The Plant Cell, Vol. 2, 1261–1272, December 1990 © 1990 American Society of Plant Physiologists

Upstream Sequences Other than AAUAAA Are Required for Efficient Messenger RNA 3′-End Formation in Plants

Bradley D. Mogen, Margaret H. MacDonald, Robert Graybosch,[1] and Arthur G. Hunt[2]

Plant Physiology/Biochemistry/Molecular Biology Program, Department of Agronomy, University of Kentucky, Lexington, Kentucky 40546–0091

We have characterized the upstream nucleotide sequences involved in mRNA 3′-end formation in the 3′ regions of the cauliflower mosaic virus (CaMV) 19S/35S transcription unit and a pea gene encoding ribulose-1,5-bisphosphate carboxylase small subunit (*rbcS*). Sequences between 57 bases and 181 bases upstream from the CaMV polyadenylation site were required for efficient polyadenylation at this site. In addition, an AAUAAA sequence located 13 bases to 18 bases upstream from this site was also important for efficient mRNA 3′-end formation. An element located between 60 bases and 137 bases upstream from the poly(A) addition sites in a pea *rbcS* gene was needed for functioning of these sites. The CaMV −181/−57 and *rbcS* −137/−60 elements were different in location and sequence composition from upstream sequences needed for polyadenylation in mammalian genes, but resembled the signals that direct mRNA 3′-end formation in yeast. However, the role of the AAUAAA motif in 3′-end formation in the CaMV 3′ region was reminiscent of mRNA polyadenylation in animals. We suggest that multiple elements are involved in mRNA 3′-end formation in plants, and that interactions of different components of the plant polyadenylation apparatus with their respective sequence elements and with each other are needed for efficient mRNA 3′-end formation.

INTRODUCTION

The polyadenylation of mRNAs is an essential step in the course of the expression of genes in eukaryotes. In animal cells, RNAs are polyadenylated in a process that is distinct from transcription termination; nascent RNAs are cleaved and a tract of polyadenylate is added to the 3′ end of the cleaved RNA (Birnsteil et al., 1985; Platt, 1986). This process appears to be limited to RNA polymerase II transcripts (Sisodia et al., 1987). These events occur shortly after the site of polyadenylation is transcribed and take place in the nucleus. Polyadenylation is directed by two RNA sequence signals: the sequence AAUAAA, located between 15 nucleotides (nt) and 40 nt upstream from polyadenylation sites, and less well characterized sequences located downstream from such sites (Birnsteil et al., 1985; Platt, 1986). In vitro, polyadenylation requires a poly(A) polymerase and several other chromatographically distinguishable components (Christofori and Keller, 1988, 1989; Gilmartin et al., 1988; McDevitt et al., 1988; Takagaki et al., 1988, 1989). These other factors confer AAUAAA specificity upon the poly(A) polymerase and carry out the endonucleolytic cleavage needed for proper polyadenylation. There are indications that polyadenylation is necessary for termination of RNA polymerase II transcription (Whitelaw and Proudfoot, 1986; Logan et al., 1987; Connelly and Manley, 1988). Thus, although polyadenylation is an RNA processing event and not a transcription termination event, it may be a necessary requisite for transcription termination by polymerase II.

In contrast to the situation in animal cells, polyadenylation and transcription termination in yeast cells seem to be more tightly coupled (Birnsteil et al., 1985; Platt, 1986; Russo and Sherman, 1989). In addition, the 3′ ends of several yeast mRNAs can be generated in vitro by processing and modification of a precursor RNA (Butler and Platt, 1988; Butler et al., 1990). The RNA sequences that direct polyadenylation in yeast differ in both composition and position from those that have been defined in animals. In particular, there does not seem to be a strict consensus sequence analogous to the AAUAAA element in animal genes. In at least two yeast genes, the functional elements involved in mRNA 3′-end formation share spatial features with those that direct rho-dependent transcription termination in *Escherichia coli* (Osborne and Guarente, 1989), suggesting mechanistic analogies between yeast and bacteria.

The process of mRNA 3′-end formation in plants appears to be somewhat different from the same process in other higher eukaryotes. For example, the animal poly(A)

[1] Current address: USDA/ARS, Department of Agronomy, Keim Hall, East Campus, University of Nebraska, Lincoln, NE 68583.
[2] To whom correspondence should be addressed.

signal AAUAAA does not occur in a number of plant genes; related sequences are seen in many plant genes but at widely varying distances from sites of polyadenylation (Messing et al., 1983; Hunt et al., 1987; Joshi, 1987). RNAs arising from single plant transcription units generally are heterogeneous at their 3′ ends because of the occurrence of multiple sites of polyadenylation (Dean et al., 1986). These and other studies (Hunt and MacDonald, 1989) suggest that mRNA 3′-end formation in plants may be more analogous to the same process in yeast than that in mammals. This suggestion would imply a role for upstream sequences other than AAUAAA in 3′-end formation in plants. Here, we present studies that indicate that upstream sequences other than AAUAAA are indeed needed for efficient mRNA 3′-end formation in plants. Moreover, we show that the canonical AAUAAA motif present in the cauliflower mosaic virus (CaMV) 3′ region is required for efficient mRNA 3′-end formation in this region. We suggest that mRNA 3′-end formation in plants requires the interaction of at least two components that recognize distinct sequence elements upstream from polyadenylation sites.

RESULTS

Assembly of Chimeric Genes Carrying Different Portions of the 3′ Regions of Two Plant Genes and Description of Experimental Strategy

The region of the CaMV genome studied here is shown in Figure 1A. This region encompasses nucleotides 7226 through 7983 of the CaMV genome (Guilley et al., 1982). In this region are the 35S promoter, the 181-nt terminal redundancy that is characteristic of the 35S RNA, the polyadenylation site for the 19S and 35S transcripts, and 369 nt of sequence downstream from this site. This region thus carries bases −389 to +369 with respect to the polyadenylation site.

The region of the pea *rbcS*-E9 gene that is the subject of this study is illustrated in Figure 1B. This region directs polyadenylation at a series of sites [noted in Figure 1B as normal poly(A) sites] either as part of the native *rbcS*-E9 gene or in the context of chimeric genes (Coruzzi et al., 1984; Nagy et al., 1985; Hunt, 1988).

To localize the sequence elements in the CaMV and *rbcS*-E9 3′ regions responsible for efficient mRNA 3′-end formation, the wild-type 3′ regions and a number of deletions derived by nuclease Bal31, restriction endonuclease digestion, or oligonucleotide-directed mutagenesis were assembled into chloramphenicol acetyltransferase (*cat*)-3′-region chimeric genes in a Ti plasmid-associated binary plasmid and introduced into *Agrobacterium tumefaciens*, and the engineered *Agrobacterium* strains were used to transform tobacco. In addition, a mutant in which 3 nt of

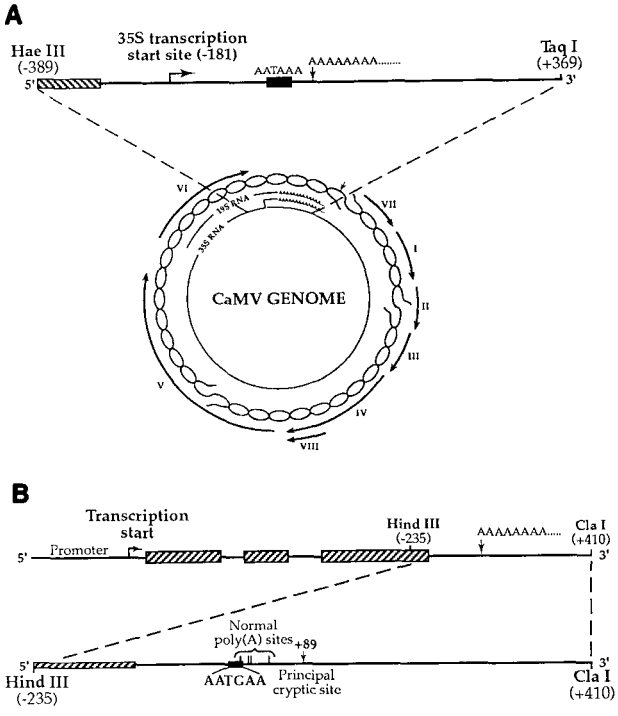

Figure 1. Structures of the 3′ Regions Characterized in This Study.

