The EMBO Journal Vol.2 No.3 pp.419–426, 1983

Identification of sequences involved in the polyadenylation of higher plant nuclear transcripts using *Agrobacterium* T-DNA genes as models

Patrick Dhaese[1], Henri De Greve[2], Jan Gielen[1], Jozef Seurinck[1], Marc Van Montagu[1,2] and Jozef Schell[1,2,3*]

[1]Laboratorium voor Genetica, Rijksuniversiteit Gent, B-9000 Gent, [2]Laboratorium voor Genetische Virologie, Vrije Universiteit Brussel, B-1640 St.-Genesius-Rode, Belgium, and [3]Max-Planck-Institut für Züchtungsforschung, D-5000 Köln, FRG

Communicated by J. Schell
Received on 4 January 1983

Sequences in the 3′-untranslated region of two different octopine T-DNA genes were analyzed with regard to their significance in polyadenylation. Poly(A) addition sites were localized precisely by S1 nuclease mapping with T-DNA-derived mRNAs isolated from tobacco. The gene encoding 'transcript 7' contains two AATAAA hexanucleotides, respectively 119 bp and 170 bp downstream of the TAA stop codon. A single poly(A) site was mapped 24–25 bp downstream of the first AATAAA. Further, we show that a mutant octopine synthase gene, which has lost part of its 3′-untranslated region by deletion, is still active. This mutant gene terminates 19 bp upstream from the major wild-type polyadenylation site. The deletion also removes the AATAAT signal preceding this site. The mutant octopine synthase gene contains a minimum of four different poly(A) sites. The most prominent of these sites is identical to the minor poly(A) site of the wild-type gene, and is preceded by a sequence AATGAATATA. Three other sites are located within the adjacent plant DNA, giving rise to hybrid T-DNA/plant DNA transcripts. The two most distal sites are probably dependent on a motif AATAAATAAA, found 29 bp away from the T-DNA/plant DNA junction.

Key words: Agrobacterium/octopine synthase/polyadenylation/S1 mapping/T-DNA sequences

Introduction

The molecular basis for crown gall, a neoplasmic transformation of plant cells, is the transfer and integration into nuclear plant DNA of a segment of the *Agrobacterium tumefaciens* Ti plasmid. The transferred segment is called T-DNA (for a recent review, see Kahl and Schell, 1982). T-DNA is expressed in plant cells, thereby determining their morphogenetic growth properties and also the biosynthesis of low mol. wt. compounds, termed 'opines', that are not found in normal plant tissue (Tempé and Goldmann, 1982). Opines formed the basis for a classification of crown gall tumors and their causative Ti plasmids (Guyon *et al.*, 1980). To date, most studies have concentrated on the octopine- and nopaline-types.

In octopine-type tumors, the TL-DNA (Thomashow *et al.*, 1980; De Beuckeleer *et al.*, 1981) encodes several distinct polyadenylated transcripts (Willmitzer *et al.*, 1982; Gelvin *et al.*, 1982) of different size and relative abundance, whose synthesis is sensitive to α-amanitin (Willmitzer *et al.*, 1981). These findings suggested that T-DNA genes are transcribed by the host RNA polymerase II, and therefore can serve as models for defining trancriptional and translational control sequences in nuclear, protein-coding plant genes. Indeed, the availability of TL-DNA genomic sequences, as fragments cloned from the Ti plasmid (De Vos *et al.*, 1981), constitutes a great advantage.

Recently, the structure of the octopine synthase gene, located near the right border of the TL-DNA (Figure 1), was elucidated (De Greve *et al.*, 1982a), and the 5′ and 3′ termini of its mature cytoplasmic mRNA, isolated from a tobacco crown gall tumor, were mapped at the nucleotide level. A 'Goldberg-Hogness' box, TATTTAA, was found 32 nucleotides upstream from the transcription initiation site. A hexanucleotide sequence AATAAT preceded the major site of poly(A) addition by 17 nucleotides. The same AATAAT sequence, representing a one-base variation of the consensus polyadenylation signal AATAAA (Proudfoot and Brownlee, 1976; Fitzgerald and Shenk, 1981) was found 54 nucleotides upstream of the 3′ end of another T-DNA gene, nopaline synthase (Depicker *et al.*, 1982).

Here we report, firstly the DNA sequence of the gene for 'transcript 7' (Willmitzer *et al.*, 1982), the smallest (~670 nucleotides) polyadenylated transcript encoded by octopine TL-DNA. The function of this particular gene in the context of crown gall tumorigenesis remains unknown, whereas specific roles have been assigned to most of the other TL-DNA-encoded genes (Garfinkel *et al.*, 1981; Leemans *et al.*, 1982). Northern hybridizations (Willmitzer *et al.*, 1982; Gelvin *et al.*, 1982) using poly(A)$^+$ mRNA from tobacco octopine tumor lines indicated that the transcript is rather abundant compared with the other mRNAs specified by octopine TL-DNA, and that it maps entirely within the right *Sma*I/*Hind*III subfragment of *Hind*III-18c (Figure 1). We have used S1 nuclease mapping (Berk and Sharp, 1977; Weaver and Weissman, 1979) to determine precisely the initiation and poly(A) sites for transcript 7.

Secondly, we have studied in detail the octopine synthase mRNA isolated from a regenerated tobacco plant, rGV1 (Otten *et al.*, 1981). In rGV1, all TL-DNA genes that determine neoplasmic growth are deleted, and only octopine synthesis is expressed. Consistently, Southern blottings showed that the remnant T-DNA portion is limited to a 1.6-kb segment (Figure 1), originating from the extreme right part of the TL-DNA (De Greve *et al.*, 1982b). Here, we define the left border of the rGV1 T-DNA, and show that, whereas the 5′ leader and entire coding sequence of the octopine synthase gene are retained, the authentic AATAAT sequence and the major poly(A) addition site are deleted. The adjacent plant DNA sequence is very AT-rich and contains a motif AATAAATAAA close to the T-DNA/plant DNA junction. Since normal amounts of wild-type size octopine synthase mRNA were detected in the cytoplasmic poly(A)$^+$ RNA fraction from rGV1 extracts (De Greve *et al.*, 1982b), the possibility of finding poly(A) addition sites within the plant DNA was examined by 3′ end mapping of the rGV1 octopine synthase mRNA.

Our results are discussed with respect to what is currently known about polyadenylation in higher plant nuclear genes.

*To whom reprint requests should be sent.

© IRL Press Limited, Oxford, England.

