Nucleic Acids Research, Vol. 19, No. 20 5575–5581

Characterization of the polyadenylation signal from the T-DNA-encoded octopine synthase gene

Margaret H.MacDonald, Bradley D.Mogen and Arthur G.Hunt*
Plant Physiology/Biochemistry/Molecular Biology Program, Department of Agronomy, University of Kentucky, Lexington, KY 40546–00914, USA

Received July 22, 1991; Accepted August 16, 1991

ABSTRACT

We have characterized the polyadenylation signal from the octopine synthase (*ocs*) gene. This signal directs mRNA 3′ end formation at a number of distinct sites. A combination of deletion and linker-substitution analyses revealed that each of these sites is controlled by multiple upstream sequence elements. Upstream sequences relatively far (>80 nt) from the *ocs* poly[A] sites were found to be needed for functioning of these sites. Upstream sequences nearer to poly[A] sites were also found to be involved in mRNA 3′ end formation in the *ocs* gene. In addition, a set of novel elements that mediates 3′ end choice was uncovered by deletion analysis of sequences downstream from the *ocs* polyadenylation sites. Our experiments indicate mRNA 3′ end formation in the *ocs* is controlled by a complex series of cis-acting signals, and suggest that the process of mRNA 3′ end formation might be linked to transcription termination.

INTRODUCTION

The nature of the sequence signals responsible for the generation of proper 3′ termini of plant mRNAs is not well understood. It has been widely assumed that the hexanucleotide AAUAAA or related sequences is needed for efficient polyadenylation in plants, as is the case in mammals (1, 2). Indeed, recent work has revealed an AAUAAA motif that lies 13–18 nt upstream from the 3′ ends of the cauliflower mosaic virus (CaMV) 19S and 35S RNAs is required for efficient utilization of this site (3–5). However, AAUAAA cannot be the sole determinant of polyadenylation in plant genes because many plant genes have no such sequence in their 3′ regions (6).

Recent studies have also revealed a requirement for less well-defined upstream sequences unusually far from polyadenylation sites for mRNA 3′ end formation in the CaMV 19S/35S and pea *rbcS*-E9 genes (4, 5). This feature clearly distinguishes the two plant polyadenylation signals studied thus far from their mammalian counterparts, with the possible exception of certain animal virus-encoded signals (7–9). The nature of these 'far-upstream elements' (FUEs) is unclear; it is not known whether the CaMV and *rbcS* elements are functionally interchangeable, or if such elements are needed for mRNA 3′ end formation in other plant genes. Also, the identity of the precise nucleotides that comprise these elements is not known.

Here, we describe an analysis of the polyadenylation signal from the *Agrobacterium tumefaciens* Ti plasmid-encoded octopine synthase (*ocs*) gene. We show that, as has been reported for the CaMV and *rbcS* polyadenylation signals, upstream sequences rather far from certain of the *ocs* polyadenylation sites are needed for functioning of these sites. In addition, we demonstrate that upstream sequences in closer proximity to polyadenylation sites also play a role in 3′ end site choice in this gene. Finally, we describe novel signals that mediate differential 3′ end selection in this gene.

MATERIALS AND METHODS

Recombinant DNA manipulations

Our strategy for the characterization of plant polyadenylation signals has been described in detail elsewhere (5, 10). Basically, different portions of the 3′ region of interest are tested for their ability to direct polyadenylation of chloramphenicol acetyltransferase (*cat*)-containing RNAs in transgenic plants.

The 'wild-type' *ocs* polyadenylation signal used here (ocs950) was cloned from a pBR322 derivative that carried the HindIII-1 fragment of pTiAch5 as a TaqI fragment into pUC18 with a resulting orientation of 5′-SstI-*ocs* 3′-PstI-3′. The ocsDP and ocsRP constructions were built by digesting the wild-type clone with DdeI or RsaI, respectively, repairing (in the case of the DdeI digestion) with Klenow, digesting with PstI, and purifying the appropriate fragments from agarose gels. These fragments were subcloned into HincII + PstI-digested pUC18, yielding constructions with the same 5′-SstI-*ocs* 3′-PstI-3′ orientation as the wild-type clone. The ocsSH, ocsSB, ocsSA, and ocsSS mutants were obtained by digesting the wild-type clone with HaeIII, BstEII, AhaIII, and SspI, respectively, repairing (in the case of the BstEII digestion) with Klenow, digesting with SstI, and purifying the appropriate fragments from agarose gels. These fragments were subcloned into SstI + SmaI-digested pUC18, yielding constructions with the same 5′-SstI-*ocs* 3′-PstI-3′ orientation as the wild-type clone. The ocs-41 mutant was derived from the ocs61/42 linker substitution mutant (see below). For

* To whom correspondence should be addressed

this, the ocs61/42 clone was digested with SstII and PstI, the appropriate fragment purified from an agarose gel and subcloned into SstII + PstI-digested pBluescript KS+ (Stratagene, La Jolla, CA), yielding a clone with the orientation 5'-SstI-*ocs* 3'-PstI-3'.

