

# Different 3' End Regions Strongly Influence the Level of Gene Expression in Plant Cells

Ivan L.W. Ingelbrecht, Lieve M.F. Herman, Rudy A. Dekeyser, Marc C. Van Montagu,<sup>1</sup> and Ann G. Depicker

Laboratorium voor Genetica, Rijksuniversiteit Gent, B-9000 Gent, Belgium

**We have investigated the functional role of a 3' end region on the expression of a reporter gene in plant cells. In stably transformed plants, expression of the reporter gene without a plant gene 3' end is variable and depends on the fortuitous presence of polyadenylation signals in the downstream sequences. When the reporter gene is flanked by pBR322 DNA, 3'-processing and polyadenylation occurs at (a) cryptic site(s) within these vector sequences. Using a transient gene expression system, we present a deletion analysis of the 3' end of the octopine synthase gene showing that the most proximal polyadenylation signal per se is not sufficient to ensure expression but that a downstream (G)T-rich sequence is also required. Optimal expression of the fusion gene requires more than 98 base pairs and at most 142 base pairs downstream from the most distal polyadenylation site. We analyzed the expression of chimeric genes with 3' end sequences originating from different plant genes. In the transient expression assay, all constructs direct similar neomycin phosphotransferase II activities. However, in stably transformed tissue, the gene constructs displayed characteristic expression levels which varied as much as 60-fold. This result suggests a role for 3' end sequences in post-transcriptional processes such as efficiency of 3'-processing and/or mRNA stability.**

## INTRODUCTION

In eukaryotic cells, biogenesis of a functional mRNA requires complex post-transcriptional modifications including 5'-capping, intron splicing, and 3' end-processing. In animal cells, formation of most mRNA 3' ends involves endonucleolytic cleavage of larger precursors followed by the addition of up to 250 adenylate residues [the poly(A) tract] (for review, see Birnstiel, Buslinger, and Strub, 1985; Proudfoot and Whitelaw, 1987). Extensive deletion analysis and site-specific mutagenesis have defined two *cis*-acting elements that are both essential for accurate and efficient 3' end formation: the highly conserved AAUAAA, located 10 nucleotides to 30 nucleotides upstream from the poly(A) addition site (Proudfoot and Brownlee, 1976; Fitzgerald and Shenk, 1981; Montell et al., 1983; Wickens and Stephenson, 1984) and the less conserved (G)T-rich sequences immediately 3' to the cleavage site (Gil and Proudfoot, 1984; Woychik et al., 1984; Hart, McDevitt, and Nevins, 1985; McLauchlan et al., 1985; McDevitt et al., 1986; Gil and Proudfoot, 1987). In addition to these sequences, the 3'-untranslated region can also contain regulatory sequences that affect the mRNA stability. For example, a family of mammalian mRNAs encoding growth regulatory functions contains a conserved AT-rich element in their 3'-noncoding region

that mediates the selective degradation of these mRNAs (Caput et al., 1986; Shaw and Kamen, 1986; Brewer and Ross, 1988). Similarly, sequences in the trailer of the *Drosophila hsp70* mRNA are important in directing its turnover (Simcox et al., 1985; Petersen and Lindquist, 1988). Together, these data indicate that 3' sequences can play an important role in the establishment of mRNA prevalence levels in animal cells.

In contrast, little information is available about the sequences governing 3'-processing and polyadenylation of plant pre-mRNAs. Compilation analyses indicate that an AATAAA-like sequence is conserved among nuclear plant gene 3'-untranslated sequences, implying a common function for this sequence in both plant and animal cells (Joshi, 1987; Elliston and Messing, 1988). Although a majority of plant genes do not contain a perfect AATAAA signal, this plant polyadenylation signal rarely deviates from the animal sequence in more than one position, most of the bias being restricted to position 6. Recently, Hunt et al. (1987) reported that tobacco cells do not recognize the polyadenylation signals of a human and two animal virus genes as animal cells do. Apparently, there must be significant differences regarding either the sequence requirements for 3'-processing and polyadenylation or the utilization of these processing signals in animal and plant cells.

In this paper we have investigated the role of a 3' end

<sup>1</sup> To whom correspondence should be addressed.

region on expression of a chimeric gene in plant cells. Using a transient gene expression system, we found that optimal expression of a chimeric gene with the 3' end of the octopine synthase gene requires both the polyadenylation signal and a (G)T-rich sequence motif present downstream from the polyadenylation site. Without 3' end sequences, mRNA production in transgenic plants is dependent on the presence of fortuitous polyadenylation signals within the 3'-flanking DNA. By analyzing the expression of chimeric genes with different plant 3' gene regions, we show that expression levels vary as much as 60-fold among the various constructs. We conclude that post-transcriptional processes, associated with 3' end formation, are important in determining mRNA prevalences.

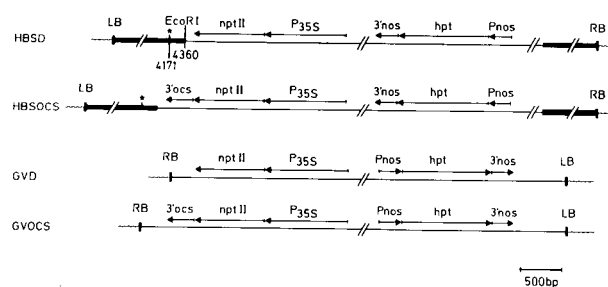
## RESULTS

### Importance of 3'-Processing Signals for Efficient Expression of a Gene

To evaluate the influence of 3' end sequences on gene expression in plant cells, chimeric genes were constructed, based on the neomycin phosphotransferase II (*nptII*) coding sequence (Herrera-Estrella et al., 1983) fused to the cauliflower mosaic virus 35S promoter (Odell, Nagy, and Chua, 1985). The plasmids pHBSOCS and pHBSD contain a *P35S-nptII* chimeric gene with and without the 3' end of the octopine synthase gene (3' *ocs*) (De Greve et al., 1982), respectively. Both plasmids also contain a hygromycin resistance gene under the control of the nopaline synthase promoter as an independent selection marker (van den Elzen et al., 1985). Using the pGV3850 *Agrobacterium* transfer system (Zambryski et al., 1983), these chimeric genes were introduced into the tobacco genome. Figure 1 shows the structure of the two T-DNAs after integration in the plant genome.

Hygromycin-resistant calli were selected and, after regeneration of plants, a neomycin phosphotransferase II (NPTII) enzymatic assay was performed on leaf extracts of independent transformants. Transformed plants containing the pHBSOCS T-DNA (the HBSOCS plants) show a 12-fold higher amount of NPTII activity than that observed in the HBSD plants (data not shown). For each construct, six transformants were assayed and, in all cases, the variation in expression level between independent transformants was less than threefold.