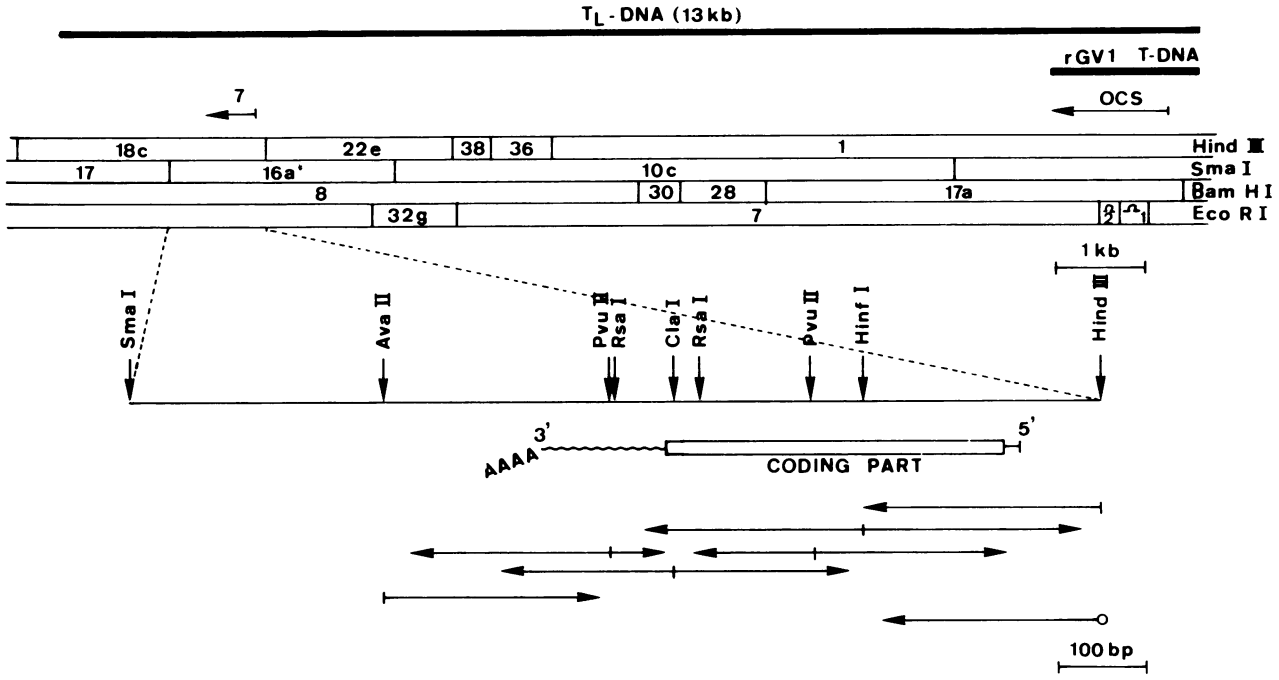

Fig. 1. Upper part: restriction map of part of the octopine Ti plasmid pTiAch5, with the indication of: (i) the extent of the TL-DNA and the rGV1 T-DNA; (ii) the map position and polarity of the two TL-DNA transcripts studied here: octopine synthase (*ocs*) and transcript 7. **Lower part**: detailed restriction map of the *Sma*I-*Hind*III subfragment from *Hind*III-18c showing the strategy for sequencing the gene encoding 'transcript 7' ◀—— : 5′ end-labeling; ◀——○: 3′ end-labeling.

Results

Structure of the octopine TL-DNA gene encoding 'transcript 7'

Plasmid pGV117 DNA was used to generate restriction fragments mapping within the *Sma*I-*Hind*III subfragment of *Hind*III-18c. Sequence determination of this region (for strategy used: see Figure 1) and subsequent computer analysis revealed an open reading frame encoding a polypeptide of 125 amino acids (Figure 2), extending from right to left on the conventional TL-DNA map.

Several lines of evidence suggest that this reading frame represents the translated part of the gene for 'transcript 7': (i) the direction is consistent with the experimentally determined polarity for transcript 7 (Willmitzer *et al.*, 1982); (ii) the sequence originates from a region outside the homology with nopaline T-DNA (Engler *et al.*, 1981) in accordance with the absence of transcript 7 in nopaline crown gall tissue; (iii) the predicted mol. wt. of the translation product is 14 129 daltons, in agreement with the detection of a polypeptide of apparent mol. wt. 14 000 in SDS gels, upon *in vitro* translation of A6-S1 tumor RNA selected by hybridization against *Hind*III fragment 18c (Schröder and Schröder, 1982).

To determine the precise initiation site(s) of 'transcript 7', a *Hinf*I digest of pGV117 was 5′-^{32}P-labeled and a 319-bp fragment was isolated. This fragment extends from a *Hinf*I site within the coding sequence (Figure 2) and passes the right *Hind*III-18c/pBR322 junction in pGV117 to a *Hinf*I site 50 bp in the vector DNA. The labeled fragment was strand-separated on a 7% polyacrylamide gel, and each strand was hybridized separately to poly(A)$^+$ RNA (± 10 μg) from the A6-S1 tumor line, and to nopaline BT37 tumor total RNA used as a control. After S1 nuclease digestion performed on each of the four samples, the products were analyzed on an 8% acrylamide gel. Figure 3A shows the results obtained for the complementary DNA strand. The tracks corresponding to the non-complementary strand remained completely blank. This was also the case in the 3′ end mapping experiment described below.

A clustered set of five bands (Figure 3A, track 2) is visible at positions corresponding to the adenosine residues in the sequence, CAAACACA, starting 17 bp upstream of the ATG codon (Figure 2). This may either reflect a slight inaccuracy inherent to the action of S1 nuclease or a true microheterogeneity in the mRNA start positions. A sequence ATATATA is present 31 − 25 nucleotides upstream of the first of these start sites. Such an arrangement is characteristic for RNA polymerase II-transcribed sequences in animal cells (Breathnach and Chambon, 1981).

Downstream of the TAA translational stop codon, two AATAAA sequences are present (+119 → 124 bp and +170 → 175 bp, respectively), each forming part of a 10-bp direct repeat (Figure 2). To determine the poly(A) addition site(s) for transcript 7, we chose the *Cla*I site, 9 bp upstream of the TAA codon, as a reference point.

pGV117 DNA was digested with *Cla*I, followed by 3′ end-labeling using Klenow polymerase and [α-^{32}P]dCTP (>3000 Ci/mmol) as the labeled nucleotide. A 1700-bp *Cla*I fragment, labeled at both 3′ ends, was isolated and strand-separated on a 3% polyacrylamide gel. Both strands were hybridized to A6-S1 poly(A)$^+$ RNA and to control RNA, then S1 nuclease-digested and analyzed as in the 5′ end mapping experiment.