The three linker substitution mutants (ocs61/42, ocs41/22, and ocs21/2) were generated by oligonucleotide-directed mutagenesis (11). The wild-type *ocs* 3' region was subcloned into pBluescript KS+ (Stratagene) as a BamHI-PstI fragment for this purpose. Using this clone, uracil-containing single-stranded ocs950 DNA was produced in *E.coli* BD 2399 and used with the following oligonucleotides for the mutagenesis reactions: 5'-TACCGC-CGGT-TTCGGGGTCACCCCGCGGGTCTAGAACCCGTT-ACTATCGT-3' was used to replace nucleotides −61 to −42 (all numbering is with respect to polyadenylation site 1; see below) with the sequence GGTCACCCCGCGGGTCTAGA; T-ATCGGTCACCCCGCGGGTCTAGATATGAATAATAAT-CT-3' was used to replace nucleotides −41 to −22 with the same sequence; and 5'-TTACTATCGTATTTTGGTCACCCCGCG-GGTCTAGACAATTTACTGATTGT-3' was used to replace nucleotides −21 to −2 with the same sequence. All mutations were verified by restriction endonuclease mapping.

Each mutant was subcloned into pAH10 or pAH10CC (5, 10) as a BamHI-PstI or SstI-PstI fragment so that the *cat* gene in pAH10 was flanked with the sequence of interest in the proper orientation. The resulting *cat-ocs* cassettes were then inserted as HindIII fragments behind the cauliflower mosaic virus (CaMV) 35S promoter in p3-1, a Ti plasmid-associated expression/shuttle vector (12), or 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 −415 to −90 with respect to the transcription initiation site (5). We have found in other studies that polyadenylation signals function identically regardless of the promoter 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 (12).

Transformation of tobacco

The bacterial strains, tobacco cultivars, and procedures for *Agrobacterium*-mediated transformation have been described in detail elsewhere (12).

Transcript mapping

The procedures for mapping the 3' ends of *cat*-containing RNAs have been described in detail elsewhere (5, 10, 13). Briefly, total RNA was isolated from pooled populations of transformants (6 or more independent plants or cell lines) and hybridized with double-stranded DNA probes labeled (with Klenow) at the NcoI or EcoRI sites in the *cat* gene in the various pAH10 derivatives. Probes were prepared by digesting the appropriate pAH10 derivative with NcoI or EcoRI, repairing the unpaired ends with Klenow in the presence of α-^{32}P-labeled deoxyribonucleotide triphosphates, excising the probes with PstI or PvuII (the latter was used for probes prepared from pAH10 derivatives carrying the ocsDP, ocs-41, and ocsRP mutations), and purifying the probes from agarose gels. These probes carry, in order from the site of labeling, 230 (for NcoI probes) or 530 (for EcoRI probes) bp of the *cat* gene and the 3' region of interest. In addition, probes excised with PvuII carry 197 bp of the *lac* operon of pUC19 that can serve to distinguish undigested probe from that portion of the probe protected by RNAs that extend through the *ocs* region. RNA-DNA hybridizations, S1 nuclease treatments, and sequencing gel analyses were as described previously (5, 10). Each sequencing gel was calibrated with DNA size standards ranging from 110 to over 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.

For exonuclease VII digestions, hybridizations were diluted into 10 volumes of ExoVII digestion buffer (30 mM KCl, 10 mM Tris HCl, pH 7.4, 10 mM EDTA) containing 4 U/ml of exonuclease VII, incubated for 60 minutes at 37°C, and the protected fragments recovered and analyzed as described for the S1 nuclease protection experiments.

For Northern blots and hybridizations, total RNA (20 μg) was separated in a 1.1% agarose gel containing 0.66% formaldehyde in MOPS buffer and electroblotted in TAE buffer (20 mM Tris-Cl, pH 7.8, 10 mM sodium acetate, 0.5 mM EDTA) onto a Nytran (Schleicher & Schuell, Keene, NH) membrane. The membrane was incubated overnight at 43°C in a solution composed of formamide (50%), 4×SSCP (1×SSCP: 0.12 M NaCl, 15 mM Na citrate, 13 mM potassium phosphate, pH 7.2, and 1 mM EDTA), denatured salmon sperm DNA (0.5 mg/ml), Denhardt's solution (10×), SDS (1%), and nonfat dried milk (1%). The probe, a *cat* antisense RNA uniformly labeled with α-^{32}P ATP using T7 RNA polymerase, was denatured with

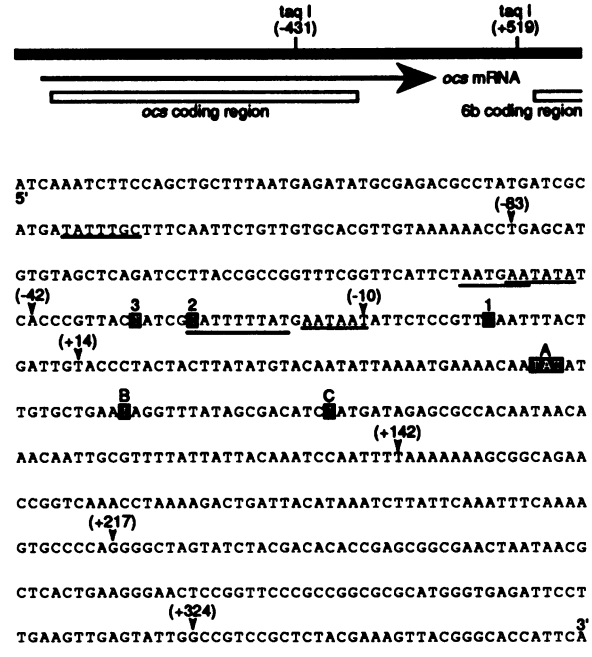

Figure 1. Nucleotide sequence of the 3' region of the Ti plasmid-encoded octopine synthase (*ocs*) gene. The illustration at the top depicts the region of the T-DNA from which the *ocs* 3' region was derived. Shown beneath this is the sequence extending from the translation termination codon to a point 330 bp beyond the polyadenylation site noted by Dhaese *et al.* (2) as the principal site (site 1 in this study). This corresponds to nucleotides −192 to +330. Highlighted are the various deletion endpoints (➤), the different polyadenylation sites described here (boxed nucleotides), and motifs related to AAUAAA (underlined nucleotides) and UAUUUGUA (bold underlines).

