayette, IN). This *B.t.*-toxin gene had been modified previously by fusing the sequence encoding the N-terminal nine amino acids of *LacZ* to the 5′ portion of the *B.t.*-toxin-coding region. The coding region was further modified in our laboratory after it was introduced into a modified pT7/T3α19 vector (GIBCO-BRL) containing the pSP64-poly(A$^+$) multiple-cloning site (Promega). The translation initiation site was altered to conform to the plant consensus sequence (Joshi, 1987; Lütcke et al., 1987), and two Pro codons were added to the 3′ end of the coding region to protect the C terminus of the toxin from proteolytic activity (Bigelow and Channon, 1982). Nucleotide position 31 in the coding region of our gene, which is the first base of the codon for amino acid 9, corresponds to position 415 in the *cryIA(c) B.t.*-toxin sequence file, m11068, of the EMBL/GenBank/DDBJ databases. The first 31 nucleotides of our *B.t.*-toxin-coding region are ATGGCTAT GATTACGCCTAGCTTGCATGCCT.

The resulting plasmid, p995, was digested with *Bgl*II and *Bam*HI to release the *B.t.*-toxin-coding region with its modified 5′ and 3′ ends (lacking the pSP64-poly[A$^+$] multiple cloning site and the poly[A$^+$] sequence). This *B.t.*-toxin-coding region fragment was then used to replace the β-globin-coding region of p1185, which originally contained an expression cassette consisting of a 2X35S-globin-coding region and the 3′ UTR from the pea Rubisco small subunit *rbcS-E9* gene to generate p1204. The construction of

p1185 is described below. The resulting 2X35S-*B.t.* toxin-E9 gene cassette from p1204 was then inserted into the *Hin*dIII site of the binary vector pBI121 (accession no. X77672) to make p1205, which was used for integration into the genome of BY-2 cells

p1185 is a pUC 8 plasmid containing a 2X35S, a globin-coding region, and the pea *rbcS-E9* 3′ UTR that was generated from pMF6 (Goff et al., 1991), a plasmid kindly provided by Michael Fromm at U.S. Department of Agriculture/University of California, Berkeley. To create the 2X35S a copy of the CaMV 35S enhancer contained on a *Hin*cII-*Eco*RV fragment of pMF6 was inserted into the *Eco*RV site of a second pMF6 CaMV 35S promoter. The nopaline-synthase polyadenylation sequence of the modified pMF6 plasmid, now called p1079, was removed by digestion with *Kpn*I and *Sca*I. After creating blunt ends at the *Kpn*I site of p1079, a *Bgl*II-*Sca*I fragment from a second p1079 plasmid, also treated to create blunt ends, was ligated to the vector to generate the plasmid p1138.

The alcohol dehydrogenase 1 (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 2X35S that was excised with the intron, the *Eco*RV-*Bam*HI fragment from p1163 was ligated into p1138 to form the plasmid p1166. The p1163 *Eco*RV-*Bam*HI fragment consisted of the region downstream of −90 (*Eco*RV) of the CaMV 35S promoter through the 5′ UTR (*Bgl*II) of the CaMV 35S-globin-E9 cassette described by Newman et al. (1993), plus irrelevant polylinker sequences between *Bgl*II and *Bam*HI that were excised in the next cloning step. Finally, the human β-globin-coding region and E9 3′ UTR on a *Bgl*II-*Cla*I fragment from p977 (described by De Rocher et al., 1998) was inserted into *Bgl*II-*Cla*I-digested p1166 behind the 2X35S to generate p1185. The 2X35S-globin-E9 gene cassette was removed from p1185 by *Hin*dIII digestion for insertion into the *Hin*dIII site of the binary vector pBI121 to make p1528.

To construct the globin-*B.t.* chimeric gene, the *Acc*I-*Bam*HI fragment (segment 4) was excised from the *cryIA(c)* *B.t.*-toxin-coding region in p995. After creating blunt 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 (GAGCTCAGATCTTCTAGAGATATCGGATTCATCGAT) containing *Bgl*II, *Xba*I, *Eco*RV, and *Bam*HI restriction sites to generate p1167. After digestion with *Bgl*II and *Bam*HI, the segment 4 DNA fragment was inserted into the unique *Bam*HI site between the globin-coding region and the E9 3′ UTR of p1160, which contained an expression cassette consisting of 2X35S-ADH1 intron 1-globin-coding region-*rbcS-E9* 3′ UTR (De Rocher et al., 1998), to create p1171. A DNA fragment carrying *B.t.*-toxin segment 4 and the E9 3′ UTR was excised from p1171 with *Bam*HI and *Sca*I and substituted for the region of p1185 between *Bam*HI and *Sca*I containing the E9 3′ UTR to create the expression cassette 2X35S-globin-*B.t.*-toxin segment 4-E9 in p1188. The expression cassette was then excised with *Hin*dIII and introduced into the *Hin*dIII site of the binary vector pBI121 to make

plasmid p1194 for standard *A. tumefaciens*-mediated transformation of BY-2 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 LiCl pellet. Twenty micrograms of total RNA or 2 μg of poly(A$^+$) RNA was denatured and separated on 2% (v/v) formaldehyde/1% (w/v) agarose gels in 1× Mops buffer (20 mM Mops, 5 mM sodium acetate, 1 mM EDTA, and 1 μg mL^{-1} ethidium bromide) before capillary transfer to membrane (BioTrace HP, Gelman Sciences, Ann Arbor, MI). Blots were prehybridized and hybridized as described by De Rocher et al. (1998), except that prehybridization was overnight. Radiolabeled DNA probes were synthesized by the random-primer method described by Feinberg and Vogelstein (1983) using restriction fragments separated on low-melting-point agarose gels. Probes corresponding to segments 1, 2, and 3 of the *cryIA(c)* *B.t.*-toxin-coding region (see Fig. 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). Blots were washed for 30 min at 65°C in 2× SSC, 0.1% (w/v) SDS, followed by a wash in 1× SSC, 0.1% (w/v) SDS under the same conditions. Radioactive bands were detected using a phosphor imager (Molecular Dynamics, Sunnyvale, CA).