Two adjacent bands are seen in the A6-S1 poly(A)$^+$ RNA lane, which are clearly absent from the control RNA track (Figure 3B). From their position relative to the sequence ladder we infer that the protected fragment ends at a TA dinucleotide 24 − 25 bp downstream of the first AATAAA sequence. The A residue might either be specified by the

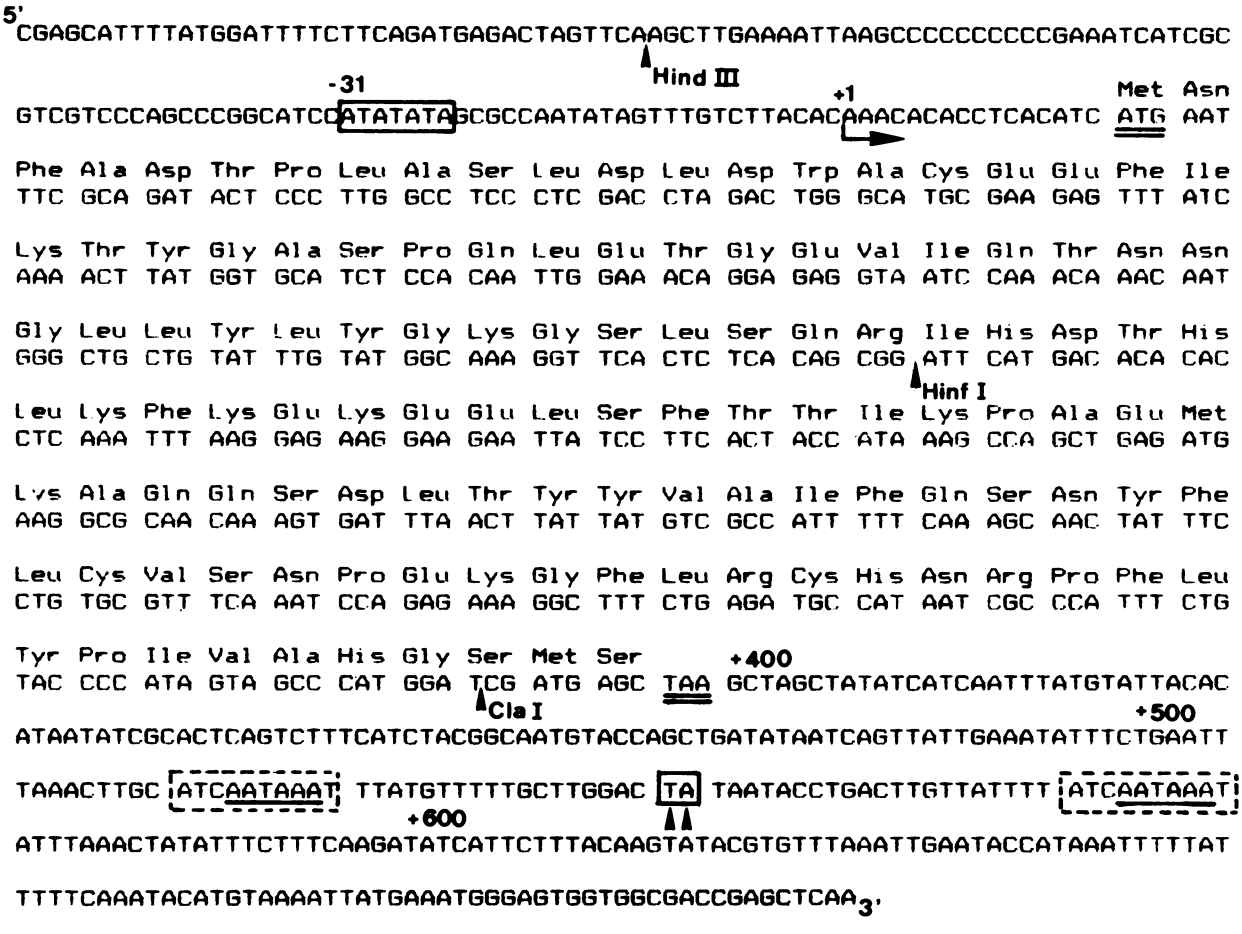

Fig. 2. Nucleotide sequence of the TL-DNA region encoding 'transcript 7'. The sequence is presented in the opposite direction with respect to the TL-DNA map. Three relevant restriction sites are indicated: the *Hind*III site bordering fragments 18c and 22e, the *Hinf*I site used in the 5′ end transcript mapping and the *Cla*I site that served in the poly(A) site localization. Nucleotides are numbered beginning at the first mRNA site start (+1). The 'TATA' sequence is boxed (−31). Translational initiation and stop codons are underlined twice. Two consecutive vertical arrows indicate the poly(A) addition site. Note further the two 10-bp repeats (broken-lined boxes) each containing a 'consensus' polyadenylation signal (underlined).

genome, or added post-transcriptionally as the first nucleotide of the poly(A) tail. Since no other significant bands are seen, it appears that only the first AATAAA hexanucleotide is functional as a recognition site for poly(A) addition. The genomic distance between the 5′ end and poly(A) site of transcript 7, as determined here, measures 542 bp, a value consistent with the Northern blot estimations (0.67 − 0.73 kb) taking into account a usual length of 100 − 200 nucleotides for the poly(A) tail. Therefore, the gene most probably does not contain introns.

The overall G + C content of the 3′-untranslated region is very low (29.9%), a feature already observed in other nuclear plant genes, including octopine synthase (De Greve *et al.*, 1982a; Croy *et al.*, 1982).

Sequence analysis of the left T-DNA border in the rGV1 regenerate

Southern-type hybridizations indicated that the T-DNA in the regenerate plant rGV1 essentially consists of only octopine synthase sequences (Figure 1) (De Greve *et al.*, 1982b). The rGV1 T-DNA encompasses, at its right end, the *Bam*HI site 120 bp upstream of the 5′ end of the mRNA. The left border was localized within a 750-bp *Pvu*II fragment, beginning at the *Pvu*II site 14 bp downstream of the coding part of the octopine synthase (Figure 4). We ought to define the precise location of the left border relative to the 3′ end of the gene by comparing rGV1 genomic sequences from this region with the corresponding TL-DNA sequence, as determined recently (De Greve *et al.*, 1982a).

The isolation of the rGV1 genomic clone cgGV1-1, consisting of a 3.2-kb *Hind*III fragment, encompassing the whole rGV1 T-DNA inserted in a Charon 21A λ vector, has been described (De Greve *et al.*, 1982b). The left *Hind*III-*Bam*HI subfragment (2.1 kb) of this insert, extending from the *Bam*HI site in the rGV1 T-DNA to the plant *Hind*III site past the left border, was subcloned in pBR325 by substitution of the *Hind*III-*Bam*HI fragment of the vector (Figure 4). The resulting plasmid was designated pGV718 (7.7 kb).