formamide, added directly to the prehybridization solution, and the hybridization allowed to proceed for 16 hours at 43°C. The membrane was then sequentially washed for 15 minutes each at room temperature in 2×SSCP/0.1% SDS, 0.5×SSCP/0.1% SDS, and 0.1×SSCP/0.1% SDS. Final high stringency washes were performed in 0.1×SSCP/1.0% SDS as follows: 50°C for 1 hour, 65°C for 1 hour, and 75°C for 1 hour. The filter was then dried and exposed to Kodak XAR-5 film.

RESULTS

The octopine synthase polyadenylation signal directs mRNA 3′ end formation at two or three predominant sites

The region of the *ocs* gene that we analyzed in this study is shown in Figure 1. This region extends from 431 base pairs upstream of the principal poly[A] addition site noted by Dhaese *et al.* (2) to 519 base pairs downstream from this site (i.e. base pairs −431 to +519). It includes the last 239 base pairs of the reading frame carried by the *ocs* mRNA and 711 base pairs of the region between the *ocs* gene and the adjacent T-DNA gene that encodes transcript 6b of octopine-type Ti plasmids. The 3′ end of the region studied here (i.e. nt +519) lies some 167 base pairs upstream from the translation initiation codon of the 6b gene.

When the 'wild-type' *ocs* 3′ region was assessed for ability to direct mRNA polyadenylation, the resulting RNAs were found to end at a distinct series of sites (Figure 2). One of these sites (noted as site 1 in Figure 2) corresponded to the major site reported by Dhaese *et al.* (2). However, instead of the single 'minor' site reported by these authors (2), two additional prominent 3′ ends (noted as sites 2 and 3 in Figure 2) were seen. Although we have not precisely mapped these 3′ ends, site 2 seemed to correspond to the previously-reported 'minor' site, judging from the electrophoretic mobilities of each protected fragment (data not shown). It is possible that site 2 is identical to the minor site noted previously, and that site 3 is a nuclease protection artifact. However, these sites can be differentially affected by distant mutations (see below) and have been consistently seen in all appropriate constructions and in repeated nuclease protection experiments. In addition, the nucleotide composition surrounding site 3 is not remarkably AT rich, making the possibility of digestion at sites of 'breathing' of the RNA/DNA hybrid unlikely. Currently, we cannot explain the differences between our nuclease protection results and those of Dhaese *et al.* (2). Because of the consistency of our results, we will present our studies in the context described for Figure 1 here, keeping in mind the possibility that sites 2 and 3 may be the same.

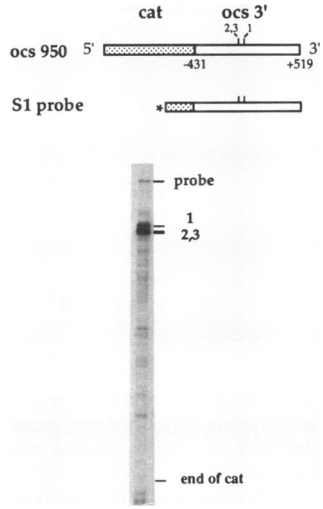

Figure 2. The wild-type *ocs* polyadenylation signal directs mRNA 3′ end formation at a series of distinct sites in tobacco. The structure of the *cat*-ocs950 chimeric gene is illustrated as is the structure of the end-labeled DNA fragment used for the nuclease protection analysis. 20 μg of total RNA from a pooled population of ocs950-carrying transgenic plants was annealed with the pictured probe (end-labeled at the NcoI site in the *cat* gene), the hybrids treated with nuclease S1, and the protected fragments separated on a 6% sequencing gel as described in Materials and Methods. The positions of the renatured probe (probe), protected fragments corresponding to sites 1, 2, and 3 (1 and 2, 3; see text), and the junction of the *cat* and *ocs* sequences (end of *cat*) are shown next to the autoradiograph.