(A) The region of the CaMV genome analyzed in this study. The positions of the various CaMV genes, the gaps, and the large intergenic region are represented on the circular CaMV genome. Above this is a diagram of that part of gene VI and the large intergenic region that is present in clone CaMV 32-1, the wild-type CaMV poly(A) signal in our studies. Shown in this diagram are the restriction enzyme sites used to subclone this region, the extent of the gene VI coding region present in this construction (▧), and the positions of the first base of the 35S RNA (35S transcription start site), AATAAA motif, and polyadenylation site (arrow below AAAAAAAA. . .). The numbers represent the positions relative to the polyadenylation site; negative numbers denote upstream sequences, positive numbers denote downstream sequences.

(B) The region of the pea *rbcS*-E9 gene analyzed in this study. The top line represents the complete *rbcS*-E9 gene (adapted from Coruzzi et al., 1984) and shows the relative location and restriction enzyme sites used to subclone the polyadenylation signal. The bottom line shows the extent of *rbcS* coding region present in the IDK8 (▨), the positions of the four polyadenylation sites normally seen in RNAs from this gene [vertical lines designated normal poly(A) sites], the location of the principal cryptic poly(A) site located downstream from the normal sites (↓), and the relative position of a potential AAUAAA-like motif located upstream from the normal sites (small box labeled AATGAA). The numbers represent the positions relative to the polyadenylation site; negative numbers denote upstream sequences, positive numbers denote downstream sequences.

the AATAAA motif in the CaMV 3′ region were changed to yield the sequence TAGAAT was assembled into the same expression system. The general structure of the micro-T-DNAs introduced into plants in this study is shown in Figure 2. These T-DNAs carry the *cat*-3′-region chimeric genes and a gene that confers kanamycin resistance upon transformed cells and plants, all flanked by T-DNA borders. Two different promoters were used to drive expression of the *cat*-3′-region genes in our studies. One was the CaMV 35S promoter and the other a derivative of the 35S promoter in which bases −416 to −90 with respect to the transcription initiation site were duplicated (noted here as the 35S^2 promoter). The orientation of these genes is such that the 3′ ends of each are adjacent to each other so that transcription of each gene begins near the respective border. This orientation places the 3′ region of interest some 2 kb from any possible host-derived sequence in transformed plants, thus minimizing potential interference of host sequences with 3′ end choice and/or efficiency in our studies. Transformed plant lines that were kanamycin resistant and carried high levels of chloramphenicol acetyltransferase activity, and thus had no gross rearrangements in the micro-T-DNA, were selected for further analyses.

Preliminary studies with plant cell lines carrying various mutants indicated that there were no appreciable differences in chloramphenicol acetyltransferase activity between any of these lines (data not shown), indicating that chloramphenicol acetyltransferase activity is a poor indicator of poly(A) site function. This was consistent with results reported by Hernandez et al. (1989) and our own earlier studies that showed that a chimeric 35S-*cat* gene with no functional polyadenylation signal, when introduced into tobacco in the context of the binary plasmid p3-1, was still able to produce stable, translatable *cat*-containing RNAs (Hunt and MacDonald, 1989; R. Graybosch and A.G. Hunt, unpublished observations). In addition, preliminary experiments revealed no consistent differences in the total levels of *cat*-containing RNAs in plants carrying different constructions (B.D. Mogen, M.H. MacDonald, R. Graybosch, and A.G. Hunt, unpublished observations). These observations indicated that poly(A) site choice and efficiency could not be reliably judged by measurements of chloramphenicol acetyltransferase activity or of comparisons of levels of *cat*-containing RNAs in different transformed plant lines. Also, comparisons of RNA gel blot and S1 nuclease analyses showed the latter procedure to be more suitable than the former for our studies because this technique permitted both the evaluation of each of the sites present in one of the 3′ regions studied here (the *rbcS* 3′ region) and the detection of RNAs with 3′ ends beyond the region of interest (see below).

For these reasons, we adopted a strategy for the assessment of poly(A) site choice and efficiency that did not rely upon quantitative comparisons of RNA abundance in different transgenic plants lines. RNA from plant lines carrying each construction was analyzed by S1 nuclease protection using probes designed to permit the determination of 3′ ends of RNAs that ended in the region of interest and the detection of RNAs with 3′ ends located in downstream vector sequences. Our expectation was that fully functional and efficient poly(A) sites would yield only RNAs with 3′ ends at previously described, expected sites (termed here as normal sites), whereas poly(A) sites with

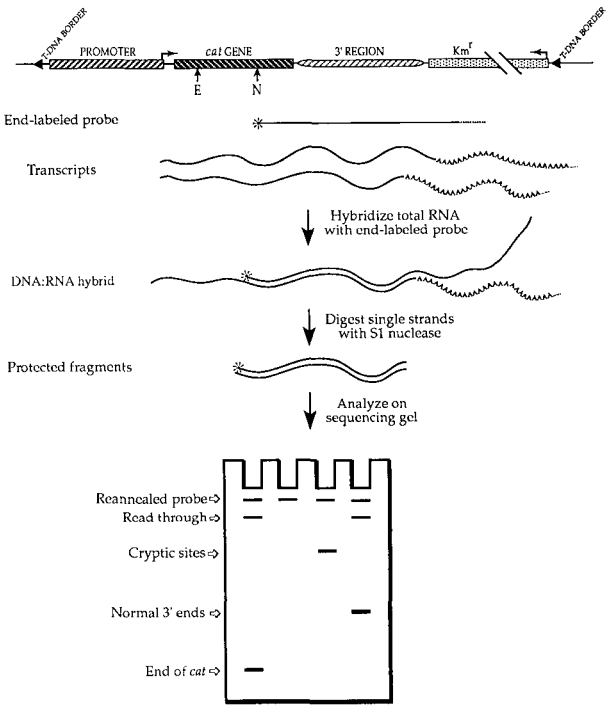

Figure 2. Schematic Depiction of the S1 Nuclease Protection Analyses Described in Methods and in Figures 3, 4, and 5.