Sequences around the left border were determined using fragments 5′-labeled at the restriction sites shown in Figure 4. The results (Figure 5) reveal that the rGV1 T-DNA ends within the 3′-untranslated sequence of the *ocs* gene, more precisely at an A residue. From this point on, the rGV1 sequence diverges completely from the TL-DNA sequence, indicating that it originates from resident plant nuclear DNA. This adjacent plant DNA is very AT-rich.

The deleted portion (19 bp) of the octopine synthase gene includes the major poly(A) site (M) and the presumptive signal for poly(A) addition, AATAAT.

3′ End mapping of rGV1 octopine synthase mRNA

Although, as shown above, the rGV1 octopine synthase

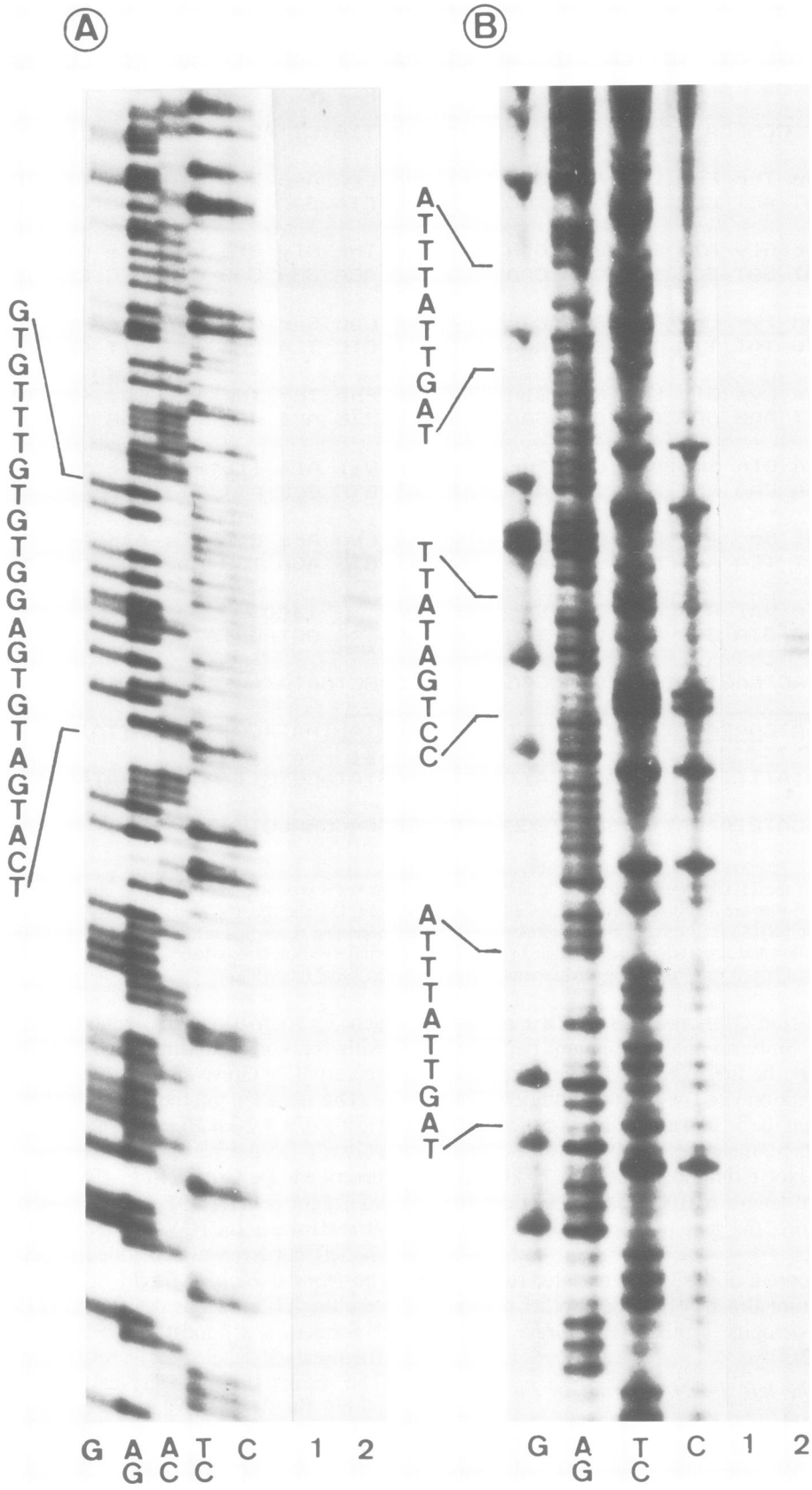

Fig. 3. S1 nuclease mapping of the initiation and poly(A) addition sites for 'transcript 7'. **A.** 5′ end mapping. S1 nuclease-resistant fragments, resulting from hybridization of the coding strand, 5′-^{32}P-labeled at the *Hinf*I site (see Figure 2), with A6-S1 poly(A)$^+$ RNA (**track 2**) were sized on an 8% polyacrylamide-urea gel. As a control, nopaline tumor RNA was also hybridized to the ^{32}P-labeled probe DNA followed by S1 digestion (**track 1**). Tracks G, A+G, A+C, C+T, C represent products of a Maxam and Gilbert sequence reaction performed on the probe fragment, serving as accurate size markers. **B.** 3′ end mapping. In this case, the coding strand was 3′ end-labeled at the *Cla*I site (see Figure 2). The tracks are arranged similarly as in **A**: hybridization with control RNA (**track 1**) and A6-S1 poly(A)$^+$ RNA (**track 2**). Relevant sequences (complementary to those displayed in Figure 2) are shown to the left of each Maxam-Gilbert ladder.