Upstream sequences unusually far from polyadenylation sites are needed for mRNA 3′ end formation at such sites

In order to identify the upstream sequences needed for mRNA 3′ end formation in the *ocs* 3′ region, a series of deletions (ocsDP, ocs-41, and ocsRP) were analyzed in the manner described above. The ocsDP and ocs-41 constructions, with respective 5′ endpoints at positions 83 and 42 nt upstream from site 1, yielded populations of RNAs with predominant 3′ ends at a series of sites some 50−100 nt downstream from site 1 (A, B, and C in Figure 3). These polyadenylation sites were apparent in constructions with the wild-type site, but were relatively rare with respect to sites 1−3 (see ocs950 in Figures 1 and 4). In the ocsDP and ocs-41 mutants, on the other hand, the 3′ ends at sites A, B, and C were the predominant ones. Because there were no striking differences between the 3′ end profiles seen with these two mutants, we

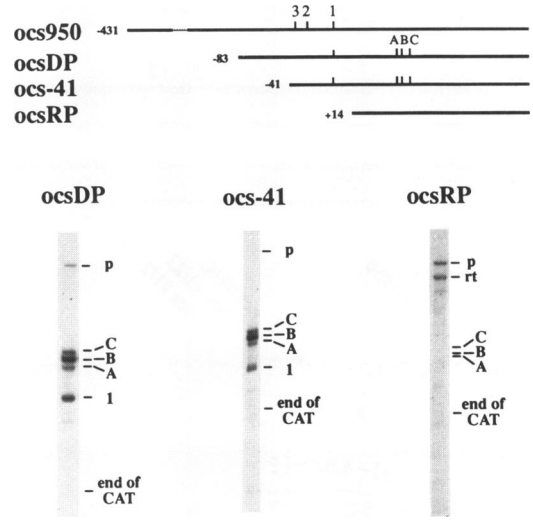

Figure 3. Definition of upstream sequences involved in mRNA 3′ end formation in the *ocs* 3′ region. The structures of the wild-type *ocs* polyadenylation signal (ocs950) and of three deletion mutants are illustrated, showing the relative positions and identities of the various protected fragments seen in the experiment. The deletion endpoints are given to the left of each construction; solid lines represent that part of the *ocs* 3′ region retained in each construction. 20 μg of total RNA from pooled populations of transgenic plants carrying each construction was annealed with probes prepared by labeling at the NcoI (for the ocsDP and ocs-41 constructions) or EcoRI (for the ocsRP mutant) sites in the *cat* gene, the hybrids treated with nuclease S1, and the protected fragments separated on a 6% sequencing gel as described in Materials and Methods. The positions of the reannealed probe (p), readthrough RNAs (rt), protected fragments corresponding to sites 1, 2, 3, A, B, and C, and the junction of the *cat* and *ocs* sequences (end of cat) are shown to the right of each autoradiograph.

conclude that sequences farther than 83 nt upstream from site 1 are important for efficient functioning of sites 1, 2, and 3, and sequences between −83 and −42 are not required for the functioning of sites A, B, and C.

The ocsRP construction was unable to direct mRNA 3′ end formation in the *ocs* region of the 35S promoter-*cat*-*ocs* chimeric gene (Figure 3). However, *cat*-containing RNAs with homology extending through the end of the *cat* gene could be detected (rt in Figure 3). Therefore, the ocsRP construction generated stable *cat* transcripts from a functional 35S promoter. However, this deletion no longer had the ability to direct polyadenylation in the *ocs* 3′ region, indicating that sequences between −42 and +14 (with respect to site 1) are needed for functioning of the three downstream sites (A, B, and C) in the *ocs* 3′ region, and that this region controls all three of these downstream sites.

Sequences within 40 nt of sites 1, 2, and 3 are involved in the functioning of these sites

Studies of the CaMV polyadenylation signal have shown that the motif AAUAAA, located 13−18 nt upstream from the CaMV polyadenylation site, is involved in efficient 3′ end formation at this site (3−5). In order to search the corresponding region of the *ocs* 3′ region for signals important for polyadenylation at sites 1, 2, and 3, three linker substitution (LS) mutants were created in which 20 nt increments beginning with nt −2 were replaced with a series of unrelated bases. These were tested for function as described above.

When nucleotides between −61 and −42 were replaced with 20 unrelated bases (ocs61/42), the resulting mutants were noticeably impaired in 3′ end formation at sites 2 and 3, but not at site 1 (Figure 4). This mutation thus defines a sequence element that specifically affects sites 2 and 3 in the *ocs* 3′ region, but not site 1.

The ocs41/22 mutant also was relatively unaffected in terms of its ability to direct 3′ end formation at site 1 (Figure 4). In contrast to the ocs61/42 mutant, the ocs41/22 mutant also yielded 3′ ends at site 2. However, protected fragments corresponding to site 3 were not seen with the ocs41/22 mutant. This linker substitution thus differentiates between sites 2 and 3, either by eliminating a cis element needed for functioning of just site 3 (if sites 2 and 3 are distinct) or by altering the nucleotide composition surrounding site 3 so that it is no longer apparent (if site 3 is derived from artifactual digestion of RNA/DNA hybrids that end at site 2).

With the ocs21/2 mutant, site 1 was eliminated but sites 2 and 3 remained relatively unaffected (Figure 4). Thus, this substitution identifies a cis element that is specific for site 1 in the *ocs* 3′ region.