To verify whether the different NPTII activities reflect different steady-state mRNA levels, RNA gel blot analysis was performed using RNA isolated from HBSOCS and HBSD leaf tissue. Figure 2A demonstrates that the *nptII* transcripts are less abundant in HBSD plants than in HBSOCS plants. Furthermore, despite the fact that the *nptII* gene in HBSD plants does not have a 3' end of a plant gene, analysis of poly(A)<sup>+</sup> RNA and poly(A)<sup>-</sup> RNA

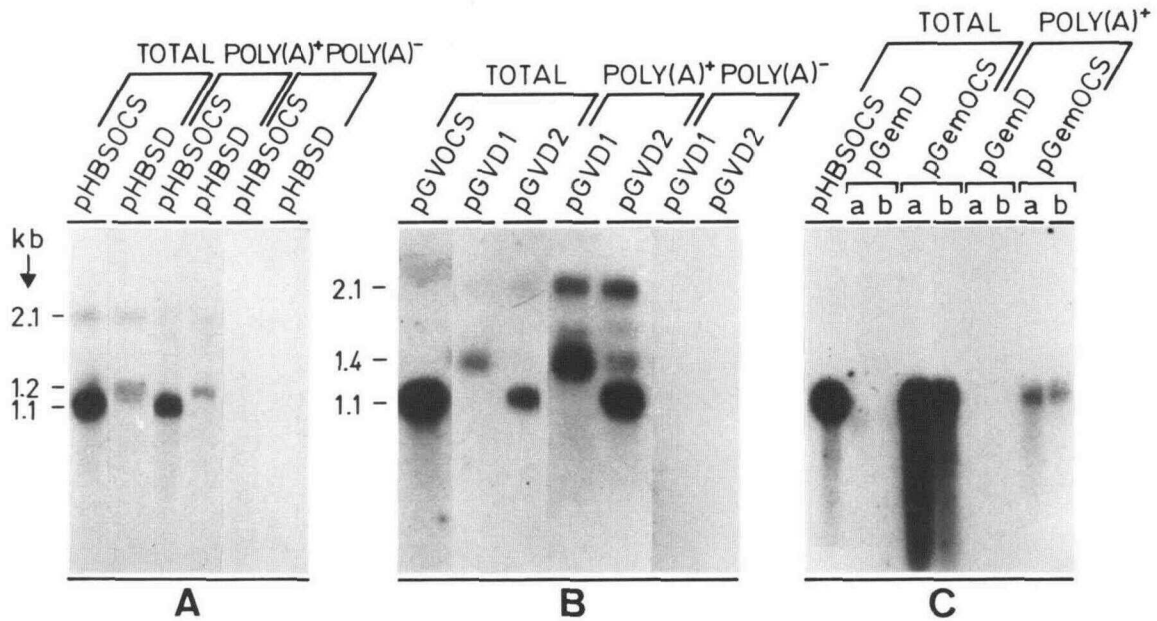


**Figure 1.** T-DNA Structure as Found in HBSD, HBSOCS, GVD, and GVOCS Transgenic Plants.

The position and orientation of the chimeric *nptII* and *hpt* genes are indicated with respect to the T-DNA borders. RB and LB designate the right and left borders, respectively; the solid lines represent pBR322 and T-DNA sequences derived from the pGV3850 vector; the wavy lines symbolize plant DNA; the asterisk at position 4171 in pBR322 (standard numbering) indicates the putative polyadenylation signal AATAAT.

demonstrates that the *nptII* transcripts are only found at detectable levels in the poly(A)<sup>+</sup> fraction. In addition, these *nptII* transcripts have a defined length, indicating that they are polyadenylated at (a) specific site(s) in the downstream sequences. As predicted from the mapped transcription start and polyadenylation site, the *nptII* transcripts in HBSOCS plants are 1050 bases long, whereas the *nptII* transcripts in HBSD plants are approximately 1150 bases long, as estimated from comparison with RNA size markers (data not shown). The length of these latter transcripts implies that polyadenylation occurs in the flanking downstream pBR322 sequences around position 4145 (standard numbering). Inspection of the surrounding pBR322 sequences reveals that an AATAAT signal is present at position 4172, which is 27 bp upstream from position 4145. Notably, Zhang, Denome, and Cole (1986) reported that polyadenylation can occur in pBR322 DNA around position 4145 with an efficiency of 5% to 8%. So, although we have not mapped precisely the polyadenylation site(s), it is very likely that the same signal is also functional in plants and that this polyadenylation event directs the low expression level of the *nptII* gene in all HBSD plants.

To determine whether gene expression can be detected when this pBR322 sequence is removed, the same *nptII* chimeric genes were introduced into plant cells via the pGV1503 T-DNA vector (see Methods). After transfer and integration of the T-DNA, the *nptII* stop codon of the chimeric gene without 3' end is located 360 bp from the right T-DNA border, as illustrated in Figure 1. Hygromycin-resistant calli were selected, and, after plant regeneration, leaf tissue was assayed for NPTII activity. As expected, the NPTII activity detected in GVOCS plants was very similar to the activity found in the HBSOCS plants (data



**Figure 2.** RNA Gel Blot Analysis of Transcripts from the *P35S-nptII* Gene with and without the 3' *ocs* Fragment in Stably Transformed Plants and after Transient Expression.

Hybridizations were carried out using the *nptII* coding region as a probe. Blots containing total RNA (A, B) and poly(A)<sup>+</sup> RNA (B) were simultaneously reacted with a probe representing the  $\beta$ -subunit mitochondrial ATP synthase gene as an internal standard; the hybridizing transcript is 2.1 kb in length.

(A) RNA gel blot containing total, poly(A)<sup>+</sup>, and poly(A)<sup>-</sup> RNA from leaf tissue transformed with the T-DNA of the plasmids pHBSOCS and pHBSD as indicated above the individual lanes. The *nptII* transcripts are 1.1 kb and 1.2 kb in length in HBSOCS and HBSD plants, respectively.

(B) RNA gel blot containing total, poly(A)<sup>+</sup>, and poly(A)<sup>-</sup> RNA from leaf tissue transformed with the T-DNA of the plasmids pGVOCS and pGVD as indicated above each lane. The hybridizing *nptII* transcripts are 1.1 kb long in GVOCS and GVD2 plants and 1.4 kb long in the GVD1 plant.