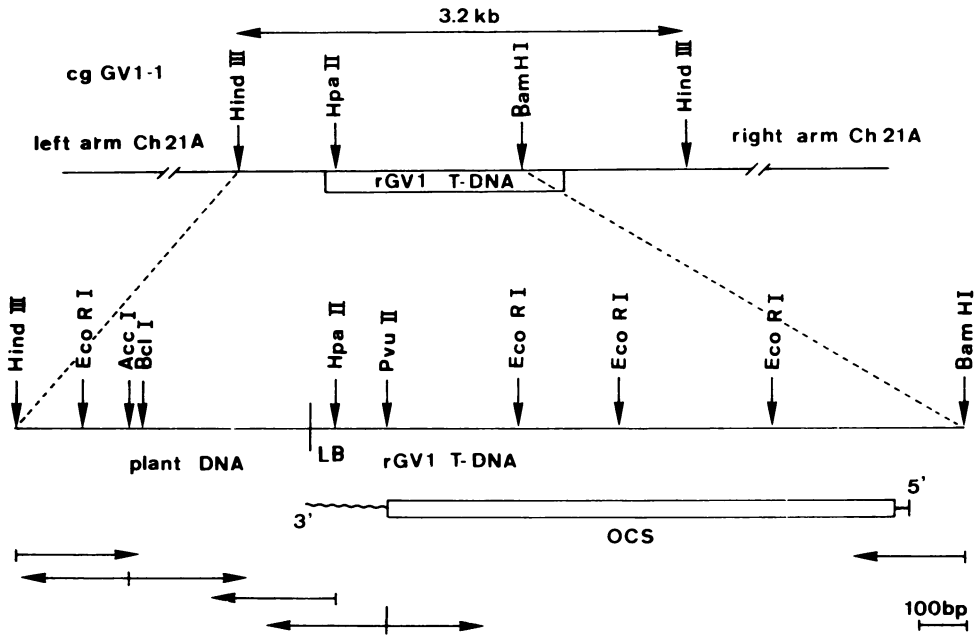

Fig. 4. Organization of genomic clones containing the T-DNA present in the rGV1 regenerate tobacco plant. The position of the octopine synthase mRNA (*ocs*) is shown relative to a restriction map of the *Hind*I-*Bam*HI fragment, which was subcloned in pBR325. Also shown is the sequencing strategy used to locate the left border (LB) of the rGV1 T-DNA.

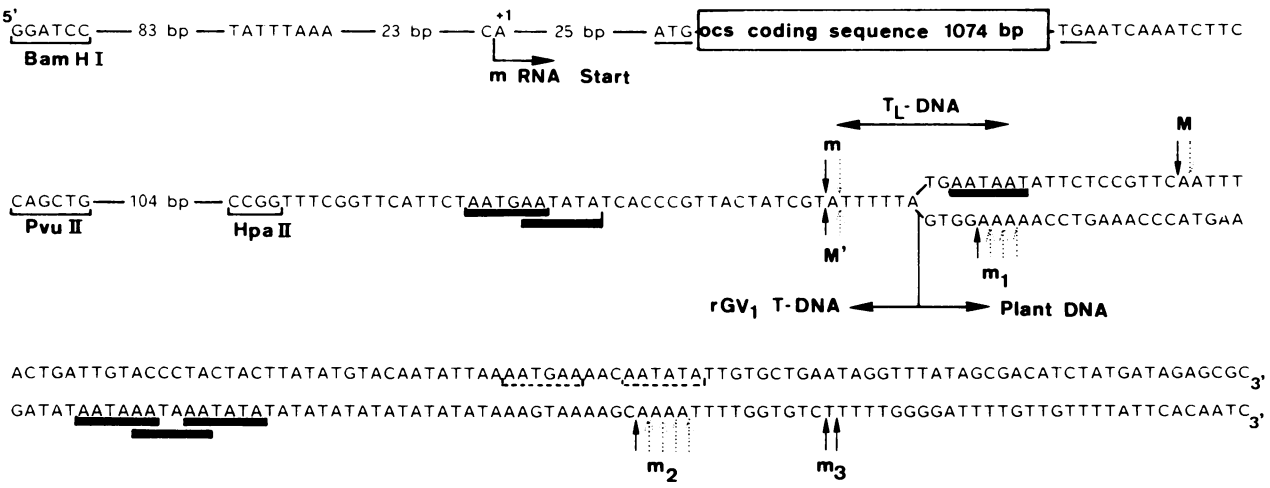

Fig. 5. Nucleotide sequences around the 3′ end of the octopine synthase gene in the TL-DNA, and in the rGV1 T-DNA/plant DNA junction region. The sequence is presented in the opposite orientation with respect to the restriction maps of Figures 1 and 4. The divergence into two sequences marks the left border of the rGV1-T-DNA. From this point on, the upper row shows the continuing Ti-plasmid sequenced as present in the TL-DNA, containing the wild-type octopine synthase gene (De Greve *et al.*, 1982a). The lower row is the plant DNA sequence flanking the rGV1 T-DNA. Poly(A) addition sites in the wild-type *ocs* gene are indicated with downward arrows: minor (m) and major (M). The upward arrows show the position of poly(A) sites in the mutant rGV1 *ocs* gene: major (M′), and minor (m1, m2, m3). The dotted arrows, always following an A residue, indicate our inability to distinguish whether these residues were specified by the genome, or added post-transcriptionally to the mRNA as part of the poly(A) tail. The hexanucleotide signals upstream of the poly(A) site are underlined. The broken lines show two hexanucleotides without corresponding poly(A) sites.

gene lacks the major wild-type polyadenylation signal and corresponding site, an apparently normal amount of specific mRNA, of unaltered size (1400–1500 nucleotides), was detected in the rGV1 poly(A)$^+$ population by Northern blot hybridization (De Greve *et al.*, 1982b). The presence of a motif AATAAATAAA, only 29–39 bp downstream from the T-DNA/plant DNA junction in the rGV1 genome, prompted us to examine whether this sequence could substitute for the loss of the wild-type signal. If so, the polyadenylated transcript would be, at most, 100 nucleotides longer than the wild-type mRNA, a difference not usually detectable on Northern gels.

To determine precisely the 3′ end(s) of rGV1 octopine synthase mRNA, we again resorted to the S1 nuclease mapping technique. A *Pvu*II/*Hind*III digest of pGV718 was 3′ end-labeled using Klenow polymerase and [α-^{32}P]dATP. We isolated the 600-bp *Pvu*II-*Hind*III fragment (Figure 4), extending from the *Pvu*II site 14 bp downstream of the octopine synthase-coding sequence, to the *Hind*III site within the adjacent plant DNA. After strand separation, both strands were hybridized separately to 25 μg poly(A)$^+$ RNA from rGV1. As control RNA we used total RNA from the BT37 nopaline tumor line, and also poly(A)$^-$ rGV1 RNA. The S1 nuclease digestion products of these six combinations were analyzed