RNA probes corresponding to the entire E9 3′ UTR or the *Acc*I/*Bam*HI segment of the *B.t.*-toxin-coding region (segment 4) were in vitro transcribed using a riboprobe system (Promega) from the 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 were purified with push columns (Stratagene). Prehybridization of the blots was performed as described above; however, hybridization with the riboprobe was done overnight at 65°C in hybridization solution (described by De Rocher et al., 1998), which was modified by increasing the formaldehyde concentration to 50% (v/v) and decreasing the SSC concentration to 1×. Blots were washed as described above but with an additional wash at 0.2× SSC, 0.1% (w/v) SDS.

Transcripts synthesized in vitro for RNase-H-cleavage experiments were generated from approximately 0.5 μg of p995 that was linearized with *Acc*I. The reaction was assembled at room temperature and consisted of 40 mM Tris-HCl, pH 8.0, 19 mM MgCl$_2$, 5 mM DTT, 2 mM spermidine, 0.01% Triton X-100, 4 mM each ATP, CTP, UTP, and GTP, 16 units of RNAsin (Promega), 1 unit of yeast pyrophosphatase, and 200 units of T7 RNA polymerase (GIBCO-BRL) in a 50-mL volume. After 6 h of incubation at 37°C, 1 μL of RNase-free DNase I (GIBCO-BRL) was added, and the reaction was incubated for another 15 min at 37°C. Sterile distilled water at a volume of 130 μL was added followed by 20 μL of 5 M NH$_2$OAc/100 mM EDTA. The RNA was then phenol/chloroform extracted and pre-

cipitated. The reaction resulted in a 1201-nucleotide in vitro transcript containing segments 1, 2, and 3. Fifty picograms of the transcript was added to 20 μg of total RNA from untransformed BY-2 cells before electrophoresis through an RNA gel.

RT-PCR Analysis

RT-PCR was performed using the 3′ rapid amplification of cDNA ends system (GIBCO-BRL) as recommended by the manufacturer. For first-strand cDNA synthesis, 1 μg of total RNA isolated from stably transformed cell lines was incubated for 10 min at 65°C with 500 nM (final concentration) adaptor primer, which was supplied with the kit. After chilling on ice for 2 min, the mixture was incubated at 42°C for 30 min in the presence of 20 mM Tris-HCl, pH 8.4, 50 mM KCl, 2.5 mM MgCl$_2$, 100 μg mL^{-1} BSA, 10 mM DTT, 500 mM each dATP, dGTP, dCTP, and dTTP, and 200 units of RT (SuperScript, GIBCO-BRL). The RNA in the reaction was then degraded with 4 units of RNase-H at 42°C for 10 min. Two microliters of the 20-μL cDNA synthesis reaction was added to 20 mM Tris-HCl, pH 8.4, 50 mM KCl, 2.5 mM MgCl$_2$, 100 μg mL^{-1} BSA, 200 mM each dATP, dGTP, dCTP, and dTTP, and 0.1 unit μL^{-1} *Taq* DNA polymerase (GIBCO-BRL) in a 50-mL volume for PCR amplification. The gene-specific primers (Macromolecular Structure Facility, Michigan State University) PG-177 (5′-CTCTCAATGGGACGCATTTCTTG-3′), which hybridizes to bases 213 to 235 relative to the *cryIA(c) B.t.*-toxin translation-initiation site, and PG-170 (5′-CTATCAGAAA GTGGTGGCTGGTGTGGCTAATG-3′), which anneals 386 to 417 bases from the globin translation start site, were added to a 200 nM final concentration to amplify the *B.t.*-toxin and globin-*B.t.* chimeric transcripts, respectively.

The reverse primer, PG-192 (5′-GGCCACGCGTCG ACTAGTAC-3′), which anneals to the adapter region of the adaptor primer used to generate the cDNA, was also added to a final concentration of 200 nM. The PCRs of 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 that genomic DNA was not amplified. The amplification protocol was for 5 min at 94°C followed by 30 cycles of 2 min at 94°C, 2 min at 55°C (for PG-177/PG-192), or 2 min at 65°C (for PG-170/PG-192) and 3 min at 72°C. A 15-min incubation at 72°C completed the amplification. The *B.t.*-toxin PCR products were digested with *Xba*I and *Sal*I, whereas the globin-*B.t.* PCR products were digested with *Bam*HI and *Sal*I. All of the PCR products were cloned into p948, the Bluescript II SK(−) vector described above. Four positive clones were identified by gel electrophoresis or by Southern-blot analysis using a DNA probe consisting of the *cryIA(c) B.t.*-toxin-coding region. Cycle sequence analysis of the clones identified the poly(A$^+$) addition sites (DNA Sequencing Facility, Michigan State University).