(C) RNA gel blot containing total and poly(A)<sup>+</sup> RNA isolated from tobacco protoplasts 1 hr (a) and 4 hr (b) after electroporation with the indicated plasmids. pHBSOCS control lane contains 5  $\mu$ g of total RNA from an HBSOCS transgenic plant.

not shown). However, of the 10 GVD transgenic plants analyzed, seven did not contain detectable NPTII activity, whereas three plants showed variable levels of expression. To determine whether the NPTII activity in two of these GVD plants was correlated with the presence of a defined transcript, RNA gel blot analysis was performed. Figure 2B demonstrates that the transcripts hybridizing with an *nptII* probe are less abundant in the GVD plants than in the GVOCS plant. More important, the *nptII* transcripts are detected only in the poly(A)<sup>+</sup> fractions and have clearly different lengths. These results show that the plant DNA downstream from the *nptII* gene must contain cryptic polyadenylation signals that permit mRNA 3'-processing and/or stabilization with reduced efficiency.

Because of this inherent background of gene expression associated with chimeric gene integration, we tested whether a transient gene expression assay system could be used to identify 3' sequences necessary for efficient 3'-processing and polyadenylation of the chimeric *nptII*

transcript. We introduced the plasmids pGemOCS and pGemD, which contain a chimeric *nptII* gene with and without the 3' *ocs*, respectively, into tobacco protoplasts by electroporation and then assayed protoplast extracts for both NPTII activity and *nptII* mRNA. Figure 2C shows that a polyadenylated *nptII* transcript was detected in pGemOCS-containing protoplast extracts that was equal in length to that present in stably transformed leaf tissue. In contrast, no *nptII* transcript was detected in either the total or poly(A)<sup>+</sup> RNA isolated from protoplasts electroporated with pGemD (Figure 2C). Similarly, the NPTII enzymatic activity level in pGemOCS-containing cells was at least 20-fold higher than in cells containing the pGemD plasmid (data not shown).

Together these data show that DNA sequences contained within the 3' *ocs* region are essential for efficient expression of the chimeric *nptII* gene in stably transformed plants. When a 3' end is absent, expression of the chimeric *nptII* gene is dependent on the presence of cryptic poly-

adenylation signals in the downstream DNA region. Moreover, 3' end sequences are also required for the expression of genes within isolated tobacco protoplasts after transient expression of electroporated plasmids.

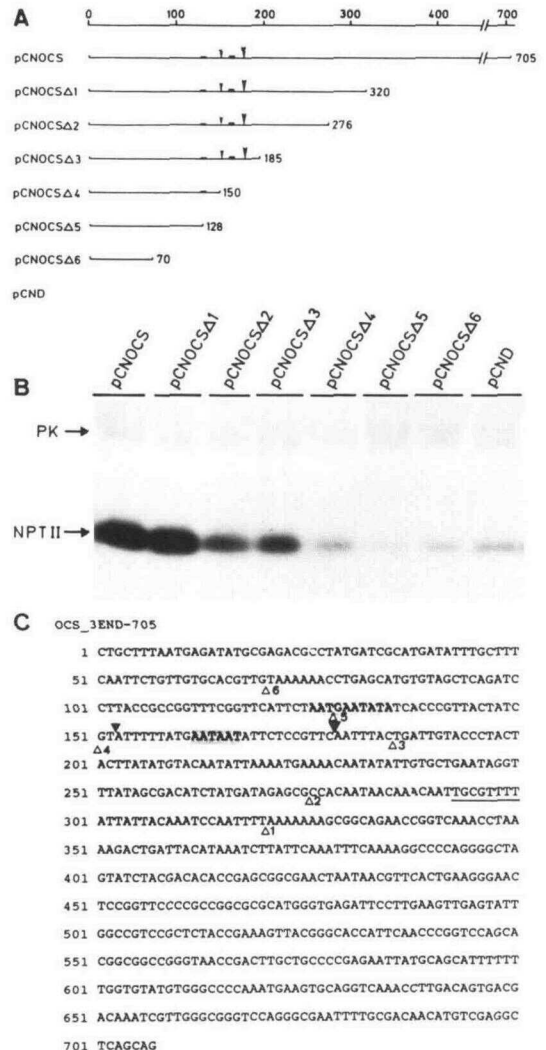
### Deletion Analysis of the 3' End of the Octopine Synthase Gene via Transient Gene Expression

To begin to localize the *cis* sequences present within the 3' *ocs* region that are necessary for optimal *nptII* fusion gene expression levels, we constructed a series of 3' deletions and then introduced these constructs into tobacco protoplasts by electroporation. As an internal control, we included a chimeric chloramphenicol acetyltransferase (*cat*) gene within each construct to reduce experimental errors due to differential uptake of the exogenously added DNA by the protoplasts. For each construction, we first measured a fraction of the protoplast extract for chloramphenicol acetyltransferase (CAT) activity using a quantitative "diffusion" assay (Neumann, Morency, and Russian, 1987) and then determined NPTII enzymatic levels in fractions containing identical amounts of CAT activity. Figure 3A shows schematically the intact 3' *ocs* fragment present in plasmid pCNOCS and the various 3' deletions present in plasmids pCNOCS $\Delta$ 1 through pCNOCS $\Delta$ 6. As illustrated in Figure 3A and shown in Figure 3C, the 3' *ocs* fragment contains two polyadenylation sites at positions 178 and 152, designated the major and minor sites, respectively (Dhaese et al., 1983). As shown in Figure 3B and summarized in Figure 3C, NPTII activity levels were not affected by deletion of the distal 385 bp (pCNOCS $\Delta$ 1 lane). In contrast, NPTII levels decreased fourfold upon deletion of sequences between 98 bp and 142 bp downstream from the major polyadenylation site (pCNOCS $\Delta$ 2 lane) and then dropped to background levels only after deletion of the 35-bp region containing the "putative" major polyadenylation signal and both minor and major sites (compare pCNOCS $\Delta$ 4 and pCND). RNA gel blot analysis showed similar reductions in relative *nptII* mRNA levels (data not shown).

Together, these data indicate that maximal *nptII* expression levels require sequences contained within a region extending 144 bp downstream from the major polyadenylation site but that the most crucial signals are found within a 35-bp fragment that contains the major polyadenylation signal and the minor and major polyadenylation sites.