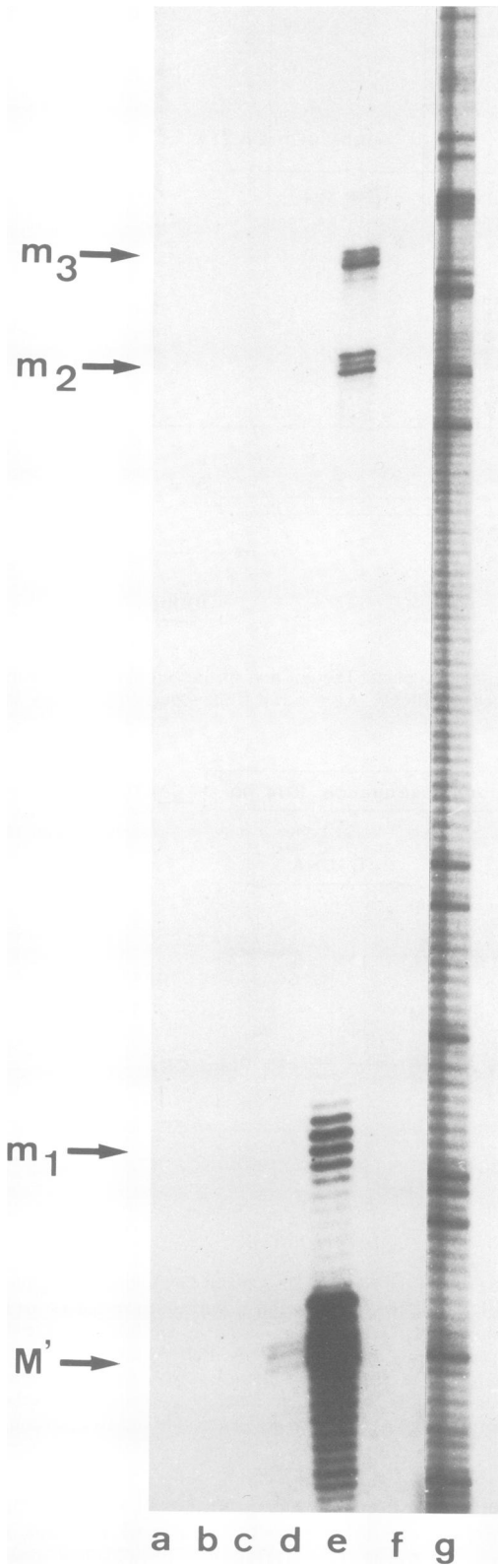

Fig. 6. 3′ end mapping of the rGV1 octopine synthase mRNA. The products of S1 nuclease digestion on the following DNA/RNA hybridization reactions were analyzed on an 8% polyacrylamide gel. (i) The non-complementary strand of the *Pvu*II-*Hind*III fragment combined with, respectively, control RNA **(track a)**, rGV1 poly(A)$^-$ RNA **(track b)** and rGV1 poly(A)$^+$ RNA **(track c).** (ii) The complementary strand of the *Pvu*II-*Hind*III fragment combined with, respectively, rGV1 poly(A)$^-$ RNA **(track d)**, rGV1 poly(A)$^+$ RNA **(track e)** and control RNA **(track f)**. **Track g** is a size marker consisting of a C reaction on the complementary strand (corresponding to a G residue in the sequence displayed in Figure 5).

next to a sequencing ladder of the corresponding strands (Figure 6). Protected fragments are visible only in tracks d and e, representing hybridization of the correct, complementary strand with rGV1 poly(A)$^-$ and poly(A)$^+$ RNA, respectively. The bands in track d probably result from a residual fraction of polyadenylated RNA not bound to oligo(dT) cellulose, since their intensity is several orders of magnitude lower than the corresponding bands in track e. The four sets of bands, visible in track e, reflect a heterogeneity at the 3′ end of rGV1 octopine synthase mRNA. The positions of the poly(A) addition sites corresponding to these four sets are indicated (M′, m1, m2 and m3) on the nucleotide sequence of rGV1 octopine synthase (Figure 5). Three poly(A) sites (m1, m2 and m3) are located within the plant DNA sequence. However, judging from the autoradiographic intensities, the major site (M′) is still within the T-DNA sequence at a TA dinucleotide 8–7 nucleotides upstream of the T-DNA/plant DNA junction.

This site corresponds to a minor poly(A) addition site (m) in the wild-type gene which we had erroneously considered as an artifact of the S1 mapping procedure, due to the extremely high local A-T content (De Greve *et al.*, 1982a). We now have evidence that these bands represent a *bona fide* poly(A) site: (i) the local A-T content in the rGV1 sequence has decreased, due to the presence of three G residues immediately past the border; (ii) lowering the hybridization and S1 digestion temperatures to 30°C and 18°C, respectively, and thus also the probability of local single strand formation, does not affect the band intensity (data not shown). A sequence AATGAATATA, respresenting two tandemly overlapping one-base variations of the consensus AATAAA signal, precedes this site (−27 to −18 bp). The second site (m1) might also depend on this signal. The third (m2) and fourth (m3) sites are located 33 bp and 48 bp, respectively, downstream of the AATAAATAAA sequence.

Discussion

Polyadenylation of the primary transcription unit is a key step in the expression of most eukaryotic protein-coding genes (Brawerman, 1981). The process occurs in the nucleus and probably consists of endonucleolytic cleavage, followed by the addition of a poly(A) tail to the newly generated 3′ end (Nevins and Darnell, 1978; Hofer and Darnell, 1981). In the 3′-untranslated region of nearly all animal, viral and cellular mRNAs, the hexanucleotide AAUAAA precedes the site of poly(A) addition by some 10–30 nucleotides (Proudfoot and Brownlee, 1976; Benoist *et al.*, 1980). Using late SV40 mRNAs, Fitzgerald and Shenk (1981) examined the effect of mutants within and around this sequence on polyadenylation, and concluded that it functions at least as part of an essential recognition site.

Here, we have attempted to assess the significance of sequences in the 3′-untranslated region for the choice of a poly(A) addition site in higher plant nuclear genes. Using S1 nuclease mapping, we have examined the polyadenylation of two octopine TL-DNA genes, each having a particularly interesting sequence arrangement around their 3′ ends.

The gene encoding 'transcript 7' contains two AATAAA sequences, separated by 42 nucleotides. We have located its poly(A) site in between these two 'canonical' polyadenylation signals, at exactly the same distance (24–25 bp) from their first A residue. Since it is generally assumed that the polyadenylation process occurs downstream of the putative signal

sequence, it appears that only the first AATAAA is functional. However, the primary transcription unit might extend considerably beyond the poly(A) site (Hofer and Darnell, 1981; Ziff, 1980) implying that the second AATAAA might also be part of the substrate offered to the polyadenylating machinery.

In a second case, we took advantage of a naturally isolated deletion mutation around the 3′ end of the TL-DNA gene, octopine synthase. The wild-type gene contains two poly(A) sites. The major site (M) occurs downstream of an AATAAT sequence, the minor (m) is preceded by -AATGAATATA- (Figure 5). The latter sequence is composed of two overlapping one-base variations of the canonical signal. We note that a copy of each AATGAA and AATATA is found further downstream in the 3′-flanking TL-DNA (broken lines in Figure 5). At the 3′ side of these sequences lies no poly(A) site.