The top line shows the basic structure of the micro-T-DNAs that were introduced into plants, with emphasis placed on the features of the *cat*-3′-region gene used to assess poly(A) site function. The arrows show the directions of transcription of the poly(A) site test gene and the kanamycin-resistance (Km^r) gene. S1 probes are prepared from the construction being analyzed, meaning that each mutant is analyzed with a different probe. Probes are prepared by labeling at either the EcoRI or NcoI sites in the *cat* gene (noted as E or N, respectively, below the structure of the chimeric gene). After hybridizing RNA with the denatured probe, S1 nuclease protection, and electrophoretic analysis, protected fragments (if any) are identified and characterized by comparison with DNA size standards. Shown are the positions of hypothetical protected fragments that represent the end of the *cat* gene, mRNA 3′ ends mapping to normal (normal 3′ ends) and cryptic sites, a band corresponding to RNAs that share homology with the probe through the entire extent of the chimeric gene (read through), and reannealed probe (which is larger than the read-through band because of the presence of 197 bp of pUC sequence in the probe that is absent from the gene introduced into plants).

less-than-optimal properties would yield RNAs with a significant portion ending downstream from the expected sites (read-through RNAs). These read-through RNAs could have 3′ ends at cryptic sites in the regions of interest (as is seen in the two regions here), or in the kanamycin-resistance gene that flanks the poly(A) test genes (see Figure 2). In the latter instance, the RNAs will protect all but 197 nt of the S1 probes. Mutations that have affected the efficiency of utilization of a polyadenylation site would thus be identified by assessing the amount of read-through RNAs produced by the respective constructions. Those constructions that yielded no read-through RNAs would define a normally functional site, whereas those constructions that gave rise to appreciable levels of read-through transcripts would define a mutant site. Such a strategy has been used in several characterizations of mammalian, yeast, and plant polyadenylation signals (see, as examples, McDevitt et al., 1984; Gil and Proudfoot, 1987; Hunt and MacDonald, 1989; Osborne and Guarente, 1989; Sanfaçon and Hohn, 1990). We would note, however, that the ratio of read-through to normal 3′ ends would not be a true quantitative estimate of poly(A) site efficiency because this parameter will vary significantly with the type of downstream sequence used to recover read-through transcripts. In other words, a ratio of 1:1 would not define a mutant poly(A) signal that is 50% as active as the wild-type signal but only a mutant signal that is now utilized as frequently as those downstream sites used to recover read-through RNAs (the normal site being utilized much more frequently than the downstream alternate sites in the wild type).

For these analyses, total [and in some cases poly(A)-enriched or poly(A)-depleted] RNA from pooled populations of the various transformed plant lines (at least six individual independent transformants for each construction) were prepared and the RNA 3′ ends of *cat*-containing transcripts determined by S1 nuclease protection analysis. RNAs from pooled populations were prepared to minimize sample-to-sample variations in *cat*-containing RNA abundance. These strategies and analyses have been described in detail elsewhere (Hunt, 1988; Hunt and MacDonald, 1989) and are represented in Figure 2.

Analysis of CaMV Mutants

When the wild-type CaMV 3′ region (defined here as bases −389 to +369, e.g., clone CaMV 32-1) was assessed for ability to direct mRNA 3′-end formation in our system, the only RNA species seen had 3′ ends identical to those reported for the 19S and 35S RNAs in CaMV-infected turnip (Guilley et al., 1982), as indicated by CaMV 32-1 in Figure 3. Therefore, the CaMV polyadenylation signal appears to function in tobacco as it does in virus-infected turnip. A mutant (CaMV-1) in which the AAUAAA motif present at bases −13 to −18 was replaced with the sequence UAGAAU was also tested in the above manner (CaMV-1 in Figure 3). CaMV-1 was able to direct mRNA 3′-end formation at the normal CaMV polyadenylation site, but also yielded RNAs with 3′ ends at a cryptic polyadenylation site in the CaMV genome (roughly 300 nt downstream from the normal CaMV site) and at uncharacterized sites in flanking vector sequences. (These RNAs yielded the S1 nuclease-resistant fragments that end at the end

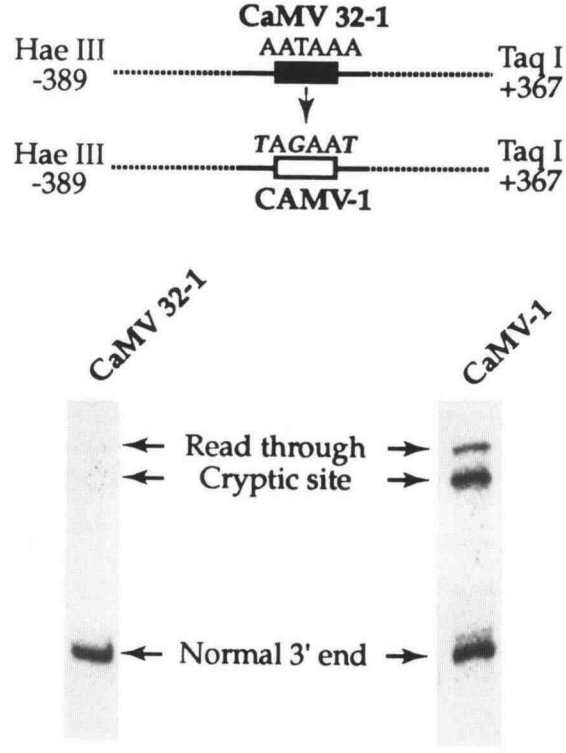

Figure 3. Alteration of an AAUAAA Motif Decreases the Efficiency of Polyadenylation in the CaMV 3′ Region.

The profiles of 3′ ends directed by a mutant carrying nucleotides −389 to +369 (CaMV 32-1) were compared with those directed by a mutant (CaMV-1) carrying the same region, but with the sequence TAGAAT at positions −13 to −18 instead of the AATAAA sequence that is present in CaMV 32-1. Twenty micrograms of total RNA from plants carrying the CaMV 32-1 or the CaMV-1 construction was annealed with end-labeled probe prepared from the corresponding mutant and processed as described in Methods. The positions of the normal polyadenylation site (normal 3′ end), an uncharacterized cryptic polyadenylation site located at roughly +300 (cryptic site), and the end of homology between the gene introduced into plants and the S1 probe (read through) are noted with arrows. (No detectable utilization of the cryptic site or read-through RNAs could be detected in the CaMV 32-1 sample; the expected positions of the corresponding protected fragments are shown for comparison with the CaMV-1 mutant.) The structures of the CaMV 3′ regions present in the chimeric genes are shown above the autoradiograms.

of homology of the probes used here and the gene introduced into plants.) With the CaMV-1 mutant, fewer than 50% of the *cat*-containing RNAs ended at the normal CaMV polyadenylation site, whereas virtually all of the *cat*-containing RNAs from the CaMV 32-1 ended at the CaMV polyadenylation site. The alteration of the AAUAAA motif did not affect the general stability of the *cat*-CaMV 3′-region RNAs because the total steady-state levels of the *cat*-containing RNAs from the two different mutants were similar (data not shown). Rather, removal of this region affected the efficiency of mRNA 3′-end formation at the normal CaMV polyadenylation site; thus, with the CaMV-1 mutant, a significant portion of the transcripts arising from the test gene did not end at the normal site but, instead, at sites downstream from the normal site. This is best explained by reasoning that read-through RNAs are either more abundant or more efficiently processed with respect to normal RNAs in the CaMV-1 mutant; in either case, this suggests a decreased efficiency of 3′-end formation at the normal site in this mutant. Therefore, the AAUAAA motif at nucleotides −13 to −18 is needed for efficient polyadenylation at the normal site.

A construction that carried bases −181 to +369 (CaMV 35STS) yielded only RNAs with 3′ ends at the normal CaMV polyadenylation site, shown as CaMV 35STS in Figure 4. This indicated that sequences upstream from the transcription initiation site in the 35S transcription unit (nucleotide −181, with respect to the polyadenylation site) are not needed for efficient use of the CaMV polyadenylation signal. A mutant that carried bases −57 to +369 (CaMV −389/57) also was able to direct mRNA 3′-end formation at the normal site (CaMV −389/57 in Figure 4). However, a significant population of read-through RNAs was also seen with the latter mutant. For the reasons stated above, we concluded that sequences between 57 nt and 181 nt upstream from the CaMV polyadenylation site are needed for efficient mRNA 3′-end formation at this site.