### Modulation of the *nptII* Gene Expression Levels by Different Plant 3' Gene Sequences

In the previous experiments, we demonstrated that cryptic polyadenylation signals can be used to produce variable protein and steady-state mRNA levels. To find out whether different plant gene 3' regions similarly can affect mRNA



**Figure 3.** Deletion Analysis of the 3' *ocs* Region.

**(A)** Schematic representation of the different deletions. The plasmids containing the different 3' *ocs* deletions are indicated on the left. The numbers at the endpoints of the lines define the last nucleotide that is still present in the deletion derivatives [see also **(C)**]. The putative minor and major polyadenylation signals, as defined by Dhaese et al. (1983), are indicated by a small and large solid box, respectively; the minor and major poly(A) addition sites are pointed out by a small and a large arrowhead, respectively. **(B)** NPTII enzymatic assay of extracts from tobacco protoplasts electroporated with the plasmids indicated above the individual lanes. The different samples were standardized on CAT activity (see Methods). PK, plant kinase; NPTII, neomycin phosphotransferase II. **(C)** DNA sequence of the 3' *ocs* fragment in pCNOCS. Deletion endpoints are indicated below the DNA sequence. The polyadenylation signals are indicated by bold lettering, and the minor and major polyadenylation sites are indicated above the DNA sequence with a large and small arrowhead, respectively. The (G)T-rich sequence is underlined.

prevalence levels, we constructed a series of chimeric *nptII* genes containing different 3' gene sequences. The various 3' gene regions used in this study are listed in Table 1 and include DNA sequences from genes encoding (1) a 2S seed storage protein, (2) a small subunit of ribulose-1,5-bisphosphate carboxylase, (3) extensin, and (4) a chalcone synthase. These genes display a characteristic expression program showing either a marked tissue specificity—for example, the 2S and *rbcS* small subunit genes—or clear physiological inducibility, as in case of the chalcone synthase gene.

To assess the functionality of each 3' gene region, we first introduced each construct into isolated tobacco protoplasts by electroporation and determined NPTII activity levels in protoplast extracts. As shown in Figure 4A, all *nptII* chimeric genes containing plant 3' regions directed the production of significantly higher NPTII activity levels (lanes pGemCHS, pGemOCS, pGemSS, pGem2S, and pGemEXT) than that observed for a gene without 3' end sequences (lane pGemD). In addition, NPTII activity levels were virtually identical for all gene constructs analyzed. These results strongly suggest that all signals required for efficient 3'-processing of the *nptII* transcripts are present within the cloned 3' end sequences.

We then introduced each gene construct into the tobacco genome via the T-DNA vector pGV1503 (see Methods) and determined NPTII activity and *nptII* mRNA levels in leaves of stably transformed tobacco plants. To eliminate variation due to position effects, we analyzed *nptII* gene expression levels in at least 10 independently transformed plants for each construct. As shown in Figure 4B and summarized in Table 2, each *nptII* fusion gene displayed a characteristic expression level in tobacco plants. If we normalize the NPTII activity found in the GVOCS plants to one, then, on average, the NPTII activity found in GVSS plants is 3 times higher. On the other hand, GV2S, GVEXT, and GVCHS plants contain a fivefold, 10-fold, and 20-fold lower NPTII activity than that found in GVOCS plants. These results indicate that NPTII enzymatic activity levels vary as much as 60-fold, depending upon the origin of the 3' gene sequences.

To determine whether the *nptII* gene expression in GVSS plants shows an organ specificity similar to that of the small subunit gene, we measured NPTII activity in callus and roots as compared with leaves. We found equally high levels of NPTII enzymatic activity in the different extracts analyzed (data not shown). Therefore, we conclude that the 3' end of the small subunit gene does not impose a differential processing or turnover rate on the *nptII* transcript in a tissue-specific manner.

To determine whether the observed NPTII activity actually reflects the steady-state mRNA content, total RNA was isolated from two representative transgenic plants for each construction. The RNA gel blot represented in Figure 4C indicates that, for each construct, the *nptII* transcript length corresponds to that predicted on the basis of the available transcriptional start and polyadenylation sites. This result confirms that the polyadenylation sites of the original genes are utilized. More important, Figure 4C also demonstrates that the steady-state mRNA content is different for the various constructs corresponding well with the NPTII enzymatic activity levels. Therefore, we can conclude that the NPTII activity as measured in at least 10 plants for each construct is a mere reflection of the mRNA content and that no unexpected translational effects are involved in establishing the NPTII activity.

## DISCUSSION

### Functional 3' End Regions Are Necessary for Efficient Gene Expression

This report focuses on the functional roles of 3' gene sequences for expression of a chimeric gene in tobacco cells.

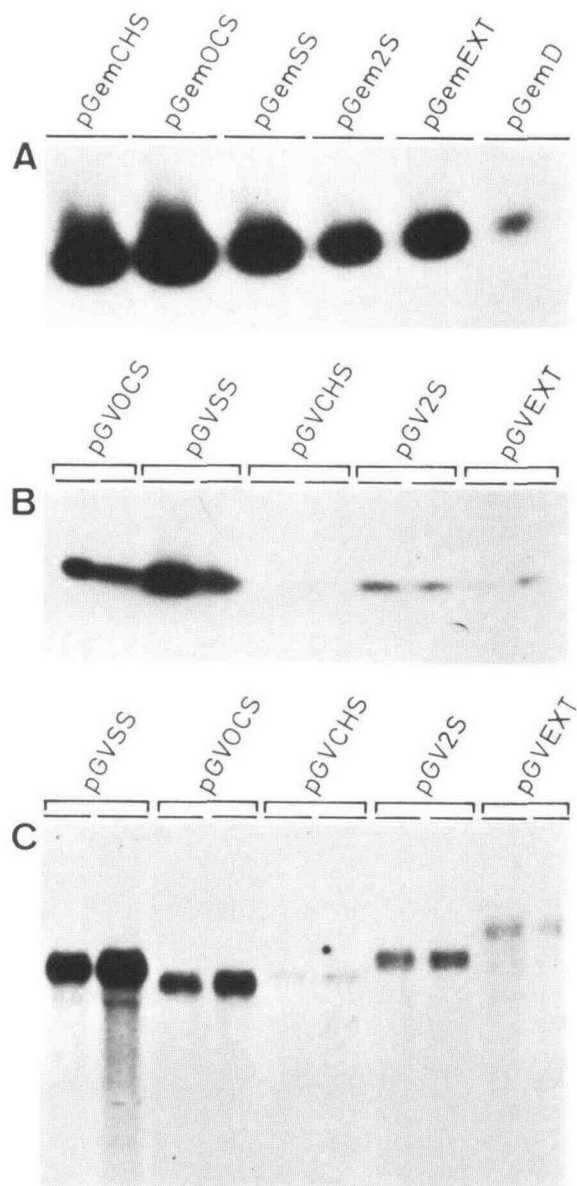
In a first experiment, we demonstrate that the expression level of an *nptII* reporter gene with the 3' end sequences of the octopine synthase gene is uniformly high, whereas expression of the same gene without 3' end sequences depends on the presence of cryptic polyade-

**Table 1.** Description of the Cloned 3' End Regions

Gene	Origin	Length of the Cloned Fragment	Length of the Noncoding Segment Relative to the Poly(A) Addition Site		References
			Upstream	Downstream	
Octopine Synthase	<i>Agrobacterium tumefaciens</i>	320	178	142	De Greve et al. (1982)
2S Seed Protein	<i>Arabidopsis thaliana</i>	±650	ND <sup>a</sup>	ND	Krebbers et al. (1988a)
Small Subunit of <i>rbcS</i>	<i>Arabidopsis thaliana</i>	±600	ND	ND	Krebbers et al. (1988b)
Extensin	<i>Daucus carota</i>	735 <sup>b</sup>	450	106	Chen and Varner (1985)
Chalcone Synthase	<i>Antirrhinum majus</i>	287	170	113	Sommer and Saedler (1986)

<sup>a</sup> ND, not determined.