Interestingly, in all three LS mutants, the relative proportion of 3′ ends at sites A, B, and C were not significantly greater than was seen in the wild-type ocs950 construction. This

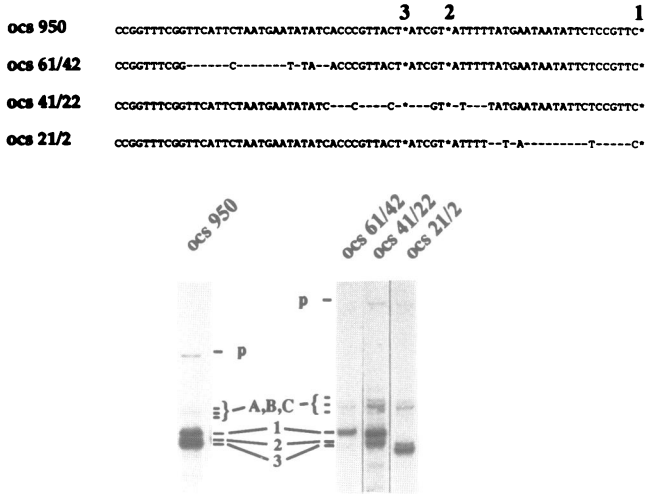

Figure 4. Sequences near the *ocs* polyadenylation sites affect mRNA 3′ end formation. The sequences between −1 and −70 of the wild-type ocs 3′ region (ocs950) and of three linker substitution mutations are shown at the top of the figure. Dashes represent wild-type bases that have been replaced by bases in the linker, and stars the positions (calculated from mobilities on sequencing gels) of sites 1, 2, and 3 (in order from right to left, respectively). 20 μg of total RNA from pooled populations of transgenic plants carrying each construction was annealed with probes prepared by labeling at the NcoI sites in the *cat* gene of the pAH10 derivative of each mutant, the hybrids treated with nuclease S1, and the protected fragments separated on a 6% sequencing gel as described in Materials and Methods. The positions of the reannealed probe (p) and protected fragments corresponding to sites 1, 2, 3, A, B, and C are shown to the right of each autoradiograph.

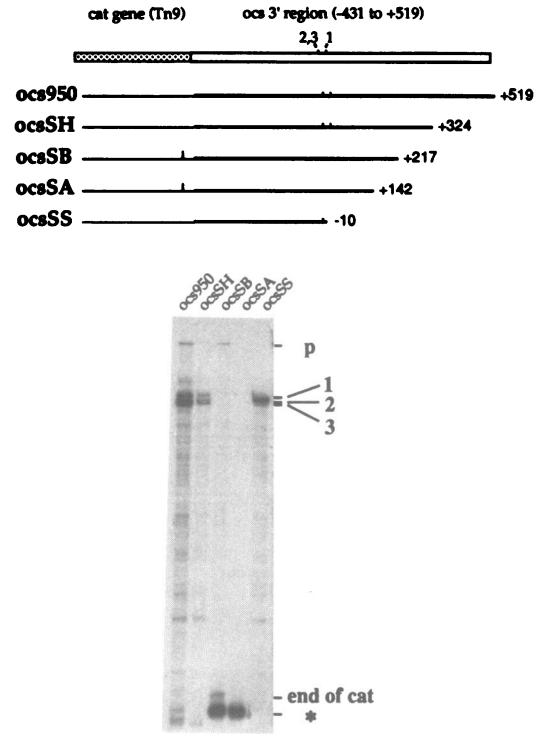

Figure 5. Analysis of sequences downstream from the *ocs* polyadenylation sites. The structures of the wild-type *ocs* polyadenylation signal (ocs950) and of four deletion mutants are illustrated, showing the relative positions and identities of the various protected fragments seen in the experiment. The deletion endpoints are given to the right of each construction; thick solid lines represent that part of the *ocs* 3′ region retained in each construction. 20 μg of total RNA from pooled populations of transgenic plants carrying each construction was annealed with a probe prepared by labeling the pAH10 derivative of ocsSH at the NcoI site in the *cat* gene, the hybrids treated with nuclease S1, and the protected fragments separated on a 6% sequencing gel as described in Materials and Methods. The positions of the reannealed probe (p), protected fragments corresponding to sites 1, 2, 3, and of the 3′ ends of the novel RNAs that end in the Tn9 region (*), and the junction of the *cat* and *ocs* sequences (end of cat) are shown to the right of the autoradiograph. Note that a protected fragment that apparently corresponds to site 1 is seen with the ocsSS mutant; this band actually represents the end homology between the probe and the 3′ region present in the transgenic plants (unpublished observations).

observation indicates that the relative efficiencies of those sites not qualitatively affected by a particular LS mutation are similar, so that the proportion of readthrough RNAs (which would end at sites A, B, and C) remains constant. This is different from what was seen with the deletion mutations described above, and suggests that the types of cis elements defined by the LS mutations are different from those defined by the deletions.

Characterization of sequences downstream from the *ocs* polyadenylation sites

In order to probe the region downstream from the *ocs* polyadenylation sites for elements involved in mRNA 3′ end formation, a series of deletions in which progressively larger parts of the downstream sequences were removed were analyzed. When the 3′-most 195 bases were deleted from the wild-type construction, the resulting mutant (ocsSH) directed polyadenylation at the same sites as did the wild-type *ocs* region (ocsSH in Figure 5). When an additional 106 bases were deleted, the resulting 3′ region (ocsSB) no longer directed polyadenylation of *cat*-containing RNAs at the expected sites (ocsSB in Figure 5). Instead, a population of RNAs with 3′ ends in the Tn9 portion of the chimeric gene was seen. The protected fragments seen in the S1 nuclease protection experiments represented true 3′ ends because similar fragments were protected against exonuclease VII, and because appropriately-sized RNAs could be seen by Northern blot analyses (data not shown).