Analysis of *rbcS* Mutants

To determine whether the requirement for upstream sequences relatively far from polyadenylation sites is a common feature of plant polyadenylation signals, we extended previous studies of the pea *rbcS*-E9 polyadenylation signal (Hunt and MacDonald, 1989). We have previously reported that sequences between −6 and −137 were needed for polyadenylation in the *rbcS*-E9 3′ region (Hunt and MacDonald, 1989). To dissect this region further, a mutant carrying bases −60 to +410 (Q2) was compared with a wild-type *rbcS*-E9 3′ region (IDK8) and two other mutants that have been described before (C2-3 and C6; Hunt and MacDonald, 1989). In contrast to IDK8 and the C2-3 mutant shown in Figure 5B, lanes 1 and 2 through 4, respectively, the Q2 mutant did not direct mRNA 3′-end formation at the normal sites in the *rbcS*-E9 3′ region (Figure 5B, lanes 5 through 7). However, it also did not yield the expected 3′ ends at the downstream cryptic sites, as was seen with the C6 mutant (Figure 5B, lane 8). Instead, a population of nonpolyadenylated RNAs with 3′

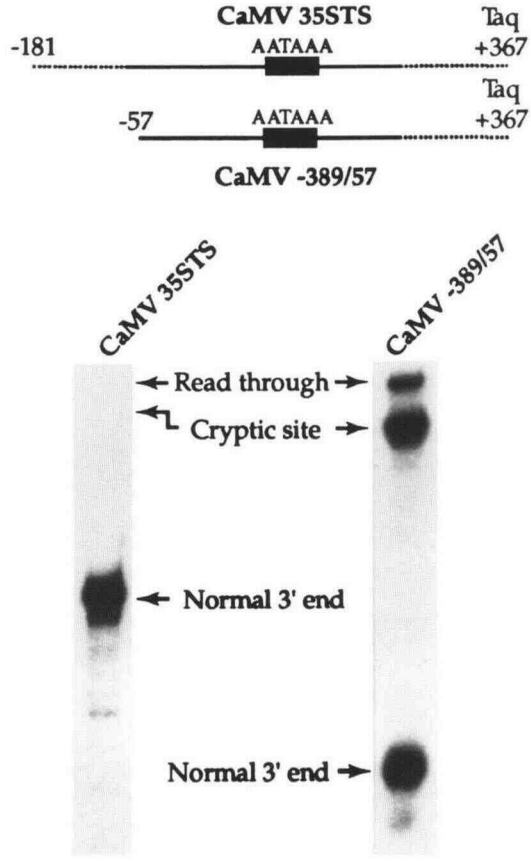

Figure 4. Sequences between −181 and −57 Are Required for Efficient Utilization of the CaMV Polyadenylation Signal.

The profiles of 3′ ends directed by a mutant carrying nucleotides −181 to +369 (CaMV 35STS) were compared with those directed by a mutant carrying nucleotides −57 to +369 (CaMV −389/57). Twenty micrograms of total RNA from plants carrying the CaMV 35STS or the CaMV −389/57 construction was annealed with end-labeled probe prepared from the corresponding mutant and processed as described in Methods. The positions of the normal polyadenylation site (normal 3′ end), an uncharacterized cryptic polyadenylation site located at roughly +300 (cryptic site), and the end of homology between the gene introduced into plants and the S1 probe (read through) are noted with arrows. (No detectable utilization of the cryptic site or read-through RNAs could be detected in the CaMV 35STS sample; the expected positions of the corresponding protected fragments are shown for comparison with the CaMV −389/57 mutant.) The structures of the CaMV 3′ regions included in the chimeric genes are shown above the autoradiograms.

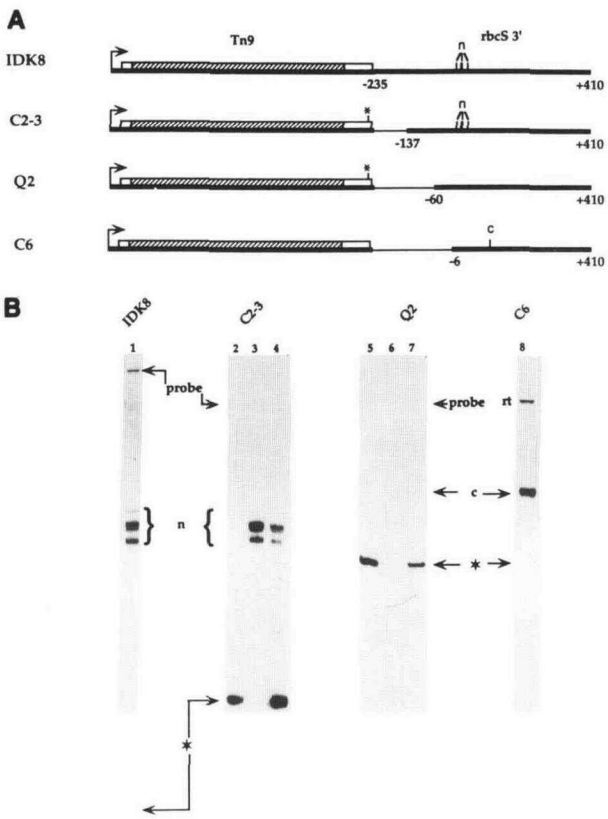

Figure 5. Sequences between −137 and −6 Affect the 3′-End Profile in the *rbcS*-E9 3′ Region.

(A) The structures of chimeric genes carrying the various *rbcS*-E9 3′ regions are shown. Portions of the *rbcS* region that were removed by deletion are represented with thin lines. A schematic representation of the Tn9 portion of these genes, including the locations of *cat* coding and noncoding regions, are shown. This region is derived from pSV2CAT (Gorman et al., 1982). The coordinates of the *rbcS* regions carried by each construction, defined with respect to one of the four polyadenylation sites in this gene (Coruzzi et al., 1984), are given beneath each construction. The positions of the 3′ ends of RNAs arising from each of these genes, as determined in Figure 5B, are denoted as n (the normal polyadenylation sites), c (the previously characterized cryptic site; Hunt, 1988), and * (the positions of the 3′ ends of the nonpolyadenylated RNAs noted below and in the text). The designation of each mutant is given to the left of each construction.

(B) The profiles of 3′ ends directed by mutants carrying nucleotides −235 to +410 of the *rbcS*-E9 3′ region (IDK8; lane 1), −137 to +410 (C2-3; lanes 2 through 4), −60 to +410 (Q2; lanes 5 through 7), and −6 to +410 (C6; lane 8) were compared. Twenty micrograms of total RNA (lanes 1, 4, 7, and 8), 20 μg of poly(A)-depleted RNA (lanes 2 and 5), or 1 μg of poly(A)-enriched RNA (lanes 3 and 6), isolated from plants carrying the respective construction, was annealed with end-labeled probe prepared from the corresponding mutant and processed as described in Methods. In the case of lanes 3 and 6, 20 μg of total RNA from control (untransformed) plants was included to better approximate the conditions used for the other samples. The positions of the normal polyadenylation sites in this gene (n), a cryptic polyadenylation site located at +89 (c), the site in the Tn9 region at which nonpolyadenylated RNAs end (*; see text), the end of homology between the gene introduced into plants and the S1 probe (rt), and the S1 probe (probe) are noted with arrows. Nonpolyadenylated RNA was not seen with the IDK8 mutant; the expected position of the corresponding protected fragment is shown for comparison with the C2-3 mutant. Likewise, no utilization of the cryptic site or read-through RNA could be detected with the Q2 mutant; the expected position of the bands is shown for comparison with the C6 mutant.

ends in the Tn9 portion of the chimeric gene, but downstream from the *cat* coding region, was seen (these RNAs are noted with an asterisk in Figure 5B). This behavior was not seen with IDK8 (Figure 5B, lane 1) nor with a more extensive deletion that directs polyadenylation at the cryptic sites present downstream from the normal sites (C6; Figure 5B, lane 8). However, the C2-3 mutant also yielded some RNAs with similar 3′ ends in addition to the normal 3′ ends (Figure 5B, lanes 2 through 4). These results indicated that the deletion in Q2 did not affect transcription of the chimeric gene. They also suggested that sequences upstream from −60 are needed for polyadenylation at the normal sites. However, the possibility that the Q2 deletion is predisposed toward an alternative mode of 3′-end formation and that the normal poly(A) sites in this construction are functional cannot be ruled out.