<sup>b</sup> This 3' fragment contains a 185-bp intron.



**Figure 4.** Comparison of the Expression Level of a Chimeric *nptII* Gene with Different 3' Ends after Transient Expression and in Transgenic Leaf Tissue.

(A) NPTII enzymatic assay of extracts from tobacco protoplasts electroporated with the plasmids indicated above the individual lanes. The samples were standardized on total protein content (see Methods).

(B) NPTII enzymatic assay of extracts from transgenic leaf tissue, transformed with the T-DNA of the plasmids indicated above the individual lanes. The samples were standardized on total protein content (see Methods).

(C) RNA gel blot with 5 µg of total RNA from transgenic leaf tissue, transformed with the T-DNA of the various plasmids indicated above the lanes. Hybridization was carried out using an *nptII* riboprobe.

nylation signals in the downstream DNA. These fortuitous polyadenylation events result in variable, usually low expression levels of the reporter gene. For example, if the *nptII* gene is followed by pBR322 DNA, *nptII* mRNA production and NPTII activity is approximately 12-fold lower than the same *nptII* gene with the 3' *ocs* fragment. Since this *nptII* transcript is polyadenylated and has a defined length, this low-level expression is due to 3'-processing and polyadenylation events centered around a fortuitous polyadenylation signal within the pBR sequences. In fact, an AATAAT signal followed by two sequence motifs similar to the YGTGTTY consensus are present within the pBR322 region corresponding to the 3' terminus of the transcript. Also, in animal systems, polyadenylation of extended transcripts at three distinct sites in pBR322 has been reported, and one of them is located in the region characterized in this study (Zhang, Denome, and Cole, 1986; Kessler et al., 1987). Polyadenylation at these cryptic sites was found to be inefficient, consistent with our results.

The fact that the same cryptic polyadenylation signal in pBR322 is used in plant and animal cells suggests not only that the structural requirements for polyadenylation are highly flexible, but also that the sequences responsible for 3' end-processing are similar in plants and animals.

#### The Polyadenylation Signal and a YGTGTTY-like Sequence Constitute Important Sequence Elements in the 3' *ocs* Region

We show that, after electroporation of the *P35S-nptII-3'ocs* fusion gene, a polyadenylated transcript is produced

**Table 2.** NPTII Activities in Plants Transformed with Chimeric *nptII* Genes Containing Different Plant Gene 3' Ends

Experiment	GVOCS	GVSS	GV2S	GVEXT	GVCHS
	( $\times 10^{-3}$ cpm)				
1	1 (8.7)	3	1/5	1/4	1/15
2	1 (18.2)	3	1/5	ND <sup>a</sup>	1/15
3	1 (17.8)	3	1/5	1/16	1/20
4	1 (11.0)	5	1/7	1/10	1/15
5	1 (13.0)	1	1/12	1/18	1/23
6	1 (5.8)	5	1/4	1/9	1/30
7	1 (3.1)	2	1/2	1/6	1/30
8	1 (1.6)	3	1/2	1/6	1/30
Average	1	3	1/5	1/10	1/22

<sup>a</sup> ND, not determined.

Identical amounts of protein were loaded in each experiment. Quantification of the NPTII activity was performed by counting the phosphorylated kanamycin bound onto the P81 paper (see Methods). The absolute counts per minute (cpm) value from the GVOCS plants, which is indicated in parentheses, has been standardized to 1 in all experiments. The counts per minute values of the other constructions are expressed relative to this standardized GVOCS value.

equal in length to the *nptII* mRNA present in stably transformed tissue. In contrast, no *nptII* transcript can be detected, and NPTII enzymatic activity levels are at least 20-fold lower after electroporation of the *P35S-nptII* fusion gene without functional 3' end sequences. Therefore, this transient expression system provides a reliable assay for the characterization of the 3' *ocs* sequences that are required for efficient expression of the gene.

The octopine synthase 3' region contains a major and a minor polyadenylation site, and the corresponding polyadenylation signals are accordingly called the major and minor polyadenylation signals (De Greve et al., 1982; Dhaese et al., 1983). Analysis of plasmids containing progressive deletions in the 3' *ocs* region indicates that the NPTII activity decreases stepwise.

When the region between 98 bp and 142 bp downstream from the major polyadenylation site is deleted, NPTII activity is reduced by a factor of 4. Since this region contains a 1-base variation of the consensus signal YGTGTTY, it is likely that this sequence is important for efficient processing of the transcript. A majority of the mammalian genes contain the YGTGTTY sequence at approximately 30 bp downstream from the polyadenylation signal, and disruption of this spatial relationship reduces the efficiency of 3'-processing. YGTGTTY-like sequences also occur in the 3' region of plant genes, but apparently not at a fixed position relative to the polyadenylation signal (Joshi, 1987). In the 3' *ocs* region, the YGTGTTY-like motif is situated some 140 bp downstream from the major polyadenylation signal, whereas, for the potato proteinase inhibitor gene, a similar sequence, situated 46 bp downstream from the polyadenylation signal, has been shown to be important for 3' end formation (An et al., 1989).

After deletion of the 35-bp region containing the major polyadenylation signal and the major and minor polyadenylation sites, the NPTII activity decreases to background levels. Therefore, this 35-bp region must contain sequences that are required for maturation of the *nptII* transcripts. Moreover, we can conclude that a polyadenylation signal, as such, is not sufficient to ensure expression, a fact that also holds true for animal genes.

### Expression Level of an *nptII* Reporter Gene Is Strongly Modulated by Different 3' Ends

We demonstrate that the expression level of *nptII* chimeric genes in transgenic tobacco plants varies as much as 60-fold, depending on the origin of the 3' end. The different 3' ends included in this comparative study were derived from genes with entirely different expression characteristics. The genes encoding extensin and chalcone synthase are both inducible by environmental stimuli. The octopine synthase gene is believed to be constitutive, whereas the 2S seed storage protein and small subunit genes are expressed primarily in seeds and leaves, respectively.