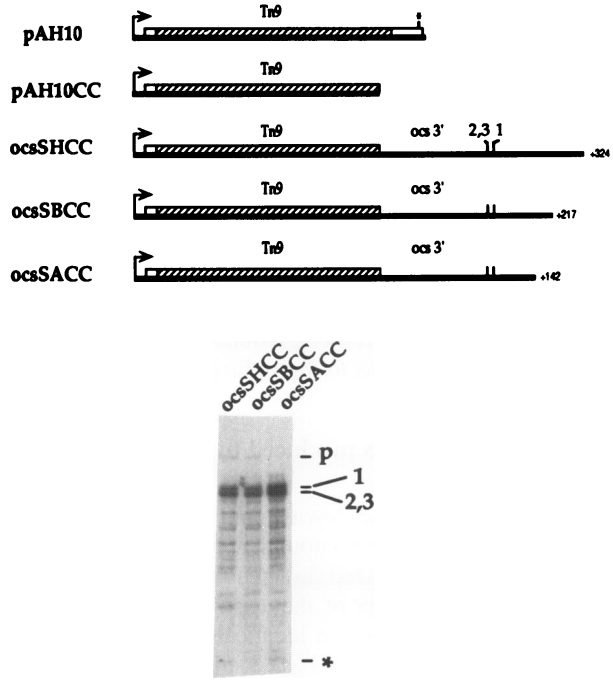

Figure 6. Sequences farther than 142 nt downstream from site 1 are not needed for mRNA 3′ end formation in the *ocs* 3′ region. The structures of the ocsSHCC, ocsSBCC, and ocsSACC constructions are shown, as are the parental constructions used to assemble these test genes. The deletion endpoints are given to the right of each construction. 20 μg of total RNA from pooled populations of transgenic plants carrying each construction was annealed with a probe prepared by labeling the pAH10CC derivative of ocsSH at the NcoI site in the truncated *cat* gene, the hybrids treated with nuclease S1, and the protected fragments separated on a 6% sequencing gel as described in Materials and Methods. The positions of the reannealed probe (p), protected fragments corresponding to sites 1, 2, 3, and of the 3′ ends of the novel RNAs that end in the Tn9 region (*) are shown to the right of each autoradiograph.

Similar RNAs were produced by a mutant that carried bases as far as 142 nucleotides downstream from site 1 (ocsSA in Figure 5). However, when all of the downstream sequences were removed (ocsSS), no such RNAs were seen (ocsSS in Figure 5). Therefore, sequences between −10 and +142 are needed for the production of these novel RNAs. This region is not absolutely necessary for polyadenylation at sites 2 and 3, since a distinct population of such RNAs are seen in the ocsSS construction. However, efficient utilization of these sites probably does require this region since appreciable quantities of readthrough RNAs do arise from the ocsSS construction. (The protected fragment seen with the ocsSS mutant that appears to correspond to site 1 in Figure 5 actually represents the end of homology between the probe and the 3′ region present in the transgenic plants and thus serves as a measure of the levels of 'readthrough' RNAs produced in these plants.)

The 3′ ends of the RNAs seen in the ocsSB and ocsSA constructions map to a position near the 3′ extremity of the Tn9 portion of these chimeric genes. To determine whether the production of these 3′ ends was due to a lack of functioning of the 'normal' *ocs* polyadenylation sites, the portion of the Tn9 region that includes the position of the novel 3′ ends was removed. The resulting constructions (ocsSBCC and ocsSACC) yielded RNAs that end at the same sites as seen in the the wild-type *ocs* construction and the ocsSHCC mutant (Figure 6). In addition, very little readthrough RNA was produced in plants with these constructions. Therefore, *ocs*-derived sequences farther than 142 nt downstream from site 1 are dispensible for efficient polyadenylation at sites 1−3, and the production of the novel 3′ ends in the Tn9 portion of our chimeric genes is not due to a defect in the wild-type *ocs* poly∼A] addition sites.

DISCUSSION

Features of the *ocs* polyadenylation signal

We have identified a number of interesting features of the *ocs* polyadenylation signal. As is the case in two other plant genes (4, 5), upstream sequences rather far from polyadenylation sites are needed for functioning of these sites. We have identified two such FUEs in the *ocs* gene. One of these, which lies between nucleotides −83 and −469, is needed for efficient mRNA 3′ end formation at sites 1, 2, and 3. The other, located between nucleotides −42 and +14, is needed for polyadenylation at sites A, B, and C. This second element lies between 40 and 125 nt upstream from sites A, B, and C (the 3′-most endpoint is about 40 nt from the 5′-most of the cluster of sites, and the 5′-most endpoint some 125 nt from the 3′-most site).

The nearest potential AAUAAA-like element in the FUE for sites 1, 2, and 3 lies some 186 nt upstream from site 1 and is probably not involved in 3′ end formation in this gene. The corresponding element for sites A, B, and C does have an AAUAAA-like element (AAUAAU) between 80 and 85 nt upstream from the 5′-most site, but this is also not needed for functioning of these sites since its removal (in the ocs21/2 construction) has no apparent effect on 3′ end formation in this region (compare the utilization of these sites in the ocs61/42, ocs41/22, and ocs21/2 mutants in Figure 4). Therefore, these two elements are analogous, if not functionally identical, to the upstream sequences identified as important in the CaMV and *rbcS*-E9 polyadenylation signals.