To understand better the behavior of the Q2 mutant, we removed the portion of the Tn9 region to where the 3′ ends of the nonpolyadenylated RNAs mapped and flanked this truncated gene with the bases −60 to +410 (Q2CC) and −137 to +410 (C2-3CC). When RNAs from the resulting chimeric genes were analyzed by S1 nuclease protection analysis, 3′ ends that mapped in the *rbcS*-E9 3′ region were seen, as shown in Figure 6. With C2-3CC, the RNAs produced ended at the normal sites (C2-3CC in Figure 6), indicating that the alteration in the *cat* gene had no effect on the function of these sites. With Q2CC, the resulting RNAs ended at the downstream cryptic sites (compare Figure 5B, lane 8, with Q2CC in Figure 6), suggesting that the normal sites present in this construction were no longer functional. Therefore, sequences between −60 and −137 are indeed needed for polyadenylation at the normal polyadenylation sites; when this region is removed, the resulting RNAs are recovered by the downstream cryptic sites.

DISCUSSION

Properties of the CaMV Polyadenylation Signal

We have identified two sequence elements in the CaMV 19S/35S 3′ region that can affect the efficiency of mRNA

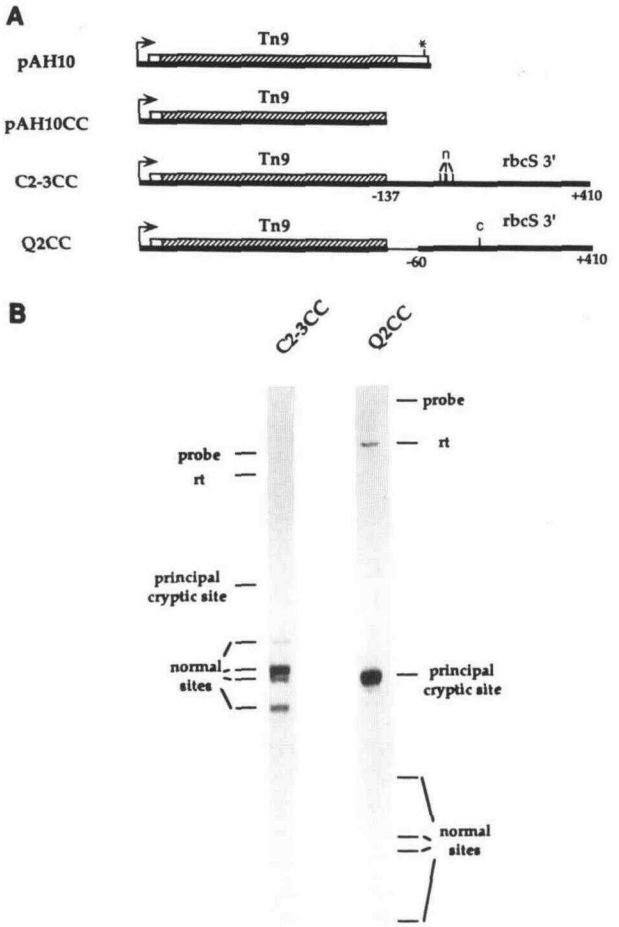

Figure 6. Sequences between −137 and −60 Are Essential for Polyadenylation at the Normal *rbcS*-E9 Sites.

(A) The structures of the C2-3CC and Q2CC constructions are shown, as are the parental Tn9 constructions used to assemble these test genes. The portion of the *rbcS* region in Q2CC that was removed by deletion is represented with a thin line. The coordinates of the *rbcS* regions carried by each construction are given beneath each construction. The positions of the 3′ ends of RNAs arising from each of these genes, as determined in Figure 6B, are denoted as n (the normal polyadenylation sites), and c (the previously characterized cryptic site). The designation of each mutant is given to the left of each construction.

(B) The profiles of 3′ ends directed by C2-3CC and Q2CC. Twenty micrograms of total RNA from plants carrying the C2-3CC or the Q2CC construction was annealed with end-labeled probe prepared from the corresponding mutant and processed as described in Methods. The positions of the normal polyadenylation sites in these genes (normal sites), the cryptic polyadenylation site (principal cryptic site), the end of homology between the gene introduced into plants and the S1 probe (rt), and the S1 probe (probe) are noted. No utilization of the cryptic site or read-through RNA could be detected with the C2-3CC mutant; the expected positions of the corresponding bands are shown for comparison with the Q2CC mutant. Likewise, the expected positions of bands corresponding to the normal polyadenylation sites are shown for the Q2CC mutant for comparison with the C2-3CC mutant.

3′-end formation at the CaMV polyadenylation site: a far-upstream region denoted by the end points of the CaMV 35STS and CaMV −389/57 end points (e.g., nucleotides −57 and −181) and a near-upstream region that may be analogous to the mammalian AAUAAA polyadenylation signal. Alteration of either of these elements reduces the efficiency of polyadenylation at the normal site, with concomitant production of RNAs with 3′ ends at alternate sites.

The far-upstream element located between 181 bases and 57 bases upstream from the CaMV polyadenylation site is similar in position to the −137 to −60 upstream element needed for polyadenylation in the *rbcS*-E9 gene (see below). In addition, it has two small sequences (CAUGUGUUG, at bases −69 to −77, and GGGUUUCGCU, at bases −78 to −87) that are similar to sequences present in the *rbcS*-E9 −137/−60 element (CAUUUGUUG, at bases −101 to −109, and GGUUUUCGCU, at bases −65 to −74), as shown in Figure 7. However, exact, or almost (one or two mismatches) exact, matches for these sequences cannot be found in many plant gene 3′ regions. There are no obvious potential secondary structures involving these elements, or any other part of this region, that might explain their roles in mRNA 3′-end formation. More extensive sequence comparisons and directed mutagenesis are needed to determine the roles, if any, of these sequences in mRNA 3′-end formation in plants.

When the AAUAAA motif located at bases −13 to −18 was changed, we found a significant reduction in the efficiency of utilization of the CaMV polyadenylation site. Although we have not systematically altered other parts of the region between −1 and −57, our data suggest that this motif has a role in mRNA 3′-end formation in the CaMV 19S/35S transcription unit. Sanfaçon and Hohn (1990) have reported similar results. Working with genes that carried the CaMV and nopaline synthase (*nos*) poly(A) signals in tandem, they found that virtually all RNAs end at the CaMV poly(A) site when an unaltered CaMV site is used, whereas no RNAs with 3′ ends at this site could be seen when the AAUAAA at positions −13 to −18 from the CaMV polyadenylation signal was removed; in the latter case, the detected RNAs ended in the downstream *nos* 3′ region. These findings are somewhat surprising because AAUAAA-like elements do not occur within 60 bases or so of roughly 30% of other plant polyadenylation sites (A.G. Hunt, manuscript in preparation). Nevertheless, it is possible that part of the recognition process that takes place during polyadenylation in plants can involve AAUAAA-like elements. This may be one factor that influences both the number and efficiencies of individual polyadenylation sites

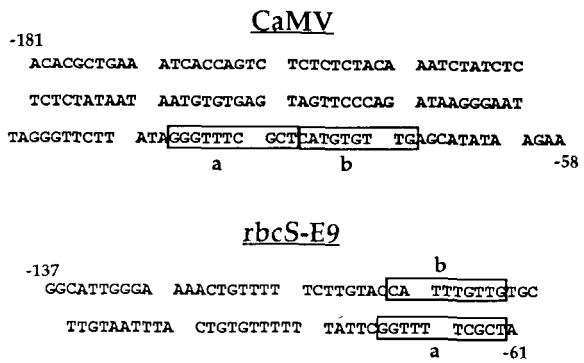

Figure 7. Sequences of the Far-Upstream Elements Needed for Efficient Polyadenylation in the CaMV and *rbcS*-E9 3′ Regions.