Although none of the 3' ends originates from tobacco genes, the resulting chimeric genes were all equally well expressed in a tobacco protoplast transient expression system, indicating that all signals required for proper 3' end formation are present. This result contrasts with the NPTII activity found in stably transformed leaf tissue. Here, we observed pronounced different expression levels of the various fusion genes. Expression of the *nptII* fusion gene with the 3' end of a ribulose-1,5-bisphosphate carboxylase small subunit gene results in approximately 3 times more NPTII enzymatic activity than that with the 3' end of octopine synthase. This latter gene is expressed 5 and 10 times higher than the fusion with the 3' end of the 2S and extensin genes, respectively. Expression of the *nptII* gene containing the 3' end of the chalcone synthase gene is uniformly low, on average 60 times less than the *nptII* gene with the 3' end of the small subunit gene. The steady-state mRNA levels correspond well with the results obtained from the NPTII assays, thus excluding unexpected translational differences between the chimeric mRNAs. The difference in steady-state mRNA levels suggests that post-transcriptional events such as efficiency of 3'-processing and polyadenylation or transcript stability differ for each 3' end used. It is not unlikely that the observed difference would, at least in some cases, be due to a different stability of the transcript. The small subunit gene is expressed mainly in chloroplast-containing tissues where its mRNA constitutes a major fraction (1% to 2%) of the poly(A)<sup>+</sup> RNA. Its gene product is a subunit of the chloroplast enzyme ribulose-1,5-bisphosphate carboxylase, which accounts for up to 50% of the total organellar protein. The obvious requirement for large amounts of the protein can be met not only by providing the gene with a strong promoter, which is in fact the case, but also by rendering its transcript stable. In a similar way, the low-level expression of the fusion gene with the 3' end of the chalcone synthase could be explained. Since this enzyme disappears quickly after induction by external stimuli such as UV illumination (Lipphardt et al., 1988), it is attractive to speculate that its mRNA might be unstable. Rapid turnover of these mRNAs would indeed facilitate their rapid removal in response to environmental stimuli. It will now be interesting to find out whether the different mRNA prevalences are due to differential processing, transport, or turnover of the transcripts.

## METHODS

### Restriction Analysis and Cloning Techniques

Restriction endonucleases, calf intestine phosphatase, the large fragment of DNA polymerase I, Bal31, phage T4 ligase, and phage T4 polynucleotide kinase were purchased from Boehringer (Mannheim, Federal Republic of Germany), Bethesda Research Labo-

ratories (N.V. Gibco, Belgium), New England BioLabs, or Pharmacia (Uppsala, Sweden) and were used according to the suppliers' instructions. All recombinant DNA techniques were performed according to Maniatis, Fritsch, and Sambrook (1982).

### Plasmids

The basic construction pGemD was obtained by fusing the coding sequence of neomycin phosphotransferase II (*nptII*) (Herrera-Estrella et al., 1983) to the cauliflower mosaic virus 35S promoter (Odell, Nagy, and Chua, 1985) in the gemini vector pGem2 (Promega Biotec). In the polylinker 18 bp downstream from the *nptII* stop codon, fragments containing the 3' end sequences of the octopine synthase gene, the *Arabidopsis* 2S-1 gene, the *Arabidopsis* *rbcS* small subunit gene, the *Daucus carota* extensin gene, and the *Antirrhinum majus* chalcone synthase gene were cloned. The resulting plasmids were called pGemOCS, pGem2S, pGemSS, pGemEXT, and pGemCHS, respectively. A description of the cloned 3' end sequences is presented in Table 1. These *nptII* fusion genes were subcloned in pGV1503, which is derived from pGV1500 (Deblaere et al., 1987) by insertion of the hygromycin resistance gene, yielding the plasmids pGVOCS, pGV2S, pGVSS, pGVEXT, and pGVCHS, respectively.

The *nptII* fusion genes from pGemD and pGemOCS were also subcloned into the pHBS plasmid, which is similar to pGV1503 but lacks T-DNA borders, yielding pHBSD and pHBSOCS.

The deletion mutants were generated by Bal31 reaction on the 3' *ocs* fragment in pGemOCS, starting from the EcoRI site in the polylinker downstream from the 3' end sequence. After treatment with the Klenow enzyme, molecules were self-ligated and transformed in *Escherichia coli* MC1061 (Casadaban and Cohen, 1980). The various deletion mutants were characterized by DNA sequencing (Maxam and Gilbert, 1980). A chloramphenicol acetyltransferase gene under control of the  $T_R2'$  promoter and 3' *ocs* region was subcloned in pGemD, in pGemOCS, and in the different deletion plasmids, yielding the plasmids pCND, pCNOCS $\Delta$ 1, pCNOCS $\Delta$ 2, pCNOCS $\Delta$ 3, pCNOCS $\Delta$ 4, pCNOCS $\Delta$ 5, and pCNOCS $\Delta$ 6, respectively. In addition, a 705-bp fragment from pAGV828 (Herrera-Estrella et al., 1983) containing the 3' end sequences of the *ocs* gene was subcloned in pCND, yielding pCNOCS.

### Conjugation and Plant Transformation

All pGV1503- and pHBS-derived plasmid constructions were mobilized to *Agrobacterium* C58C1Rif<sup>r</sup> by triparental mating (Van Haute et al., 1983). pGV1503-derived plasmids were stabilized in *Agrobacterium* C58C1Rif<sup>r</sup> by cointegration into the resident pGV2260 virulence plasmid (Deblaere et al., 1985). The pHBS-derived plasmids were introduced into the *Agrobacterium* strain C58C1Rif<sup>r</sup> containing the plasmid pGV3850 (Zambryski et al., 1983). DNA gel blot analysis on total *Agrobacterium* DNA was performed to confirm the structure and the integrity of the cointegrates (Dhaese et al., 1983).

*Nicotiana tabacum* var. SR1 was transformed via the leaf disc infection method or via co-cultivation of regenerating protoplasts (De Block et al., 1987). All transformed plant lines were selected in medium containing 50  $\mu$ g/mL hygromycin.

### NPTII Assay

NPTII activity was determined by in situ detection of NPTII activity in nondenaturing polyacrylamide gels (Van den Broeck et al., 1985). The samples were standardized on either total protein content or on the CAT activity levels (see CAT Assays). The protein concentration was determined by the method of Bradford (1976) using the kit supplied by Bio-Rad.

Quantification of the NPTII activity was performed as follows: the phosphorylated kanamycin spots were cut out of the P81 blot together with a control piece isolated from a location below the signals to estimate background levels of radioactivity. Radioactivity in each piece of P81 paper was determined using nonaqueous scintillation counting (5 min) in the <sup>32</sup>P channel of a scintillation counter. The counts per minute values obtained were between 500 cpm and 100,000 cpm. Because the absolute amount of <sup>32</sup>P counts differed significantly in independent experiments due to experimental variables, the counts per minute values obtained from the different constructs were expressed relative to the counts per minute values of the GVOCS plasmids.