Oligo-Directed RNase-H-Cleavage Analysis

To remove the poly(A$^+$) tail, cleavage experiments were performed using 1 μg of oligo(dT)$_{12-18}$ and either 20 μg of total RNA or 50 pg of in vitro-transcribed *B.t.*-toxin transcript. The oligo(dT)$_{12-18}$ and RNA mixtures were incubated at 65°C for 5 min and allowed to cool to room temperature for 5 min. For mapping the segment 3 poly(A$^+$) site, approximately 2 μg 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 μg of total RNA isolated from CHX-treated tobacco cells. Hybridization for these gene-specific oligonucleotides was performed in a 400-mL 65°C water bath that was allowed to cool to room temperature. After annealing, RNA samples for both types of reactions were incubated at 37°C in 4 mM Tris-HCl, pH 8.0, 10 mM MgCl$_2$, 20 mM KCl, 1 mM DTT, and 1 unit of RNase-H (GIBCO-BRL) for 30 min with the oligo(dT), and for 1 h for the gene-specific oligonucleotides. The reactions were precipitated and the pellets resuspended in formaldehyde loading buffer.

RNase-H-cleavage experiments were also performed using 3.5 μg of poly(A$^+$) RNA from untreated tobacco cells annealed to 0.3 μg of either PG-229 or PG-234. The hybrids were incubated in 20 mM Tris-HCl, pH 7.5, 10 mM MgCl$_2$, 100 mM KCl, and 5% (w/v) Suc, with 3 units of RNase-H for 60 min at 37°C. All samples were alcohol precipitated and resuspended in formaldehyde loading buffer before electrophoresis through a 2% (v/v) formaldehyde/1.7% (w/v) agarose gel in 1× Mops buffer (described above). The gel was blotted and the membrane probed with the full-length *B.t.*-toxin-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 *cryIA(c)* protoxin was among the first *B.t.*-toxin genes to be isolated from *B.t.* (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 *cryIA(c) B.t.*-toxin transcript does not accumulate to detectable levels in mature tobacco plants (Murray et al., 1991), making it a good candidate for further investigation of the factors limiting *B.t.*-toxin gene expression. Figure 1A shows the derivative of the *cryIA(c) B.t.*-toxin 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 use 3′ truncations that contain only the insecticidal domain. As shown in Figure 1A, transcription of the gene used in this study is controlled by a modified CaMV 35S promoter containing a duplicated enhancer region (2X35S), and the polyadenylation signal is provided by the well-

A

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

B

|	Segment 4	
2 X 35S	Globin	E9
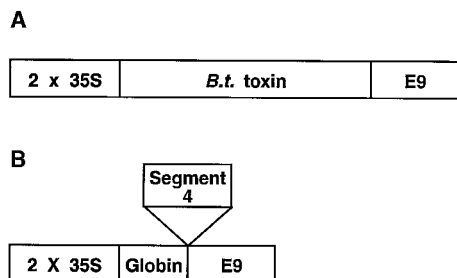
Figure 1. Structure of the genes stably introduced into tobacco cells. A, Portion of the wild-type *B.t.* subsp. *kurstaki cryIA(c)* gene encoding the insecticidal domain (amino acids 9–613; Schnepf and Whiteley 1985) used in this study. Transcription of the chimeric gene was controlled by the CaMV 35S promoter modified by duplicating the upstream enhancer region (2X35S). The E9 3′ UTR provides the elements necessary for polyadenylation. B, Chimeric globin-*B.t.*-toxin gene used to identify the polyadenylation site in segment 4 of the *cryIA(c) B.t.*-toxin-coding region. A 650-bp *Acc*I-*Bam*HI restriction fragment of the *B.t.*-toxin-coding region (segment 4) was inserted between a β-globin reporter gene under the control of the 2X35S and the E9 3′ UTR. Use of any poly(A$^+$) addition sites within the segment 4 insert will result in polyadenylated transcripts that lack the E9 3′ UTR sequences.

characterized pea *rbcS-E9* 3′ UTR (Hunt and MacDonald, 1989; Mogen et al., 1990, 1992; Li and Hunt, 1995).

Tobacco cells stably transformed with this *B.t.*-toxin derivative were analyzed for expression on RNA gel blots. However, as shown in the example in Figure 2A, lane 1, little or no accumulation of the full-length *B.t.*-toxin tran-

script (i.e. transcripts approximately 2300 nucleotides in length terminating in the E9 3′ UTR) could be observed in total RNA samples. The full-length transcript could not be visualized in most of the transformed cell lines that were examined, presumably because the transcript was below the level of detection (data not shown). This is despite the fact that transcription is directed by the 2X35S. 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 *B.t.*-toxin-coding region contains a repressor sequence capable of inhibiting transcription initiation or elongation (De Rocher et al., 1998). The discrepancy between the RNA-gel-blot and the run-on-transcription results suggests that the mechanisms responsible for the low abundance of the *B.t.*-toxin transcript in plant cells are posttranscriptional.