In the rGV1 plant, T-DNA is limited to a segment containing only octopine synthase sequences. Moreover, the rGV1 T-DNA lacks the terminal 19 bp of the gene, including the major poly(A) site and signal. As a result, the minor poly(A) site of the wild-type *ocs* gene (m) has become the major site (M′) of the rGV1 mutant gene. In addition, a new poly(A) site (m1) is created in the plant DNA four nucleotides downstream of the rGV1 T-DNA border. Interestingly, this site which probably also depends on the signal AATGAATATA, has no counterpart at a corresponding distance from the AATGAATATA signal in the wild-type *ocs* gene (see Figure 5). A simple explanation for both effects would be that removal of the AATAAT sequence, which might serve as a protein-binding site in the wild-type gene, has made this region more accessible for cleavage in the rGV1 mutant gene. Two other new sites (m2, m3) are found in the plant DNA, downstream of the sequence AATAAATAA, which is composed of two overlapping canonical signals. These sites are used at least 10 times less efficiently than the T-DNA site, as estimated from the autoradiographic intensities of the bands in Figure 6. The above results suggest that larger, or other structural elements, besides the hexanucleotide sequences mentioned above, play a role in determining the polyadenylation efficiency. One possibility is that sequences immediately surrounding the poly(A) site affect the frequency of endonucleolytic cleavage. In this respect, we note that three of the four sites in the rGV1 *ocs* gene are closely followed by a stretch of three or more T residues.

The poly(A) addition sites of several T-DNA genes from tobacco crown gall tumors have now been determined. These include wild-type octopine synthase (De Greve *et al.*, 1982a), nopaline synthase (Depicker *et al.*, 1982), rGV1 octopine synthase and transcript 7 (this work). All these sites are preceded, within a distance of ~50 bases or less, by the hexanucleotide AATAAA or a variation thereof, differing by one nucleotide. A survey of sequences upstream of the 3′ ends of other nuclear plant genes such as leghemoglobins (Hyldig-Nielsen *et al.*, 1982; Wiborg *et al.*, 1982), the zein family (Marks and Larkins, 1982; Geraghty *et al.*, 1981), legumin (Croy *et al.*, 1982) leads to a similar conclusion. There are, however, two exceptions: alcohol dehydrogenase (Gerlach *et al.*, 1982) containing the sequence AATGAG and the small subunit of ribulose 1,5-bisphosphate carboxylase (Bedbrook *et al.*, 1980) showing no sequence resembling AATAAA upstream of the poly(A) site. On the other hand, not every AATAAA serves as a recognition site, as exemplified by the second AATAAA in the 3′-flanking DNA of transcript 7, and by the AATAAA in the nopaline synthase (Depicker *et al.*, 1982). This agrees again with other nuclear plant genes where in many cases an apparently non-functional AATAAA is found further upstream, much closer to the translational stop codon than to the poly(A) addition site (Messing *et al.*, 1983).

From the data summarized above, it seems that the poly(A) addition process in plants tolerates somewhat more variance in the actual structure of the recognition sites as compared with animal genes. However, multiple poly(A) sites do also occur in animal genes (Setzer *et al.*, 1980; Early *et al.*, 1980), and in several instances the putative signals differ from the canonical sequence by one or two nucleotides (Tosi *et al.*, 1981; Valenzuela *et al.* 1981; Hanukoglu and Fuchs, 1982; Setzer *et al.*, 1982). We are, therefore, probably dealing with a universal eukaryotic processing event of which the essential characteristics and function are identical in animals and plants. The *in vitro* polyadenylation system recently reported by Chen-Kiang *et al.* (1982) might constitute a breakthrough in dissecting the nature and substrate requirements of this process.

Finally, our understanding of the DNA sequence requirements for efficient polyadenylation might be beneficial for the construction of Ti plasmid-based expression vectors in plants.

Materials and methods

Enzymes

DNA polymerase I (EC 2.7.7.6) ('large fragment', according to Klenow), S1 nuclease and T4 polynucleotide kinase were from Boehringer Pharma (Mannheim, FRG).

Restriction endonucleases were from Boehringer Pharma (Mannheim, FRG) or New England Biolabs (Beverly, MA), and were used according to the suppliers' instructions.

Plants and tissue cultures for RNA isolation

The A6-S1 octopine tumor suspension culture of tobacco has been described (Willmitzer *et al.*, 1981), as well as the BT37 nopaline tobacco teratoma line (Braun and Wood, 1976).

The rGV1-derived plant used in this work was a homozygous isolate obtained after selfing of the original rGV1 regenerant (Otten *et al.*, 1981).

Isolation of poly(A)$^+$ RNA from these different sources was as described (De Greve *et al.*, 1982a).

Bacterial strains and plasmids

E. coli strain K514 r^-m^+ was used for all transformations and plasmid preparations. The *Hind*III fragment 18c (see map, Figure 1) was subcloned in pBR325 from pGV0153 (De Vos *et al.*. 1981). The resulting plasmid was designated pGV117.

S1 nuclease mapping

DNA fragments were ^{32}P-labeled either at the 5′ end using [γ-^{32}P]ATP and T4 polynucleotide kinase, or at the 3′ end using Klenow polymerase and [α-^{32}P]dNTP, the identity of which depended on the particular sequence (see details in Results section).

Strand separation was accomplished as described (Maxam and Gilbert, 1980), using acrylamide concentrations varying between 3% and 7%, according to the fragment mol. wt.

Single strands were eluted from the gel in 50 mM Tris pH 8, 20 mM EDTA, 0.5 M NaCl overnight at room temperature. The single-stranded DNA probes were hybridized to poly(A)$^+$ and subsequently digested with S1 nuclease as described (De Greve *et al.*, 1982a), unless otherwise specified in the Results section.

Sequence analysis

DNA sequences were determined using the Maxam and Gilbert (1980) procedure.

Acknowledgements

We wish to thank Dr. Michel Lemmers for computer analysis of sequence data, and Francine Deboeck for expert technical assistance. The help of Mar-

tine De Cock and Albert Verstraete in preparing the manuscript was highly appreciated. This research was supported by grants from the 'Kankerfonds van de Algemene Spaar- en Lijfrentekas', from the 'Instituut tot Aanmoediging van het Wetenschappelijk Onderzoek in Nijverheid en Landbouw' (I.W.O.N.L. 3849A), from the Services of the Prime Minister (O.O.A./12052179), and from the 'Fonds voor Geneeskundig Wetenschappelijk Onderzoek' (F.G.W.O. 3.0001.82) to J.S. and M.V.M. P.D. is indebted to the Belgian National Fund for Scientific Research for a Senior Research Assistant fellowship.

