Structure of the octopine synthase upstream activator sequence

(Agrobacterium tumefaciens/transcriptional activator)

SCOTT M. LEISNER AND STANTON B. GELVIN*
Department of Biological Sciences, Purdue University, West Lafayette, IN 47907

Communicated by H. Edwin Umbarger, December 31, 1987 (received for review November 9, 1987)

ABSTRACT We have identified a transcriptional activating element within the 5' flanking sequence of the Agrobacterium tumefaciens octopine synthase (ocs) gene that is necessary for ocs expression in transformed tobacco calli. This element is located between 333 and 116 base pairs upstream from the transcription initiation site and functions independent of orientation when placed upstream of the ocs gene. It does not function in either orientation when placed downstream of the gene, nor can it activate its promoter when separated by a distance of 608 base pairs. Deletion analysis indicates that sequences essential for activator function are localized between 222 and 177 base pairs upstream of the transcription initiation site. Another region, located between 333 and 249 base pairs upstream of the transcription initiation site, does not as a monomer activate the ocs promoter, but it can as a dimer.

Transcriptional control regions of eukaryotic genes consist of three major classes of sequences: (i) proximal control signals, (ii) distal control signals, and (iii) activating sequences (1–3). Proximal control signals, such as the "TATA" box, are located =30 base pairs (bp) upstream of the transcription initiation site. Distal control signals, such as the "CAATTT" box, are located 50–90 bp upstream of the transcription start. Activating sequences, such as transcriptional enhancers and the upstream activating sequences of yeast (4, 5), are usually located >100 bp upstream of the transcription start site. These sequences function in both orientations upstream of homologous and heterologous promoters to enhance transcription. Many activators can also confer tissue specificity or inducibility on promoters (1–3). Although many activator sequences have the ability to function over long distances, only enhancers stimulate transcription when placed downstream of genes.

Transcriptional activators are composed of short conserved sequence motifs that serve as transcription factor binding sites (1–3, 6). Two common motifs are the simian virus 40 (SV40) enhancer core GTGGAWW (where W is either A or T) and alternating stretches of purines and pyrimidines (with the potential to form Z-DNA).

Workers in our laboratory have been interested in the transcriptional regulation of genes transferred from Agrobacterium tumefaciens into plant cells, where they are expressed (7). One of these is the octopine synthase (ocs) gene that is constitutively expressed in a wide variety of plant tissues (8). This gene has a TATA box situated 32 bp upstream of the transcription initiation site (Fig. 1) (9, 10). In addition, sequences located between 170 and 294 bp upstream from the transcription start are also required for octopine synthase activity in tobacco crown gall tumors (11). This region contains two sequence motifs that are homologous to the SV40 enhancer (6), including an SV40 enhancer core sequence (GTGGAAAG) and a 10-bp stretch of alternating purines and pyrimidines capable of forming Z-DNA (Fig. 2).

1). The presence of these motifs in an upstream region required for transcription of the ocs gene suggested to us that the nucleotides between –170 and –294 may contain a transcriptional enhancer-like element. In this report, we show that the ocs element activates its own promoter in a manner similar to other activator sequences. We also show that the ocs activator has a complex structure.

MATERIALS AND METHODS

Materials. Restriction endonucleases were purchased from Bethesda Research Laboratories and used according to the manufacturer's specifications. DNA polymerase I Klenow fragment and T4 DNA ligase were obtained from Pharmacia and used according to the manufacturer's specifications. [α-32P]dCTP was purchased from Amersham. DNA fragments were labeled with an Amersham nick-translation kit. Reagents for the octopine synthase assay and antibiotics were purchased from Sigma. Reagents for determining protein concentrations were purchased from Bio-Rad and used according to the manufacturer's specifications.

Strains and Culture Conditions. Escherichia coli strains were grown in LB medium (13). Agrobacterium strain LBA4404 (14) was grown in AB minimal medium (15). Antibiotic concentrations, when used, were for E. coli: kanamycin, 50 μg/ml; ampicillin, 100 μg/ml; for A. tumefaciens: kanamycin, 100 μg/ml; rifampicin, 10 μg/ml. These experiments were conducted under Pl containment conditions as specified by the National Institutes of Health Recombinant DNA Guidelines.

Construction of Plasmids to Test the Effects of Position and Orientation on the ocs Gene Activator. The ocs structural gene is located within a 2.57-kbp BamHI/Sma I subfragment [base pairs 13774–11207 (10)] of BamHI fragment 17 [base pairs 13774–9062 (10)] from the Ti plasmid pTiA6. This 2.57-kbp fragment was cloned into the BamHI and Sma I sites of the binary plasmid pBin19 (16) to yield the plasmid pOCSΔ2 (Fig. 2B). This plasmid contains 116 bp upstream from the transcription start, the entire ocs structural gene, and ~1200 bp downstream from the polyadenylation site.

The activator was isolated as a 217-bp Acc I/BamHI subfragment [base pairs 13991–13774 (10)] of BamHI fragment 2 (10). This fragment contains the sequence from –333 to –116 bp relative to the ocs mRNA initiation site (9). The Acc I site was converted to a BamHI site by linker attachment (15) and the fragment was cloned, in both orientations, into the unique BamHI site upstream of the ocs gene in pOCSΔ2. Alternatively, the ends were filled in with Klenow fragment and the resulting fragment was cloned into the unique Sma I site downstream of the ocs gene (Figs. 2 and 3). Cloning the activator sequence upstream of the ocs gene in the correct orientation (pEN1) reconstitutes the ocs 5' flanking sequence to –333 (Fig. 3). pENR1 contains the

Abbreviations: SV40, simian virus 40; CAT, chloramphenicol acetyltransferase.