Figure 2A, lane 2, shows the poly(A$^+$) RNA fraction from the same cell line shown in Figure 2A, lane 1. The full-length *B.t.*-toxin transcript can be easily detected in this fraction, indicating that it is polyadenylated, as expected. However, two short transcripts of 900 and 600 nucleotides can also be detected. The abundance of these two short transcripts were significantly increased by treating the cell line with the translation inhibitor CHX (compare Figure 2A, lanes 1 and 3 and lanes 2 and 4). CHX can increase the abundance of unstable transcripts because many of these transcripts are degraded in a translation-dependent manner. Nearly all of the stably transformed cell lines analyzed accumulated the two short transcripts in the presence of
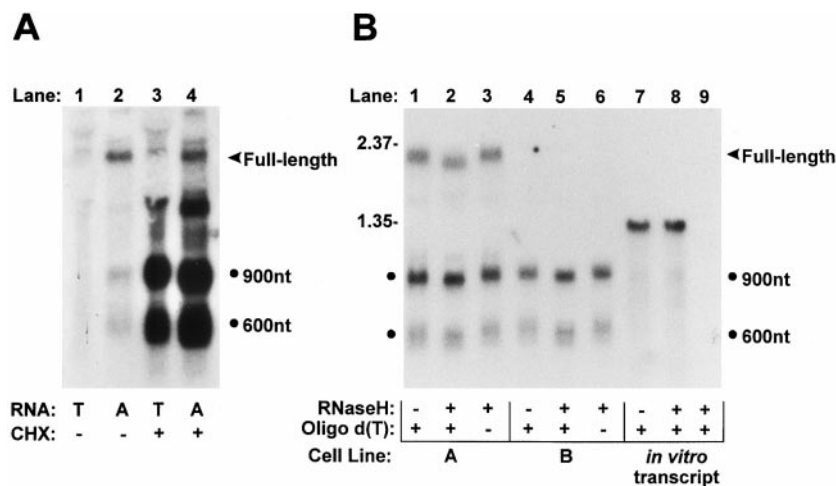
Figure 2. Poor accumulation of *B.t.*-toxin mRNA and the detection of short, polyadenylated transcripts in tobacco cells. A, Total RNA (T) from stably transformed tobacco cells was isolated from a pool of kanamycin-resistant calli growing in liquid culture and electrophoresed next to poly(A$^+$) RNA (A) isolated from the same cells. Cell cultures were either treated (+) or not treated (−) with CHX. The RNA gel blot was probed with a 730-bp DNA fragment corresponding to the 5′ portion of the *B.t.*-toxin-coding region. The autoradiograph was overexposed to show the *B.t.*-toxin transcripts in tobacco cells not treated with CHX. B, Detection of poly(A$^+$) tails on the *B.t.*-toxin transcripts. Total RNA was isolated from two different CHX-treated tobacco cell lines (A and B) expressing the *B.t.*-toxin gene. Both cell lines accumulate the 900- and 600-nucleotide (nt) *B.t.*-toxin transcripts, but only one cell line (A) accumulates the full-length *B.t.*-toxin transcript. RNA from each line was incubated with RNase-H, oligo(dT)$_{12-18}$, or both, and electrophoresed next to a 1227-nucleotide in vitro-transcribed *B.t.*-toxin transcript that was incubated in the presence or absence of RNase-H. Hybridization of oligo(dT)$_{12-18}$ to the poly(A$^+$) tail of a transcript results in cleavage of the poly(A$^+$) tail by RNase-H. As a consequence, deadenylated transcripts will have an increased mobility in an RNA gel. The RNA gel blot was hybridized with a probe for the *B.t.*-toxin-coding region. Lane 9 contains total RNA from untransformed BY-2 cells that was used as a carrier for the in vitro-transcribed transcript.

CHX, even when the full-length *B.t.*-toxin transcript was below the limit of detection (Fig. 2B, lanes 4–6). This is consistent with the rapid degradation of the 900- and 600-nucleotide transcripts in plant cells. The instability of *cryIA(c) B.t.*-toxin transcripts is the focus of the accompanying paper (De Rocher et al., 1998).

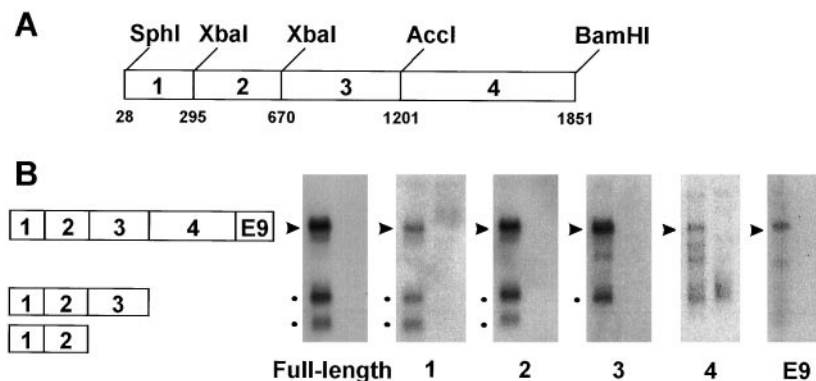## Characterization of the Short *B.t.*-Toxin Transcripts

The abundance of the 900- and 600-nucleotide transcripts in transformed tobacco cells suggests that the mechanism responsible for the generation of these transcripts may play a major role in limiting the accumulation of the full-length *cryIA(c) B.t.*-toxin transcript in tobacco. Because *B.t.*-toxin transcripts have been suggested to be unstable in plant cells (Fischhoff et al., 1987; Vaeck et al., 1987), the 900- and 600-nucleotide 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-nucleotide transcripts were characterized further. Figure 2A, lane 4, shows the presence of the short transcripts and the full-length transcript in the poly(A$^+$) fraction from CHX-treated cells. This indicates that the 900- and 600-nucleotide transcripts are polyadenylated. The 900- and 600-nucleotide transcripts were also present in the poly(A$^+$) RNA fraction from tobacco cells that were not treated with CHX (Fig. 2A, lane 2).