Numbers indicate the positions relative to the respective polyadenylation sites. Regions of sequence homology noted in the text are boxed. a, the GGUUUUCGCU motif; b, the CAUUUGUUG motif.

in a transcription unit as well as the relative contributions of the various sequence elements to efficiency of mRNA 3′-end formation.

The CaMV 35S RNA has a 181-nt terminal redundancy; for this RNA to be produced, the polyadenylation site present in the first 200 nt of this transcript is not utilized, but the identical site present some 8000 nt farther downstream is recognized. One possible hypothesis that explains how the terminally redundant 35S RNA is made postulates that sequences upstream from −181 (and, therefore, absent from the 5′ end of the 35S RNA but not the 3′ end) are needed for polyadenylation at the CaMV polyadenylation site. We see that sequences upstream from −181 [with respect to the polyadenylation site; this corresponds to the previously reported transcription initiation site (Guilley et al., 1982)] are not needed for efficient polyadenylation at the CaMV polyadenylation site (Figure 3). Therefore, this hypothesis can be ruled out. A more likely explanation is that offered by Sanfaçon and Hohn (1990): that proximity to the site of transcription initiation interferes with mRNA 3′-end formation.

Properties of the *rbcS*-E9 Polyadenylation Signal

We have found that a region between 60 bases and 137 bases upstream from one of the *rbcS*-E9 polyadenylation sites is needed for functioning of the sites normally utilized by this gene. Thus, when this region is removed, alternative RNA 3′ ends are produced (with Q2 and Q2CC). This region does not function to stabilize RNAs because stable, translatable RNAs that lack this region but not downstream portions of the *rbcS*-E9 3′ region can be produced (with C6, for example). Instead, the Q2CC–C2-3CC comparison (Figure 6) indicated that this region has a specific role in the formation of mRNA 3′ ends in the *rbcS*-E9 gene.

The region between −60 and −137 has no motifs similar to the AAUAAA consensus that is necessary for polyadenylation in mammalian systems. In addition to the similarities with the CaMV −181/−57 element noted above, this region has two UG motifs (UUUGUUGUG and UGUGUUUUU) that are similar to motifs that have been noted in the 3′-untranslated regions of several plant genes (Dunsmuir et al., 1983; Joshi, 1987). Similar sequence elements have been suggested to be required for mRNA 3′-end formation in two other plant genes (An et al., 1989; Ingelbrecht et al., 1989). However, the ubiquity and degree of conservation of these sequences in the 3′ regions of plant genes remain to be established. Further work is needed to clarify the functions, if any, of these sequences.

We have found that two of the mutants analyzed here (C2-3 and Q2) direct the production of stable, nonpolyadenylated RNAs that end in the Tn9 region of the appropriate chimeric gene. The means by which these RNAs are made are not known, but their properties suggest a number of interesting possibilities. These RNAs could arise from 3′ to 5′ exonucleolytic activities that are halted by a fortuitous block in the Tn9 sequences. However, the fact that removal of the proposed nuclease block (in the Q2CC mutant) results not in the abolition of stable *cat* RNA production but rather in the production of RNAs that end at poly(A) sites in the *rbcS*-E9 region suggests that nuclease action is not the mechanism by which these 3′ ends arise. Instead, the nonpolyadenylated RNAs probably represent the products of alternative processing or transcription termination at sites in the Tn9 sequences. Interestingly, this alternative mode of 3′-end formation predominates over polyadenylation at the downstream poly(A) sites in the Q2 mutant because removal of the site of alternative 3′-end formation is needed for utilization of the sites in the *rbcS* portion of this gene.

The *rbcS*-E9 3′ region has two elements that modulate 3′-end formation in the Tn9 region: one element, between −235 and −60 (probably sequences near −137, considering the distribution of 3′ ends seen in the C2-3 mutant), prevents RNA 3′-end formation at these sites; the other element, between −60 and −6, is needed for 3′-end formation at these sites, provided that the −235 to −137 element is absent. At present, the roles of these two elements in the formation of these nonpolyadenylated RNAs cannot be defined. In any case, however, it is clear that the regions between −235 and −137 and between −60 and −6 can affect the production of stable mRNAs in the context described here.

Messenger RNA 3′-End Formation in Plants

We have identified RNA sequences responsible for efficient mRNA 3′-end formation in two plant genes (CaMV −57/

−181 sequence and the *rbcS*-E9 −60/−137 element) that are located unusually far from their respective polyadenylation sites when compared with upstream sequences that are responsible for polyadenylation in mammalian genes; in the latter, the AAUAAA signal is generally the only upstream signal needed for efficient polyadenylation (Birnsteil et al., 1985; Platt, 1986). There seems to be no highly conserved sequence that may serve as a signal in the CaMV and *rbcS*-E9 far-upstream elements, although some similarities (noted above) may represent motifs needed for efficient 3′-end formation. However, because we do not know whether the far-upstream elements mapped in the CaMV and *rbcS*-E9 3′ regions are functionally equivalent, or whether they represent different classes of signals, we may not be able to identify a consensus sequence signal.

Yeast polyadenylation/termination signals apparently have no highly conserved consensus signal (Birnsteil et al., 1985; Osborne and Guarente, 1989). Moreover, yeast signals are situated at locations similar to the upstream elements we have mapped in the CaMV and *rbcS*-E9 3′ regions. These shared characteristics suggest that the process of mRNA 3′-end formation in plants and yeast may have mechanistic similarity, perhaps including a close link between transcription termination and mRNA polyadenylation in plants. We are currently exploring these possibilities.

We have also found that an AAUAAA motif is involved in efficient utilization of the CaMV polyadenylation signal. Sanfaçon and Hohn (1990) have reported similar results. In their studies, deletion of these nucleotides apparently abolished functioning of the CaMV polyadenylation signal. These observations suggest that mRNA 3′-end formation in plants may share features with the analogous process in mammalian cells. However, AAUAAA-like sequences cannot be absolutely required for polyadenylation of mRNAs in plants because such motifs are absent in many plant genes. There are two possible mechanistic explanations for these observations. AAUAAA-like elements may serve as auxiliary *cis* elements for polyadenylation in plants; in certain sequence contexts, or perhaps in concert with other elements, they would enhance cleavage and polyadenylation of pre-mRNAs. Alternatively, the putative plant AAUAAA-recognizing factor may be able to recognize sequences other than AAUAAA, or it may be one of a family of factors, other members of which would have sequence specificities other than AAUAAA. This would imply that sequences within 40 bases of plant polyadenylation sites in general would contain important determinants of poly(A) site choice and efficiency. However, efficient 3′-end formation would also require interactions with the factor that recognizes the upstream element noted above and perhaps with components that recognize sequences downstream from polyadenylation sites.

Plants have been found to recognize animal gene polyadenylation signals with reduced efficiency and at somewhat different sites (Hunt et al., 1987). This is probably due to the absence in animal gene polyadenylation signals of upstream elements corresponding to the CaMV −57/−181 sequence or the *rbcS*-E9 −60/−137 element. Indeed, the low efficiency with which mammalian polyadenylation signals are utilized in plants is qualitatively similar to the efficiency with which the CaMV −389/57 mutant is recognized. The observation that plants polyadenylate animal pre-mRNAs at sites different from those seen in animal cells would indicate that the spatial relationship between the proposed specificity factor that recognizes AAUAAA and the endonucleolytic cleavage site is different in plants from the analogous relationship in animals.