Upstream sequences other than AAUAAA that are involved in mRNA 3′ end formation have now been identified in three different plant genes (4, 5, this study). Sanfaçon *et al.* (4) have found that the sequence UUAGUAUGUAUUUGUAUUUGUA can 'activate' an otherwise unused AAUAAA when placed at an appropriate site upstream from the latter, and have suggested that the motif UAUUUGUA may be involved in the functioning of this element. Related sequences (UAUUUGC, at nts −132 to −138, and UAUUUUUAU, at nts −19 to −27) do occur in the two upstream elements identified in this study. However, mutation of the second of these, which occurs in the FUE for sites A, B, and C, does not noticeably impair the functioning of these sites (ocs21/2 and ocs41/22 in Figure 4). Thus, sequences related to UAUUUGUA cannot be the sole determinant of functioning of this FUE.

Pairwise analyses of the different elements thus far described do yield conserved sequences, but none that are shared by all of the FUEs identified to date (unpublished observation). There are a number of possible explanations for this observation. The various elements thus far described may be functionally analogous but not interchangeable, representing instead different members of a family of polyadenylation-related motifs. Alternatively, they may consist of groups of active sequence elements, much as eucaryotic promoters generally are collections of cis elements whose number and type determine the properties of any given promoter. One such element may consist of sequences related to the motif UAUUUGUA. A third interesting possibility is that there is no highly conserved sequence responsible for the functioning of these elements. Rather, the regions we have mapped may have some less obvious feature, such as a lack of secondary structure, or a particular nucleotide composition (high U or low C, for example). In this respect, these elements would be similar to the sequences downstream from mammalian polyadenylation sites that are needed for proper mRNA 3′ end formation in mammals (14), or to sequences that direct rho-dependent transcription termination in *E.coli* (15, 16).

We have found that upstream sequences nearer to particular polyadenylation sites can also play a role in mRNA 3′ end formation in the *ocs* gene. The ocs61/42 mutation, for example, selectively reduces 3′ end formation at sites 2 and 3 but has little effect on site 1. The ocs41/22 mutation seems to affect just site 3. The ocs21/2 mutation, on the other hand, eliminates usage of site 1 but not sites 2 and 3. There are two interesting explanations for the selective effects of these three linker-substitution mutations. These may define cis elements analogous in function to the mammalian polyadenylation signal AAUAAA. The ocs61/42 and ocs21/2 mutations are appropriately situated for such elements, being between 2 and 30 nt from the sites that each respective mutation affects. Although none of these regions has an AAUAAA motif, each has a related sequence (AAUAAU, 12−17 nt upstream from site 1, AAUAUA, 23−28 nt upstream from site 2, and AAUGAA, 22−27 nt upstream from site 3) whose alteration could explain the observed phenotypes. This would be consistent with the observation that an AAUAAA motif is important for mRNA 3′ end formation in the CaMV 19S/35S transcription unit (3−5) and the finding that AAUAAA-like elements, when present in the 3′-non-coding regions of plant genes, are usually situated between 11 and 40 nt from polyadenylation sites (10).

Alternatively, these mutations may affect 3′ end formation at particular sites by altering the context of the putative cleavage site involved. This is a likely explanation for the effect of the ocs41/22 mutation on site 3; this mutation changes the region from approximately 10 nt upstream to 10 nt downstream from site 3. Along these lines, Joshi (18) has noted that the region between −12 and +12 with respect to plant polyadenylation sites has a high average U content and a distinct preference for a YA dinucleotide at the site of cleavage. The ocs41/22 mutation changes the region surrounding site 3 from UUACUAUCGU to CCCCGCGGGU, thereby decreasing the U content from 50% to 10% and modifying the proposed cleavage site from UA to GC. However, this is a less likely explanation for the effects of the ocs61/42 and ocs21/2 mutations, since neither of these affects the cleavage sites of their respective 'targets' or the U content between −5 and +5 of these sites.

Most plant transcription units yield populations of mRNAs that end at multiple sites (17). The *ocs* gene is typical in this respect in that it gives rise to mRNAs that end at two or three predominant sites. Our studies indicate that multiple, and partially overlapping, upstream sequence elements are responsible for 3′ end formation at these various sites. As far as we can ascertain, a single FUE is responsible for polyadenylation at sites 1, 2, and 3, although our set of deletions may not have resolved separate elements. This element is clearly distinct from the corresponding element for sites A, B, and C. In addition to this complexity, it is probable that each site at which mRNAs end in the *ocs* gene is determined by separate near-upstream signals. Interestingly, some of the latter may overlap significantly with the FUE for sites A, B, and C.

The significance of this arrangement of polyadenylation-related signals in the *ocs* gene is not clear at this time. Selective removal of a subset of these does not seem to affect the total levels of mRNAs produced, or of their translatability, since all of the constructions described here yield similar levels of CAT activity in transgenic plants (data not shown). Therefore, overall efficiency of mRNA formation or translation does not seem to require this arrangement. It is possible that the individual signals may be utilized with different efficiencies during development or under differing environmental conditions. The functional redundancy seen in polyadenylation signals could be a mechanism to ensure efficient mRNA production over a broad range of growth conditions. In this respect, redundancy in polyadenylation signals would be analogous to the redundancy seen in promoter elements.