### CAT Assay

CAT activity was determined using a modified version of the "diffusion CAT assay protocol" (Neumann, Morency, and Russian, 1987). At 20-min intervals, the samples were counted for 45 sec in the <sup>14</sup>C channel of a liquid scintillation counter. Three to five measurements were routinely used to plot counts per minute values versus time for the different reactions. CAT activity was determined by calculating the reaction rate, and these values were used to standardize the amount of crude cell extracts to be analyzed for NPTII activity. A linear correlation was found between the reaction rate and CAT concentration if final counts per minute values were not exceeding  $\pm 40,000$ .

### Electroporation Conditions

Generally, 1 to 2  $\times 10^8$  protoplasts were electroporated with 10  $\mu$ g of plasmid DNA using a homemade 200- $\mu$ F capacitor at 200 V across a 0.6-cm path (Dekeyser et al., 1989). For the CAT and NPTII assays, the protoplasts were incubated for 48 hr at 28°C. For RNA preparations, the protoplasts were incubated only for 1 hr or 4 hr.

### RNA Analysis

Total RNA was extracted from transformed *N. tabacum* leaves and from electroporated protoplasts essentially as described by Jones, Dunsmuir, and Bedbrook (1985). Poly(A)<sup>+</sup> RNA was purified by oligo(dT)-cellulose affinity chromatography according to Slater (1984).

RNA was electrophoresed in a 1.5% agarose/formaldehyde gel and transferred to nylon filters (Hybond-N; Amersham Corp.). The blotting and hybridization were performed as recommended by Amersham Corp. The coding sequence of the *nptII* gene and of the  $\beta$ -subunit gene of the mitochondrial ATP synthase (Boutry and Chua, 1985) were cloned in antisense orientation in an SP6



vector and transcribed into a riboprobe according to Melton et al. (1984).

#### ACKNOWLEDGMENTS

We thank Dr. Diane Jofuku and Dr. Allan Caplan for critically reading the manuscript and Dr. John Van Emmelo and Geert Angenon for stimulating discussions. We also thank Anni Jacobs for excellent technical assistance, Katia Van Compernelle for help with restriction analysis of the deletion mutants, Martine De Cock for typing the manuscript, Karel Spruyt, Stefaan Van Gijsegem, and Vera Vermaercke for figures and photographs, and Jeroen Coppieters for computer assistance. Finally, we thank Dr. Nam-Hai Chua, Dr. Enno Krebbers, Dr. Heins Saedler, and Dr. Joseph Varner for the generous gift of plasmids. This work was supported by grants from the Services of the Prime Minister (O.O.A. 12.0562.84 and U.I.A.P. 12.OC01.87). I.I. and L.H. are indebted to the I.W.O.N.L. for a predoctoral fellowship and a postdoctoral fellowship, respectively, and R.D. is a Research Assistant of the National Fund for Scientific Research (Belgium).

Received April 21, 1989; revised May 8, 1989.

#### 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.
- Birnstiel, M.L., Busslinger, M., and Strub, K. (1985). Transcription termination and 3' processing: The end is in site! *Cell* **41**, 349–359.
- Boutry, M., and Chua, N.-H. (1985). A nuclear gene encoding the beta subunit of the mitochondrial ATP synthase in *Nicotiana plumbaginifolia*. *EMBO J.* **4**, 2159–2165.
- Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254.
- Brewer, G., and Ross, J. (1988). Poly(A) shortening and degradation of the 3' A+U-rich sequences of human *c-myc* mRNA in a cell-free system. *Mol. Cell. Biol.* **8**, 1697–1708.
- Caput, D., Beutler, B., Hartog, K., Thayer, R., Brown-Shimer, S., and Cerami, A. (1986). Identification of a common nucleotide sequence in the 3' untranslated region of mRNA molecules specifying inflammatory mediators. *Proc. Natl. Acad. Sci. USA* **83**, 1670–1674.
- Casadaban, M.J., and Cohen, S.N. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J. Mol. Biol.* **138**, 179–207.
- Chen, J.-C., and Varner, J.E. (1985). An extracellular matrix protein in plants: Characterization of a genomic clone for carrot extensin. *EMBO J.* **4**, 2145–2151.
- Deblaere, R., Bytebier, B., De Greve, H., Deboeck, F., Schell, J., Van Montagu, M., and Leemans, J. (1985). Efficient octopine Ti plasmid-derived vectors for *Agrobacterium*-mediated gene transfer to plants. *Nucl. Acids Res.* **13**, 4777–4788.
- Deblaere, R., Reynaerts, A., Höfte, H., Hernalsteens, J.-P., Leemans, J., and Van Montagu, M. (1987). Vectors for cloning in plant cells. In *Recombinant DNA, Part D (Methods in Enzymology, Vol. 153)*, R. Wu and L. Grossman, eds (New York: Academic Press), pp. 277–292.
- De Block, M., Botterman, J., Vandewiele, M., Dockx, J., Thoen, C., Gosselé, V., Movva, R., Thompson, C., Van Montagu, M., and Leemans, J. (1987). Engineering herbicide resistance in plants by expression of a detoxifying enzyme. *EMBO J.* **6**, 2513–2518.
- De Greve, H., Dhaese, P., Seurinck, J., Lemmers, M., Van Montagu, M., and Schell, J. (1982). Nucleotide sequence and transcript map of the *Agrobacterium tumefaciens* Ti plasmid-encoded octopine synthase gene. *J. Mol. Appl. Genet.* **1**, 499–512.
- Dekeyser, R., Claes, B., Marichal, M., Van Montagu, M., and Caplan, A. (1989). Evaluation of selectable markers for rice transformation. *Plant Physiol.* **90**, 217–223.
- Dhaese, P., De Greve, H., Gielen, J., Seurinck, J., Van Montagu, M., and Schell, J. (1983). Identification of sequences involved in the polyadenylation of higher plant nuclear transcripts using *Agrobacterium* T-DNA genes as models. *EMBO J.* **2**, 419–426.
- Elliston, K., and Messing, J. (1988). The molecular architecture of plant genes: A phylogenetic perspective. In *Architecture of Eucaryotic Genes*, G. Kahl, ed (Weinheim: VCH Verlag), pp. 21–56.
- Fitzgerald, M., and Shenk, T. (1981). The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* **24**, 251–260.
- Gil, A., and Proudfoot, N.J. (1984). A sequence downstream of AAUAAA is required for rabbit  $\beta$ -globin mRNA 3'-end formation. *Nature* **312**, 473–474.
- Gil, A., and Proudfoot, N.J. (1987). Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit  $\beta$ -globin mRNA 3' end formation. *Cell* **49**, 399–406.
- Hart, R.P., McDevitt, M.A., and Nevins, J.R. (1985). Poly(A) site cleavage in a HeLa nuclear extract is dependent on downstream sequences. *Cell* **43**, 677–683.
- Herrera-Estrella, L., Depicker, A., Van Montagu, M., and Schell, J. (1983). Expression of chimaeric genes transferred into plant cells using a Ti-plasmid-derived vector. *Nature* **303**, 209–213.
- 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.* **8**, 23–35.
- Jones, J.D.G., Dunsmuir, P., and Bedbrook, J. (1985). High level expression of introduced chimeric genes in regenerated transformed plants. *EMBO J.* **4**, 2411–2418.
- Joshi, C.P. (1987). Putative polyadenylation signals in nuclear genes of higher plants: A compilation and analysis. *Nucl. Acids Res.* **15**, 9627–9640.
- Kessler, M.M., Westhafer, M.A., Carson, D.D., and Nordstrom, J.L. (1987). Polyadenylation at a cryptic site in the pBR322 portion of pSV2-neo: Prevention of its utilization by the SV40