Although we can now draw certain analogies with other systems with regard to mRNA 3′-end formation in plants, there are still features of plant polyadenylation signals that are not understood. An example presented in this study is the ability of certain *rbcS*-E9 deletions to direct the production of nonpolyadenylated RNAs with 3′ ends within prokaryotic portions of our chimeric genes. We cannot explain this behavior, although it may reflect a link between 3′-end formation and transcription termination. However, these observations clearly indicate the complexity of the process of mRNA 3′-end formation in plants and magnify the importance of a clear understanding of this process. A necessary requisite for genetic engineering of plants is the ability to predict accurately the 3′ ends of engineered genes. This may not be as routine as normally assumed because sequences relatively far from polyadenylation sites can have dramatic effects on 3′-end formation, and changes distant from polyadenylation sites can yield RNAs with 3′ ends at unpredicted, and perhaps undesirable, locations.

METHODS

Recombinant DNA Manipulations

Our strategy for the mutational analysis of plant polyadenylation signals has been described in detail elsewhere (Hunt, 1988; Hunt and MacDonald, 1989). Basically, different portions of the 3′ region of interest are tested for their ability to direct polyadenylation of *cat*-containing RNAs in transgenic plants.

The construction of the IDK8, C2-3, and C6 *rbcS*-E9 clones has been described previously (Hunt, 1988, Hunt and MacDonald, 1989). The Q2 mutant was obtained by subcloning a TaqI fragment carrying bases−60 to +410 into AccI-digested pUC18 and selecting recombinants with the orientation BamHI-*rbcS*-PstI. The *rbcS* region was then moved into the 3′ end of the *cat* gene in pAH10 (Hunt, 1988) as a BamHI-PstI fragment and the resulting *cat*-Q2 cassette moved into the expression/binary shuttle vector p3-1 (Hunt, 1988) as a HindIII fragment.

To remove the alternative processing site in the 3′-untranslated part of the Tn9 sequences present in our constructions (see Figure 5), pAH10 was partially digested with ScaI and completely digested with SmaI, and the appropriate fragment was gel purified and ligated. These manipulations remove the last 26 nt of the

chloramphenicol acetyltransferase coding region and all of the noncoding part of the Tn9 sequences, including the alternative processing site. The C2-3 and Q2 mutants were moved into this construction (pAH10CC) as BamHI-PstI fragments and the resulting abbreviated *cat-rbcS* cassettes moved into p3-1 as HindIII fragments and recombinants with the proper orientations identified by restriction enzyme analysis of miniprep DNA.

The CaMV sequences were derived from a 758-bp HaeIII-TaqI fragment of the CaMV genome that carries sequences extending from 389 bases upstream from the normal CaMV poly(A) site (Guilley et al., 1982) to 369 bases downstream from this site. The clone CaMV 32-1 carries this entire region and is considered to be our wild-type control. All CaMV mutants were generated by oligonucleotide-directed mutagenesis (Kunkel, 1985). Mutants were obtained by producing uracil-containing single-stranded CaMV 32-1 DNA in *Escherichia coli* BD 2399 and using the following oligonucleotides for the mutagenesis reactions: 5′-ATCCTCTAGAGTCGACACGCTGAAATCACC-3′ was used to remove all sequences 5′ to the normal 35S transcription start site at −181 (clone CaMV 35STS); 5′-ATCCTCTAGAGTCGACCCT-TAGTATGTATT-3′ was used to remove all sequences upstream from position −57 including those of gene VI (clone CaMV −389/57); and 5′-TAAAATACTTCTATCTAGAATATTTCTAATTC-CTAA-3′ was used to mutate specifically the AAUAAA site to TAGAAT (clone CaMV-1). All mutations were verified through restriction mapping and sequence analysis.

Each mutant was subcloned into pAH10 as a BamHI-PstI fragment so that the *cat* gene in pAH10 was flanked with the sequence of interest in the proper orientation. The resulting *cat-rbcS* cassettes were then inserted as HindIII fragments behind the CaMV 35S promoter in p3-1, a Ti plasmid-associated expression/shuttle vector, or, for the experiments in Figures 3 and 4, behind a derivative of p3-1 (p3-1:35S^2) in which the CaMV 35S promoter was replaced with a 35S promoter containing a duplication of bases −416 to −90 with respect to the transcription initiation site; this duplication is similar to one reported by Kay et al. (1987) and yields greater levels of expression of foreign genes in transgenic plants. We have found in other studies that polyadenylation signals function identically regardless of which of these promoters is used to drive expression of the *cat*-3′-region genes. Recombinants with the proper orientation were identified by selection on 4 μg/mL chloramphenicol as described previously (Hunt, 1988; Hunt and MacDonald, 1989).

Transformation of Tobacco

The bacterial strains, tobacco cultivars, and procedures for *Agrobacterium*-mediated transformation of these cultivars used in this study have been described in detail previously (Schardl et al., 1987; Hunt, 1988; Hunt and MacDonald, 1989).

Transcript Mapping

RNA was isolated from pooled populations of transformants (six or more independent plants or cell lines), in some instances enriched for polyadenylated RNAs using oligo(dT)-cellulose, and hybridized with double-stranded DNA probes labeled (with Klenow) at the NcoI site in the *cat* gene in the various constructions as described previously (Hunt, 1988; Hunt and MacDonald, 1989). Probes were prepared by digesting the appropriate pAH10 derivative with NcoI, repairing the unpaired ends with Klenow in the presence of α-^{32}P-labeled deoxyribonucleotide triphosphates, excising the probe from the pAH10 derivative with PvuII, and purifying the probe from agarose gels. These probes carry, in order from the site of labeling, 230 bp of the *cat* gene, the 3′ region of interest, and 197 bp of the *lac* operon of pUC19 that serve to distinguish undigested probe from that portion of the probe protected by RNAs that extend through the *rbcS* region. For RNA-DNA hybridizations, RNA [20 μg to 30 μg of total or poly(A)-depleted RNA, or 1 μg to 2 μg of poly(A)-enriched RNA] and probe (20,000 cpm to 100,000 cpm, depending on the specific activity of the probe; this corresponds to roughly 20 fmol to 100 fmol of double-stranded probe) were dissolved in 10 μL of hybridization buffer (80% formamide + 0.04 M Pipes, pH 6.4, 0.4 M NaCl, 1 mM EDTA), incubated at 85°C to 90°C for 5 min, and then at 52°C (or 47°C for the analysis of the C6 mutant) for 12 hr to 18 hr. These temperatures and formamide concentrations were empirically determined to optimize hybrid formation and minimize probe renaturation. After 12 hr to 16 hr at the appropriate temperature, the hybridization mixtures were diluted into 10 volumes of S1 digestion buffer (0.28 M NaCl, 0.05 M sodium acetate, 4.5 mM $ZnSO_4$) containing 100 units of S1 nuclease and incubated for 40 min at 30°C, and the protected fragments were recovered by ethanol precipitation. After dissolving and reprecipitating, each sample was analyzed on 6% polyacrylamide sequencing gels. Each sequencing gel was calibrated with DNA size standards ranging from 110 nt to more than 1300 nt; from these standards, the expected positions of bands that could not be detected were inferred and the identities of protected fragments confirmed. Each different probe was also tested with RNA prepared from untransformed tobacco; in all cases, no protected bands were seen, indicating that the data presented here describe transcripts arising from the 35S-*cat*-3′-region constructions described in detail in the text.