The nature, of the RNAs produced by the ocsSB and ocsSA constructions

Our attempts to dissect the downstream sequence requirements for polyadenylation were confounded by the observation that some of the mutants analyzed directed the production of RNAs that end in the Tn9 region of the appropriate chimeric genes. These RNAs end at a position identical to non-polyadenylated RNAs produced by certain *cat-rbcS* 3′ mutant constructions (5). We have found that the production of these novel RNAs is controlled by positive and negative elements; these elements may be located downstream from normally-utilized polyadenylation sites (as we have described here) or upstream from polyadenylation sites, as has been reported in a pea *rbcS* gene (5).

The means by which these RNAs are made are not known, but their properties suggest some interesting possibilities. These RNAs might represent the products of alternative processing at sites in the Tn9 sequences. This alternative processing would be negatively controlled by *ocs* sequences located between 217 and 324 bp downstream from site 2 (these are the endpoints of the

ocsSB and ocsSH mutants). In addition, this processing would require sequences between −10 and +142 (the endpoints of the ocsSS and ocsAS deletions). Although we cannot at this time say what the nature of the alternative processing might be, it may not be related to normal 3′ end processing since these 'aberrant' RNAs are largely or completely non-polyadenylated, based on oligo-dT chromatography (5, M.H. MacDonald and A.G. Hunt, unpublished observations).

Another interesting possibility is that these novel RNAs might be the products of premature transcription termination events. In this case, the *ocs* sequences located between +217 and +324 bp would define a sort of transcriptional antiterminator. When removed, transcription might be unable to proceed through certain types of strong pause sites, of which the Tn9 site mapped here might be one. In the presence of the −10 to +142 region, which would play a positive role in transcription termination, the paused RNA polymerase would cease transcription and release the nascent RNA. When this latter element, or the strong pause site, is removed, then transcription might again be able to proceed through the region that is needed for normal mRNA 3′ end formation. This model implies a close spatial and temporal linkage between mRNA 3′ end formation and transcription termination in plants. However, the behavior of the ocsSS deletion indicates that this linkage is not essential for mRNA polyadenylation. This model, although highly speculative, is similar to suggestions that have been made concerning the relationship between mRNA polyadenylation and transcription termination in yeast (19). Since at least one plant polyadenylation signal can apparently function in yeast (20), it is tempting to speculate that the process of mRNA 3′ end formation in plants may be remarkably similar to the same process in yeast.

ACKNOWLEDGEMENTS

We thank Rebecca Richardson and Molly Goodenow for excellent technical assistance, and Brian Rymond for helpful suggestions. This work was supported by USDA Competitive Grants 85-CRCR-1-1810 and 89-37262-4835.

REFERENCES

1. Messing,J., Geraghty,D., Heidecker,H., Hu,N.T., Kridl,J. and Rubenstein,I. (1983) In T.Kosuge, P.Meredith and A.Hollaender (eds.) *Genetic Engineering of Plants: An Agricultural Perspective*. Plenum Press, New York, pp. 211−227.
2. Dhaese,P., DeGreve,H., Gielen,J., Seurinck,J., Van Montagu,M. and Schell,J. (1983) *EMBO J.* **2**, 419−426.
3. Sanfaçon,H. and Hohn,T. (1990) *Nature* **346**, 81−84.
4. Sanfaçon,H., Brodmann,P. and Hohn,T. (1991) *Genes Develop.* **5**, 141−149.
5. Mogen,B.D., MacDonald,M.H., Graybosch,R. and Hunt,A.G. (1990) *The Plant Cell* **2**, 1261−1272.
6. Hunt,A.G., Chu,N.M., Odell,J.T., Nagy,F. and Chua,N.-H. (1987) *Plant Molec. Biol.* **7**, 23−35.
7. Carswell,S. and Alwine,J.C. (1989) *Mol. Cell. Biol.* **9**, 4248−4258.
8. Russnak,R. and Ganem,D. (1990) *Genes Develop.* **4**, 764−776.
9. DeZazzo,J.-D. and Imperiale,M.J. (1989) *Mol. Cell. Biol.* **9**, 4951−4961.
10. Hunt,A.G. and MacDonald,M. (1989) *Plant Mol. Biol.* **13**, 125−138.
11. Kunkel,T.A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 488−492.
12. Schardl,C., Byrd,A.D., Benzion,G., Altschuler,M., Hildebrand,D.F. and Hunt,A.G. (1987) *Gene* **61**, 1−11.
13. Hunt,A.G. (1988) *DNA* **7**, 329−336.
14. Birnsteil,M.L., Busslinger,M. and Strub,K. (1985) *Cell* **41**, 349−359.
15. Alifano,P., Rivellini,F., Limauro,D., Bruni,C.B. and Carlomagno,M.S. (1991) *Cell* **64**, 553−563.
16. Platt,T. (1986) *Ann. Rev. Biochem.* **55**, 339−372.
17. Dean,C., Tamaki,S., Dunsmuir,P., Favreau,M., Katayama,C., Dooner,H. and Bedbrook,J. (1986) *Nucl. Acids Res.* **14**, 2229−2240.
18. Joshi,C.P. (1987) *Nucl. Acids Res.* **15**, 9627−9640.
19. Russo,P. and Sherman,F. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8348−8352.
20. Hirt,H., Kögl,M., Murbacher,T. and Heberle-Bors,E. (1990) *Curr. Genet.* **17**, 473−479.