The existence of a poly(A$^+$) tail on the short transcripts was verified using oligo-directed RNase-H cleavage, as shown in Figure 2B. RNase-H cleaved the RNA strand of an RNA/DNA hybrid. Therefore, annealing oligo(dT) 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 incre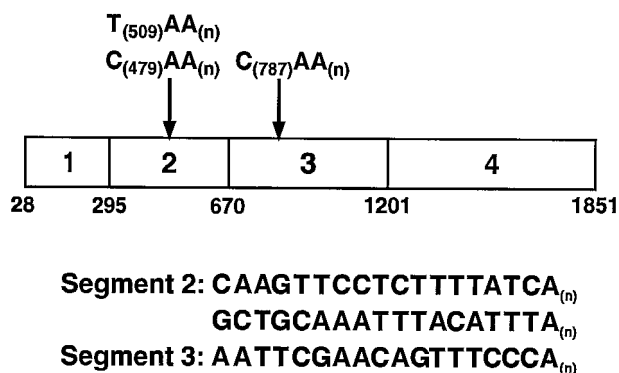ased mobility of the transcripts on an RNA gel blot. As shown in Figure 2B, lanes 2 and 5, the 900- and 600-nucleotide transcripts from two different cell lines decreased in size upon treatment with oligo(dT) and RNase-H. This is consistent with the removal of the poly(A$^+$) tail. The migration of these transcripts was unaltered when RNase-H or oligo(dT) was not present in the reactions (Fig. 2B, lanes 1 and 3 and lanes 4 and 6). The full-length *B.t.*-toxin transcript also had an increased mobility in the presence of RNase-H and oligo(dT), confirming that it is polyadenylated (Fig. 2B, lane 2). An in vitro-synthesized *cryIA(c) B.t.*-toxin transcript corresponding to the first 1201 nucleotides (segments 1, 2, and 3; see below) was also incubated with oligo(dT) in the presence or absence of RNase-H (Fig. 2B, lanes 7 and 8). The migration of this transcript was unaltered in the presence of RNase-H, indicating that oligo(dT) did not anneal nonspecifically to the 900- and 600-nucleotide transcripts or to other sequences within the first 1201 nucleotides of the full-length transcript.

The RNA gel blot in Figure 2A was probed with the first 730 nucleotides of the 1850-nucleotide *B.t.*-toxin-coding region. Detection of the 900- and 600-nucleotide transcripts with this probe and the fact that these transcripts are polyadenylated argues against their 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 which sequences were present in the 900- and 600-nucleotide transcripts, the *B.t.*-toxin-coding region was divided into four segments using convenient restriction sites, as shown in Figure 3A. Each segment was then used to generate probes to hybridize against poly(A$^+$) RNA isolated from CHX-treated cells producing the full-length transcript and both short transcripts (Fig. 3B). All of the probes hybridized to the full-length transcript, as expected. However, the 600-nucleotide transcript hybridized

Figure 3. Characterization of the short *B.t.*-toxin transcripts by hybridization with different segments of the coding region. A, The *cryIA(c) B.t.*-toxin-coding region was divided into four segments using convenient restriction sites. Segment 1 consists of an *Sph*I-*Xba*I restriction fragment from positions 28 to 295; segment 2 is a restriction fragment extending from positions 295 to 670 and is defined by *Xba*I sites at each end; segment 3 is an *Xba*I-*Acc*I restriction fragment from positions 670 to 1201; and segment 4 is the remainder of the coding region from the *Acc*I site at position 1201 to the *Bam*HI site at position 1851. B, An RNA gel blot consisting of alternating lanes of poly(A$^+$) RNA 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 *B.t.*-toxin-coding region (designated below each panel). The arrowheads indicate the full-length transcript, and the dark circles indicate the 900- and 600-nucleotide transcripts. The box diagrams indicate the segments of the *B.t.*-toxin-coding region found in each transcript.

$T_{(509)}AA_{(n)}$

$C_{(479)}AA_{(n)}$ $C_{(787)}AA_{(n)}$

| 1 | 2 | 3 | 4 |

28 295 670 1201 1851

Segment 2: CAAGTTCCTCTTTTATCA$_{(n)}$

GCTGCAAATTTACATTA$_{(n)}$

Segment 3: AATTCGAACAGTTTCCCA$_{(n)}$

Figure 4. Identification of three poly(A$^+$) addition sites within the *B.t.*-toxin-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.

to only the segment 1 and 2 probes and the 900-nucleotide transcript hybridized to the segment 1, 2, and 3 probes. Hybridization was detected in the 900-nucleotide region with the segment 4 probe. However, this was also evident in the untransformed control and, therefore, does not correspond to the 900-nucleotide *B.t.*-toxin transcript. Neither the 600- nor the 900-nucleotide transcript hybridized with the probe spanning the E9 3′ UTR (Fig. 3B, last panel), indicating that the poly(A$^+$) tail is attached directly to the *B.t.*-toxin-coding region. The simplest explanation of these data is that sequences within the *cryIA(c)* *B.t.*-toxin-coding region are recognized as polyadenylation signals. The 600-nucleotide transcript is consistent with polyadenylation within segment 2 of the coding region, and the 900-nucleotide transcript is consistent with polyadenylation in segment 3. Other minor transcripts can be observed hybridizing to some of the *B.t.*-toxin probes (Fig. 2A); however, unlike the 900- and 600-nucleotide transcripts, these transcripts were not consistently found in the stably transformed cell lines. Thus, these transcripts were not pursued further.

Identification of Polyadenylation Sites within the *B.t.*-Toxin-Coding Region

If tobacco uses 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(dT)-adapter primer. Aliquots of the cDNA were used as a template for PCR with primers that hybridize to the 3′ portion of segment 1 and to the adapter region at the 5′ end of the oligo(dT) primer. PCR products were cloned into a Bluescript vector and sequenced as described in "Materials and Methods." As shown in Figure 4, the poly(A$^+$) tail of the 120-bp PCR product mapped to position 787 in segment 3, and the approximately 195-bp PCR products mapped to two sites in segment 2 at positions 479 and 509. These sites are consistent with the sizes of the PCR products and the short transcripts in Figure 2 when assuming a poly(A$^+$) tail of 110 to 120 bases. The finding of two nearby poly(A$^+$) sites

in segment 2 is consistent with the diffuse nature of the 600-nucleotide transcript and the apparent resolution of two bands in some gels (e.g. Fig. 2B).