- late poly(A) signal. *Nucl. Acids Res.* **15**, 631–642.
- Krebbers, E., Herdies, L., De Clercq, A., Seurinck, J., Leemans, J., Van Damme, J., Segura, M., Gheysen, G., Van Montagu, M., and Vandekerckhove, J.** (1988a). Determination of the processing sites of an *Arabidopsis* 2S albumin and characterization of the complete gene family. *Plant Physiol.* **87**, 859–866.
- Krebbers, E., Seurinck, J., Herdies, L., Cashmore, A.R., and Timko, M.P.** (1988b). Four genes in two diverged subfamilies encode the ribulose-1,5-bisphosphate carboxylase small subunit polypeptides of *Arabidopsis thaliana*. *Plant Mol. Biol.* **11**, 745–759.
- Lipphardt, S., Brettschneider, R., Kreuzaler, F., Schell, J., and Dangel, J.L.** (1988). UV-inducible transient expression in parsley protoplasts identifies regulatory *cis*-elements of a chimeric *Antirrhinum majus* chalcone synthase gene. *EMBO J.* **7**, 4027–4033.
- Maniatis, T., Fritsch, E.F., and Sambrook, J.** (1982). *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).
- Maxam, A.M., and Gilbert, W.** (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. In *Nucleic Acids, Part I* (Methods in Enzymology, Vol. 65), L. Grossman, and K. Moldave, eds (New York: Academic Press), pp. 499–559.
- McDevitt, M.A., Hart, R.P., Wong, W.W., and Nevins, J.R.** (1986). Sequences capable of restoring poly(A) site function to define two distinct downstream elements. *EMBO J.* **5**, 2907–2913.
- McLauchlan, J., Gaffney, D., Whitton, J.L., and Clements, J.B.** (1985). The consensus sequence YGTGTTY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucl. Acids Res.* **13**, 1347–1368.
- Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., and Green, M.R.** (1984). Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucl. Acids Res.* **12**, 7035–7056.
- Montell, C., Fisher, E.F., Caruthers, M.H., and Berk, A.J.** (1983). Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3' consensus sequence AAUAAA. *Nature* **305**, 600–605.
- Neumann, J.R., Morency, C.A., and Russian, K.O.** (1987). A novel rapid assay for chloramphenicol acetyltransferase gene expression. *BioTechniques* **5**, 444–447.
- Odell, J.T., Nagy, F., and Chua, N.-H.** (1985). Identification of DNA sequences required for the activity of the cauliflower mosaic virus 35S promoter. *Nature* **313**, 810–812.
- Petersen, R., and Lindquist, S.** (1988). The *Drosophila hsp70* message is rapidly degraded at normal temperatures and stabilized by heat shock. *Gene* **72**, 161–168.
- Proudfoot, N.J., and Brownlee, G.G.** (1976). 3' Non-coding region sequences in eucaryotic messenger RNA. *Nature* **263**, 211–214.
- Proudfoot, N.J., and Whitelaw, E.** (1987). Termination and 3' end processing of eukaryotic RNA. In *Frontiers in Molecular Biology—Transcription and Splicing*, D.M. Glover and B.D. Hames, eds (Oxford: IRL Press), pp. 97–129.
- Shaw, G., and Kamen, R.** (1986). A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46**, 659–667.
- Simcox, A.A., Cheney, C.M., Hoffman, E.P., and Shearn, A.** (1985). A deletion of the 3' end of the *Drosophila melanogaster hsp70* gene increases stability of mutant mRNA during recovery from heat shock. *Mol. Cell. Biol.* **5**, 3397–3402.
- Slater, R.J.** (1984). The purification of poly(A)-containing RNA by affinity chromatography. In *Methods in Molecular Biology*, Vol. 2, J.M. Walker, ed (Clifton, NJ: Humana Press), pp. 117–120.
- Sommer, H., and Saedler, H.** (1986). Structure of the chalcone synthase gene of *Antirrhinum majus*. *Mol. Gen. Genet.* **202**, 429–434.
- Van den Broeck, G., Timko, M.P., Kausch, A.P., Cashmore, A.R., Van Montagu, M., and Herrera-Estrella, L.** (1985). Targeting of a foreign protein to chloroplasts by fusion to the transit peptide of ribulose 1,5-bisphosphate carboxylase. *Nature* **313**, 358–363.
- van den Elzen, P.J.M., Townsend, J., Lee, K.Y., and Bedbrook, J.R.** (1985). A chimaeric hygromycin resistance gene as a selectable marker in plant cells. *Plant Mol. Biol.* **5**, 299–302.
- Van Haute, E., Joos, H., Maes, M., Warren, G., Van Montagu, M., and Schell, J.** (1983). Intergeneric transfer and exchange recombination of restriction fragments cloned in pBR322: A novel strategy for the reversed genetics of Ti plasmids of *Agrobacterium tumefaciens*. *EMBO J.* **2**, 411–418.
- Wickens, M., and Stephenson, P.** (1984). Role of the conserved AAUAAA sequence: Four AAUAAA point mutants prevent messenger RNA 3' end formation. *Science* **226**, 1045–1051.
- Woychik, R.P., Lyons, R.H., Post, L., and Rottman, F.M.** (1984). Requirement for the 3'-flanking region of the bovine growth hormone gene for accurate polyadenylation. *Proc. Natl. Acad. Sci. USA* **81**, 3944–3948.
- Zambryski, P., Joos, H., Genetello, C., Leemans, J., Van Montagu, M., and Schell, J.** (1983). Ti plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. *EMBO J.* **2**, 2143–2150.
- Zhang, F., Denome, R.M., and Cole, C.N.** (1986). Fine-structure analysis of the processing and polyadenylation region of the herpes simplex virus type 1 thymidine kinase gene by using linker scanning, internal deletion, and insertion mutations. *Mol. Cell. Biol.* **6**, 4611–4623.