Steady-state levels of *cat*-containing RNAs from different mutants were compared by analyzing equal amounts (20 μg) of total RNA with a probe prepared from a heterologous mutant (e.g., CaMV mutants were analyzed with an *rbcS* probe and vice versa) as described above. In the resulting hybrids, the RNA from each sample would protect the probe only through the end of the Tn9 region of the probe (yielding a protected fragment of 244 nt). Because the same probe was used for each hybrid, and each hybrid has the same length, it is possible to compare directly the steady-state levels of all *cat*-containing RNAs that arise from the poly(A) test genes.

ACKNOWLEDGMENTS

We thank Joe Chappell and Robert Klein for helpful suggestions. This work was supported by U.S. Department of Agriculture Competitive Grant No. 85-CRCR-1-1810 and U.S. Department of Agriculture-University of Kentucky Cooperative Agreement No. 58-43YK-7-0025. The investigation reported in this paper (No. 90-3-133) is in connection with a project of the Kentucky Agricultural Experiment Station and is published with approval of the Director.

Received August 23, 1990; accepted October 9, 1990.

REFERENCES

An, G., Mitra, A., Choi, H.K., Costa, M.A., An, K., Thornburg, R.W., and Ryan, C.A. (1989). Functional analysis of the 3′ control region of the potato wound-inducible proteinase inhibitor II gene. Plant Cell **1,** 115–122.

Birnsteil, M.L., Busslinger, M., and Strub, K. (1985). Transcription termination and 3′ processing: The end is in site! Cell **41,** 349–359.

Butler, J.S., and Platt, T. (1988). RNA processing generates the mature 3′ end of yeast *CYC1* messenger RNA in vitro. Science **242,** 1270–1274.

Butler, J.S., Sadhale, P.P., and Platt, T. (1990). RNA processing in vitro produces mature 3′ ends of a variety of *Saccharomyces cerevisiae* mRNAs. Mol. Cell. Biol. **10,** 2599–2605.

Christofori, G., and Keller, W. (1988). 3′ cleavage and polyadenylation of mRNA precursors in vitro requires a poly(A) polymerase, a cleavage factor, and a snRNP. Cell **54,** 875–889.

Christofori, G., and Keller, W. (1989). Poly(A) polymerase purified from HeLa cell nuclear extract is required for both cleavage and polyadenylation of pre-mRNA in vitro. Mol. Cell. Biol. **9,** 193–203.

Connelly, S., and Manley, J.L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev. **2,** 440–452.

Coruzzi, G., Broglie, R., Edwards, C., and Chua, N.-H. (1984). Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase. EMBO J. **3,** 1671–1679.

Dean, C., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H., and Bedbrook, J. (1986). mRNA transcripts of several higher plant genes are polyadenylated at multiple sites in vivo. Nucl. Acids Res. **14,** 2229–2240.

Dunsmuir, P., Smith, S.M., and Bedbrook, J. (1983). The major chlorophyll *a/b* binding protein of petunia is composed of several distinct polypeptides encoded by a number of distinct nuclear genes. J. Mol. Appl. Genet. **2,** 285–300.

Gil, A., and Proudfoot, N.J. (1987). Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit β-globin mRNA 3′ end formation. Cell **49,** 399–406.

Gilmartin, G.M., McDevitt, M.A., and Nevins, J.R. (1988). Multiple factors are required for specific RNA cleavage at a poly(A) addition site. Genes Dev. **2,** 578–587.

Gorman, C.M., Moffat, L.F., and Howard, B.H. (1982). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2,** 1244–1251.

Guilley, H., Dudley, R.K., Jonard, G., Balazo, E., and Richards, K.E. (1982). Nucleotide sequence of cauliflower mosaic virus DNA. Cell **30,** 763–773.

Hernandez, G., Cannon, F., and Cannon, M. (1989). The effect of presumptive polyadenylation signals on the expression of the cat gene in transgenic tobacco. Plant Cell Rep. **8,** 195–198.

Hunt, A.G. (1988). Identification and characterization of cryptic polyadenylation sites in the 3′ region of a pea *rbcS* gene. DNA **7,** 329–336.

Hunt, A.G., and MacDonald, M. (1989). Deletion analysis of the polyadenylation signal of a pea ribulose-1,5-bisphosphate carboxylase small subunit gene. Plant Mol. Biol. **13,** 125–138.

Hunt, A.G., Chu, N.M., Odell, J.T., Nagy, F., and Chua, N.-H. (1987). Plant cells do not properly recognize animal gene polyadenylation signals. Plant Mol. Biol. **7,** 23–35.

Ingelbrecht, I.L.W., Herman, L.M.F., Dekeyser, R.A., Van Montagu, M.C., and Depicker, A.G. (1989). Different 3′ end regions strongly influence the level of gene expression in plant cells. Plant Cell **1,** 671–680.

Joshi, C.P. (1987). Putative polyadenylation signals in nuclear genes of higher plants: A compilation and analysis. Nucl. Acids Res. **15,** 9627–9640.

Kay, R., Chan, A., Daly, M., and McPherson, J. (1987). Duplication of CaMV 35S promoter sequences creates a strong enhancer for plant genes. Science **236,** 1299–1302.

Kunkel, T.A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA **82,** 488–492.

Logan, J.E., Falk-Pederson, J.E., Darnell, J.E., and Shenk, T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse β-globin gene. Proc. Natl. Acad. Sci. USA **84,** 8306–8310.

McDevitt, M.A., Imperiale, M.J., Ali, H., and Nevins, J.R. (1984). Requirement of a downstream sequence for generation of a poly(A) addition site. Cell **37,** 993–999.

McDevitt, M.A., Gilmartin, G.M., Reeves, W.H., and Nevins, J.R. (1988). Multiple factors are required for poly(A) addition to a mRNA 3′ end. Genes Dev. **2,** 588–597.

Messing, J., Geraghty, D., Heidecker, H., Hu, N.T., Kridl, J., and Rubenstein, I. (1983). Plant gene structure. In Genetic Engineering of Plants: An Agricultural Perspective, T. Kosuge, P. Meredith, and A. Hollaender, eds (New York: Plenum Press), pp. 211–227.

Nagy, F., Morelli, G., Fraley, R.T., Rogers, S.G., and Chua, N.-H. (1985). Photoregulated expression of a pea *rbcS* gene in leaves of transgenic plants. EMBO J. **4,** 3063–3068.

Osborne, B.I., and Guarente, L. (1989). Mutational analysis of a yeast transcriptional terminator. Proc. Natl. Acad. Sci. USA **86,** 4097–4101.

Platt, T. (1986). Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55,** 339–372.

Russo, P., and Sherman, F. (1989). Transcription terminates near the poly(A) site in the *CYC1* gene of the yeast *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **86,** 8348–8352.

Sanfaçon, H., and Hohn, T. (1990). Proximity to the promoter inhibits recognition of cauliflower mosaic virus polyadenylation signal. Nature **346,** 81–84.

Schardl, C., Byrd, A.D., Benzion, G., Altschuler, M., Hildebrand, D.F., and Hunt, A.G. (1987). Design and construction of a versatile system for the expression of foreign genes in plants. Gene **61,** 1–11.

Sisodia, S.S., Sollner-Webb, B., and Cleveland, D.W. (1987). Specificity of RNA maturation pathways: RNAs transcribed by RNA polymerase III are not substrates for splicing or polyadenylation. Mol. Cell. Biol. **7,** 3602–3612.

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1988). Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. Cell **52,** 731–742.

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1989). Four factors are required for 3′-end cleavage of pre-mRNAs. Genes Dev. **3,** 1711–1724.

Whitelaw, E., and Proudfoot, N. (1986). α-Thallassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3′ end processing in the human α-2 globin gene. EMBO J. **5,** 2915–2922.