The difference in size between the 900- and 600-nucleotide transcripts and that expected based on the mapped poly(A$^+$) addition sites is likely due to the length of the poly(A$^+$) tail. To test this possibility and to confirm that the segment 3 polyadenylation site identified by RT-PCR corresponds to the same polyadenylation site used to generate the 900-nucleotide 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 5A, an oligonucleotide that hybridizes starting 65 bases upstream of the segment 3 polyadenylation site directs cleavage of the full-length and 900-nucleotide transcripts in the presence of RNase-H. The mobility of the 600-nucleotide transcript, which lacks sequences complementary to the oligonucleotide, was unaltered in the presence of RNase-H, as expected.

The 180-base decrease in the size of the 900-nucleotide transcript to approximately 720 nucleotides in the presence of RNase-H (Fig. 5A) indicates that 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- to 120-base poly(A$^+$) tail predicted from the discrep-
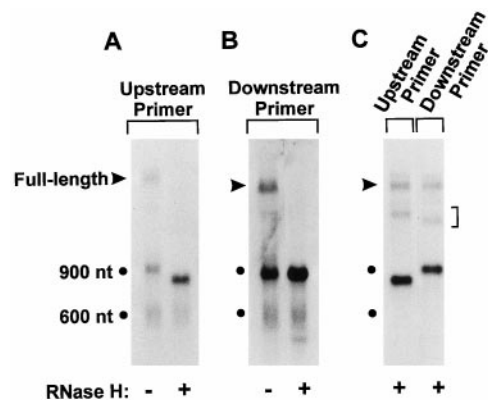
Figure 5. The mapped poly(A$^+$) addition site in segment 3 corresponds to the polyadenylation site of the 900-nucleotide (nt) in vivo transcript. Total RNA from CHX-treated tobacco cells expressing the *B.t.*-toxin gene was incubated with oligonucleotides that hybridize either upstream (A) or downstream (B) of the segment 3 polyadenylation site identified by RT-PCR. Poly(A$^+$) RNA from transgenic tobacco cells not treated with CHX was also incubated with the upstream and downstream oligonucleotides (C). Incubations were performed in the presence (+) or absence (−) of RNase-H. If sequences complementary to the oligonucleotides are present in the 900-nucleotide transcript, RNase-H will cleave the RNA strand of the RNA/DNA duplex, resulting in a band shift on the RNA gel blots. The bracketed bands, which may correspond to the 3′ ends of the full-length transcript, were detected only when poly(A$^+$) RNA was digested with RNase-H.

ancy between the position of the segment 3 polyadenylation site and the in vivo size of the transcript.

Similar RNase-H-cleavage experiments were carried out using an oligonucleotide that hybridizes starting approximately 10 bases downstream of the polyadenylation site mapped in segment 3 (Fig. 5B). This oligonucleotide should not anneal to the 900-nucleotide transcript unless it actually extends beyond the mapped poly(A^+) site. Figure 5B shows that the size of the 900-nucleotide transcript was not altered after incubation with the oligonucleotide and RNase-H, demonstrating that sequences immediately downstream of the segment 3 polyadenylation site are not present in the 900-nucleotide 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 (see Fig. 5C). These data indicate that the segment 3 polyadenylation site identified by RT-PCR is the same site used to generate the 900-nucleotide transcript. More importantly, they also indicate that the same polyadenylation site is used in both CHX-treated and untreated transformed tobacco cells. RNase-H-cleavage experiments were also performed to confirm the segment 2 polyadenylation sites. However, the diffuse nature of the 600-nucleotide transcript made it difficult to assess precisely the shift in bands (data not shown). Yet, an oligonucleotide complementary to a region upstream of the cleavage site was able to shift the 600-nucleotide transcript to a smaller size, whereas an oligonucleotide hybridizing downstream of the cleavage sites 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. In addition, plants do not require a *cis*-regulatory element downstream of the cleavage site, as do animal systems. However, as shown in Figure 6A, plant polyadenylation signals do require two *cis*-regulatory elements: the FUE and the NUE (for review, see 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 among known FUEs is the presence of multiple U/G-rich regions (for review, see Hunt, 1994; Wu et al., 1995; Rothnie, 1996). Several U/G-rich stretches can be found in the *B.t.*-toxin-coding region upstream of the segment 2 and segment 3 poly(A^+) addition sites in positions that correspond to a putative FUE (Fig. 6B).

The NUE is an A/U-rich element typically found 10 to 30 bases upstream of the cleavage site. These elements are essential for polyadenylation and control poly(A^+) addition at specific cleavage sites. Therefore, a plant transcript
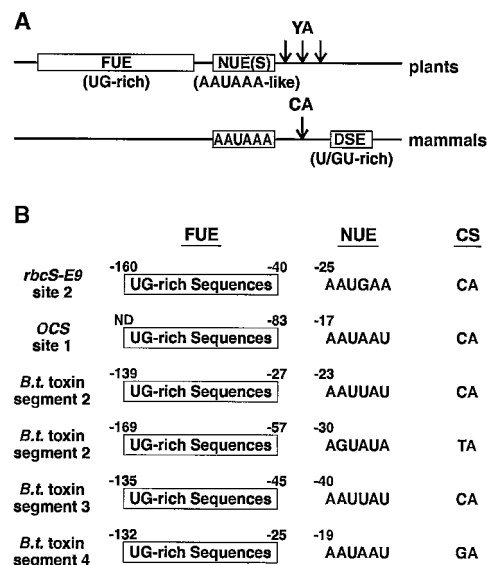
Figure 6. The *B.t.*-toxin-coding region contains elements characteristic of plant-polyadenylation signals. A, Schematic representation comparing the structure of typical plant and mammalian polyadenylation signals. Sequence motifs characteristic of the plant FUE and NUE, as well as the mammalian downstream element (DSE), are indicated. The poly(A^+) addition sites are represented by arrows. Plants can use multiple poly(A^+) addition sites downstream of specific NUEs within a transcript. The cleavage of a plant transcript usually occurs at a Py/A dinucleotide (YA). Mammalian transcripts are usually cleaved at a single site corresponding to a C/A dinucleotide (CA). B, Identification of elements characteristic of plant polyadenylation signals upstream of the poly(A^+) addition sites in the *B.t.*-toxin-coding region. The putative plant polyadenylation signals in segments 2, 3, and 4 of the *B.t.*-toxin-coding region are compared with the most commonly used polyadenylation sites in the *rbcS-E9* and octopine synthase (*OCS*) genes. The positions of the FUE and NUE relative to the cleavage site (CS) are indicated.