References

Bedbrook,J.R., Smith,S.M. and Ellis,R.J. (1980) *Nature,* **287**, 692-697.
Benoist,C., O'Hare,K., Breathnach,R. and Chambon,P. (1980) *Nucleic Acids Res.,* **8**, 127-142.
Berk,A.J. and Sharp,P.A. (1977) *Cell,* **12**, 721-732.
Braun,A.C. and Wood,H.N. (1976) *Proc. Natl. Acad. Sci. USA,* **73**, 496-500.
Brawerman,J. (1981) *CRC Crit. Rev. Biochem.,* **10**, 1-38.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.,* **50**, 349-383.
Chen-Kiang,S., Wolgemuth,D.J., Hsu,M.T. and Darnell,J.E.,Jr. (1982) *Cell,* **28**, 575-584.
Croy,R.R.D., Lycett,G.W., Gatehouse,J.A., Yarwood,J. and Boulter,D. (1982) *Nature,* **295**, 76-79.
De Beuckeleer,M., Lemmers,M., De Vos,G., Willmitzer,L., Van Montagu, M. and Schell,J. (1981) *Mol. Gen. Genet.,* **183**, 283-288.
De Greve,H., Dhaese,P., Seurinck,J., Lemmers,M., Van Montagu,M. and Schell,J. (1982a) *J. Mol. Appl. Genet.,* **1**, 499-512.
De Greve,H., Leemans,J., Hernalsteens,J.P., Thia-Toong,L., De Beuckeleer,M., Willmitzer,L., Otten,L., Van Montagu,M. and Schell,J. (1982b) *Nature,* **300**, 752-755.
Depicker,A., Stachel,S., Dhaese,P., Zambryski,P. and Goodman,H.M. (1982) *J. Mol. Appl. Genet.,* **1**, 561-574.
De Vos,G., De Beuckeleer,M., Van Montagu,M. and Schell,J. (1981) *Plasmid,* **6**, 249-253.
Early,P., Rogers,J., Davis,M., Calame,K., Bond,M., Wall,R. and Hood,L. (1980) *Cell,* **20**, 313-319.
Engler,G., Depicker,A., Maenhaut,R., Villarroel-Mandiola,R., Van Montagu,M. and Schell,J. (1981) *J. Mol. Biol.,* **152**, 183-208.
Fitzgerald,M. and Shenk,T. (1981) Cell, **24**, 251-260.
Garfinkel,D.J., Simpson,R.B., Ream,L.W., White,F.F., Gordon,M.P. and Nester,E.W. (1981) *Cell,* **27**, 143-153.
Gelvin,S.B., Thomashow,M.F., McPherson,J.C., Gordon,M.P. and Nester, E.W. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 76-80.
Geraghty,D., Peifer,M.A., Rubenstein,I. and Messing,J. (1981) *Nucleic Acids Res.,* **9**, 5163-5174.
Gerlach,W.L., Pryor,A.J., Dennis,E.S., Ferl,R.J., Sachs, M.M. and Peacock,W.J. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 2981-2985.
Guyon,P., Chilton,M.-D., Petit,A. and Tempé,J. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 2693-2697.
Hanukoglu,I. and Fuchs,E. (1982) *Cell,* **31**, 243-252.
Hofer,E. and Darnell,J.E. (1981) *Cell,* **23**, 585-593.
Hyldig-Nielsen,J.J., Jensen,E.Ø, Paludan,K., Wiborg,O., Garrett,R., Jørgensen,P. and Marcker,K.A. (1982) *Nucleic Acids Res.,* **10**, 689-695.
Kahl,G. and Schell,J. (1982) *Molecular Biology of Plant Tumors*, published by Academic Press, NY, p. 615.
Leemans,J., Deblaere,R., Willmitzer,L., De Greve,H., Hernalsteens,J.P., Van Montagu,M. and Schell,J. (1982) *EMBO J.,* **1**, 147-152.
Marks,H.D. and Larkins,B.A. (1982) *J. Biol. Chem.,* **257**, 9976-9983.
Maxam,A.M. and Gilbert,W. (1980) in Grossman,L. and Moldave,K. (eds.), *Methods in Enzymology, Nucleic Acids, Part I,* Vol. **65**, , Academic Press, NY, pp. 499-559.
Messing,J., Geraghty,D., Heidecker,G., Hu,N.-T., Kridl,J. and Rubenstein, I. (1983), in Hollaender,A. (ed.), *Genetic Engineering of Plants*, Plenum Press, NY, in press.
Otten,L., De Greve,H., Hernalsteens,J.P., Van Montagu,M., Schieder,O., Straub,J. and Schell,J. (1981) *Mol. Gen. Genet.,* **183**, 209-213.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature,* **263**, 211-214.
Setzer,D.R., McGrogan,M., Nunberg,J.H. and Schmike,R.T. (1980) *Cell,* **22**, 361-370.
Setzer,D.R., McGrogan,M. and Schimke,R.T. (1982) *J. Biol. Chem.,* **257**, 5143-5147.
Schröder,G. and Schröder,J. (1982) *Mol. Gen. Genet.,* **185**, 51-55.
Tempé,J. and Goldmann,A. (1982), in Kahl,G. and Schell,J. (eds.), *Molecular Biology of Plant Tumors,* Academic Press, NY, pp. 427-449.
Thomashow,M.F., Nutter,R., Montoya,A.L., Gordon,M.P. and Nester, E.W. (1980) *Cell,* **19**, 729-739.
Tosi,M., Young,R.A., Hagenbüchle,O. and Schibler,V. (1981) *Nucleic Acids Res.,* **9**, 2313-2322.
Valenzuela,P., Quiroga,M., Zaldivar,J., Rutter,W.J., Kirschmer,M.W. and Cleveland,D.W. (1981) *Nature,* **289**, 650-655.
Weaver,R. and Weissmann,C. (1979) *Nucleic Acids Res.,* **7**, 1175-1193.
Wiborg,O., Hyldig-Nielsen,J., Jensen,E., Paludan,K. and Marcker,K. (1982) *Nucleic Acids Res.,* **10**, 3487-3494.
Willmitzer,L., Schmalenbach,W. and Schell,J. (1981) *Nucleic Acids Res.,* **9**, 4801-4812.
Willmitzer,L., Simons,G. and Schell,J. (1982) *EMBO J.,* **1**, 139-146.
Ziff,E.B. (1980) *Nature,* **287**, 491-499.