with multiple polyadenylation sites will have a NUE corresponding to each site. NUEs can contain the mammalian canonical AAUAAA sequence, as in the CaMV polyadenylation signal (Sanfacon et al., 1991; Rothnie et al., 1994), but more often an AAUAAA-like sequence is present in which one or two of the bases do not match (for review, see Hunt, 1994; Wu et al., 1995; Rothnie, 1996). Upstream of the three 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 with other known NUEs revealed sequence similarity (Fig. 6B).

Cleavage of the *B.t.*-toxin transcript in both segments 2 and 3 occurred at Py/A dinucleotides (Fig. 6B). This is consistent with other known plant poly(A^+) addition sites. Again, no strict consensus sequence is known that defines the cleavage site in plants. However, cleavage typically occurs at a Py/A dinucleotide in plant transcripts (for review, see Hunt, 1994; Wu et al., 1995; Rothnie, 1996).

Further sequence analysis of the *B.t.*-toxin gene revealed the presence of other possible plant polyadenylation signals in the coding region. In particular, segment 4 (Fig. 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 the template. However, these transcripts may be very unstable or 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 *B.t.*-toxin-coding region contains a functional plant polyadenylation signal, a chimeric gene was constructed consisting of the segment inserted between a 2X35S-driven β-globin reporter gene and the E9 3′ UTR (Fig. 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 7A shows that stably transformed tobacco cells expressing the globin-*B.t.* gene accumulated only a small amount of transcript at a position consistent
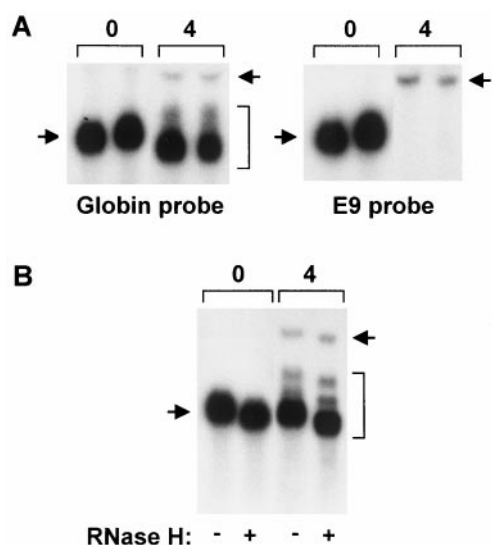
Figure 7. Identification of a fourth polyadenylation site in the *B.t.*-toxin-coding region. A, Truncated transcripts are produced in tobacco cells expressing a chimeric globin-*B.t.*-toxin gene containing nucleotides 1201 to 1851 of the *B.t.*-toxin-coding region. An RNA gel blot of total RNA extracted from two pools of 100 tobacco calli stably transformed with the globin-segment 4 chimeric gene (4) (p1194) described in "Materials and Methods" 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 (p1528) described in "Materials and Methods," but lacking the segment 4 insert, was used as a control (0). The arrows indicate the positions of the transcripts terminating in the E9 3′ UTR for the control and globin-*B.t.*-toxin genes. The bracket marks the position of the globin-*B.t.*-toxin transcripts polyadenylated in the segment 4 insert. The autoradiographs were overexposed to show the globin-*B.t.*-toxin transcripts terminating in the E9 3′ UTR. 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(dT)$_{12-18}$ 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 positions of the transcripts terminating in the E9 3′ UTR for the control and globin-*B.t.*-toxin transcripts. The bracket indicates the position of the globin-*B.t.*-toxin transcripts polyadenylated in the segment 4 insert.

with termination in the E9 region. Most of the globin-*B.t.* transcripts in these cells accumulated as discrete bands at sizes more consistent with termination in segment 4. Hybridization with the E9 3′ UTR showed that these abundant transcripts lack the E9 region (Fig. 7A). Similar 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(dT). As shown in Figure 7B, the short transcripts decreased in size in the presence of RNase-H, indicating that 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 13 bases downstream of a putative NUE in segment 4 (Fig. 6B). Taken together, these data show that segment 4 of the *B.t.*-toxin-coding region does contain sequences that function as polyadenylation signals in plants. How much polyadenylation at this site contributes to the poor accumulation of the full-length *B.t.*-toxin mRNA is not known. We cannot rule out the possibility that our globin-*B.t.* construct activates a cryptic polyadenylation site within segment 4 (Luehrsen and Walbot, 1994). However, when a chimeric *B.t.*-toxin gene was created by fusing synthetic plant-like versions of segments 1, 2, and 3 to a wild-type version of segment 4, most of the resulting mRNA was polyadenylated in segment 4, rather than at the E9 poly(A⁺) site located downstream (S.H. Diehn, W.-L. Chiu, E.J. De Rocher, and P.J. Green, unpublished results). This suggests that the segment 4 poly(A⁺) site does function within a *B.t.*-toxin gene context.

DISCUSSION

The goal of this study and the one that follows (De Rocher et al., 1998) 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 low accumulation of this transcript may make it easier to resynthesize novel *B.t.*-toxin genes, but more importantly, it may provide an understanding of why the transcripts of some foreign genes fail to accumulate in plants. In this report we have demonstrated that the *cryIA(c) B.t.*-toxin-coding region contains multiple sequence elements that are recognized by plant cells as polyadenylation signals. We suggest that use of these polyadenylation signals appears to be at least partially responsible for the low accumulation of the *cryIA(c) B.t.*-toxin 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 observed previously that tobacco plants expressing a *cryIA(c)* protoxin gene or a 3′-truncated version produced a polyadenylated 1.7-kb transcript. The size of this transcript was shorter than expected for either gene, prompting the suggestion that the *B.t.*-toxin-coding region contains plant polyadenylation signals (Adang et al., 1987). However, the transcript disappeared as the plants matured (Murray et al., 1991); therefore, the mechanism responsible for the production of the 1.7-kb transcript could not be

determined, and the disappearance of the transcript during development was not 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 accumulated polyadenylated transcripts about 900 and 600 nucleotides in length. Hybridization, RT-PCR, and RNase-H-mapping experiments all confirmed that these short transcripts were the result of polyadenylation at two nearby sites in segment 2 and at another site in segment 3. Although the short transcripts could not be detected in total RNA preparations, the abundance of these transcripts could be increased by treating the cells with CHX to a point at which they were the most abundant *B.t.*-toxin transcripts in both the total and poly(A$^+$) RNA fractions. In addition, these transcripts accumulated to detectable levels in the poly(A$^+$) RNA fractions of untreated tobacco cells. As discussed in the accompanying paper (De Rocher et al., 1998), the full-length and short *B.t.*-toxin transcripts are rapidly degraded in tobacco. Although they are unstable, we suggest that the short *B.t.*-toxin transcripts are produced in significant amounts and thereby contribute to the low accumulation of the full-length mRNA in plants.

The high A/T content of *B.t.*-toxin genes raises the possibility that regions of the *cryIA(c) B.t.*-toxin 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 that the 900- and 600-nucleotide transcripts are not a result of splicing. The hybridization data, the RT-PCR analysis, and the RNase-H experiments were consistent with the conclusion that these transcripts are a result of polyadenylation in the *B.t.*-toxin-coding region. This is not to suggest that splicing of the *cryIA(c) B.t.*-toxin transcript does not occur in tobacco cells. Splicing could account for some of the minor transcripts hybridizing to the various *B.t.*-toxin probes observed on our RNA gel blots. However, most of these transcripts were not reproducible in our transformed tobacco cell lines and, therefore, were not pursued further.

The presence of plant poly(A$^+$) addition sites within the *cryIA(c) B.t.*-toxin-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 the *cryI* class, one of six classes of *B.t.*-toxin genes (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 that are active against the insect order Lepidoptera. A subclass of the *cryI* genes, the *cryIA* genes, contains members designated *cryIA(a)*, *cryIA(b)*, *cryIA(c)*, and *cryIA(d)*, which are more than 80% identical at the nucleotide level. Alignment of the *cryIA(c)-* and *cryIA(b)*-coding regions shows that a region extending 230 bp upstream from the second segment 2 polyadenylation site (position 509) is identical between the two genes (data not shown). This

230-bp region should be of sufficient length to contain the elements necessary for polyadenylation in plants. A similar alignment with the *cryIA(a)* gene shows that the same segment 2 poly(A$^+$) signals are 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 signals (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. Other *B.t.*-toxin genes, such as *cryIE(a)*, *cryIF*, *cryID*, and *Prt A*, share lower levels of nucleotide identity in the region of the poly(A$^+$) signals (e.g. 57%–87% for the segment 2 poly[A$^+$] signals), 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.

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 to our knowledge, 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. 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 transposase 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. Like *B.t.*-toxin genes, these genes have a high A/U bias. 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.

LITERATURE CITED

Adang MJ, Firoozabady E, Klein J, DeBoer D, Sekar V, Kemp JD, Murray E, Rocheleau TA, Rashka K, Staffeld G, and others (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, New York, pp 345–353

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

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* δ-endotoxin expressed in transgenic *Nicotiana tabacum* provides resistance to lepidopteran insects. Plant Physiol **85:** 1103–1109

Bigelow CC, Channon M (1982) *In* GD Fasman, ed. CRC Handbook of Biochemistry and Molecular Biology, Ed 3: Proteins, Vol 1. CRC Press, Boca Raton, FL, pp 209–224

Cheng J, Bolyard MG, Saxena RC, Sticklen MB (1992) Production of insect resistant potato by genetic transformation with a δ-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, and others (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 (1998) Direct evidence for rapid degradation of *Bacillus thuringiensis* toxin mRNA as a cause of poor expression in plants. Plant Physiol **117:** 1445–1461

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, Germany

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, and others (1987) Insect tolerant transgenic tomato plants. Bio/Technology **5:** 807–813

Goff SA, Cone KC, Fromm ME (1991) Identification of functional domains in the maize transcriptional activator C1: comparison of wild-type and dominant inhibitor proteins. Genes Dev **5:** 298–309

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*. Microbiol Rev **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) β-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

Luehrsen KR, Walbot V (1994) Intron creation and polyadenylation in maize are directed by AU-rich RNA. Genes Dev **8:** 1117–1130

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 (1996) Plant mRNA 3′-end formation. Plant Mol Biol **32:** 43–61

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

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. PhD 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