*To whom reprint requests should be addressed.
activator in the inverted orientation at the 5' end, pEN2 contains the activator in the native orientation but downstream of the ocs gene, pENR2 contains the activator in the inverted orientation downstream of the ocs gene, and pENR22 contains a tandem duplication of the activator in the inverted orientation downstream of the ocs gene (Fig. 3).

The 2.57-kbp BamHI/Smal I fragment harboring the ocs gene was cleaved in the TATA box of the ocs promoter and 160 bp downstream from the polyadenylation site by Dra I (Fig. 2). This 1.5-kbp Dra I fragment [base pairs 13685–12221 (10)] was cloned into the unique Smal I site of pBin19 (16) to yield the plasmid pOCsA1 (Fig. 2A). HindIII linkers were attached to the 217-bp activator and the resulting fragment was cloned into the HindIII site of pOCsA1, in the correct orientation, to yield the plasmid pENΔ1 (Fig. 3). pENΔ1 contains the entire activator and coding sequences of the ocs gene but has a truncated promoter (Fig. 3).

A 608-bp HindIII/Bgl II fragment [base pairs 6631–6023 (10)] from EcoRI fragment 7 [EcoRI fragment B; base pairs 5545–12823 (10)] was cloned into the unique HindIII and BamHI sites of pOCsAΔ2 to give the plasmid pSPOC4. This fragment is derived from the protein-coding region of the tmsl gene (10, 17). HindIII linkers were attached to the 217-bp activator fragment, which was subsequently cloned into the HindIII site of pSPOC4 in the correct orientation to yield the plasmid pSPEN0.6 (Fig. 3). In this plasmid, 608 bp separate the activator from the ocs promoter. The activator with HindIII ends also was cloned into the HindIII site of pOCsAΔ2 giving rise to plasmids in which the activator is in the correct (pEN4) or inverted (pENR4) orientation. In these constructions, the activator and the promoter are separated by 30 bp.

Construction of Plasmids Harboring Deletions of the Octopine Synthase Activator. The 217-bp activator fragment was cleaved with various restriction endonucleases, BamHI linkers were added, and the resulting fragments were cloned individually into the BamHI site of pOCsAΔ2, yielding the deletions shown in Fig. 4. The two Sau3A fragments (Sau133 and Sau84) were cloned without linker attachment since they had BamHI-compatible cohesive ends.

Conjugation of the Constructs into A. tumefaciens. Plasmids were conjugated into A. tumefaciens strain LBA4404 by a
triparental mating (16) involving the *E. coli* strain harboring the recombinant plasmid, *E. coli* strain MM294/pRK2013 (18), and *A. tumefaciens* strain LBA4404. The *Agrobacterium* transconjugants were selected on AB minimal plates containing 0.5% glucose, rifampicin (10 μg/ml), and kanamycin (100 μg/ml).

**Transformation of Plants and Octopine Assays.** Tobacco leaf disks were infected with the *Agrobacterium* strains harboring recombinant plasmids by the method of Horsch et al. (19). Kanamycin-resistant calli were assayed for octopine synthase activity by the method of Otten and Schilperoot (20). Twenty different calli were assayed for each construction. Extracts contained 20–30 μg of protein as determined by the method of Bradford (21). Constructs were denoted as octopine synthase-positive when >40% of the calli analyzed exhibited detectable octopine synthase activity. Constructs denoted as octopine synthase-negative never induced calli that exhibited detectable octopine synthase activity. S1 nuclease analysis of pEN1, pENR1, and *Sau*84 dimer octopine synthase mRNA initiation sites was performed as described by Gelvin et al. (22) with strand-specific probes made from an M13 template (23).

**RESULTS**

The Upstream Region of the *ocs* Gene Contains a Transcriptional Activator Sequence. To test whether the sequence between - 294 and - 170 functions as an enhancer of the *ocs* promoter, the constructions shown in Fig. 3 were made using the binary plasmid pOCSΔ2 (Fig. 2). After mobilization of these plasmids into *A. tumefaciens* strain LBA4404, tobacco leaf disks were infected and kanamycin-resistant calli were selected (19). These calli were assayed for octopine synthase activity (20).

Calli containing the construct pOCSΔ2, that lacks the *ocs* upstream element, had no detectable octopine synthase activity (Fig. 3). When the upstream element was cloned upstream of the *ocs* gene in the correct orientation (pEN1), reconstructing the octopine synthase upstream region, octopine synthase activity was restored. Cloning the sequence upstream of the *ocs* gene in the opposite orientation (pENR1) also resulted in octopine synthase activity. S1 nuclease protection experiments indicated that the transcripts from pEN1 and pENR1 start from the correct transcription initiation site (data not shown). Together these results show that the *ocs* upstream sequences activate the *ocs* promoter independent of orientation.

When the activator sequence was cloned downstream of the *ocs* gene in either the correct (pEN2) or reverse (pENR2) orientation and introduced into tobacco calli, octopine synthase activity was not detected (Fig. 3). Similarly, when two copies of this element were cloned downstream of the *ocs* gene (pENR22), or when the activator was moved 1 kbp closer to the 3' end of the *ocs* gene (data not shown),

**Fig. 3.** The *ocs* activator sequence functions in both orientations at the 5' end of the *ocs* gene. Electrophoretogram of octopine synthase assays performed on calli incited with the constructions noted above. Positions of octopine (OCT) and arginine (ARG) are marked. Numbers above indicate the time of incubation in hours. Origin (+) is at the bottom of the photo. A schematic diagram is shown of the plasmids pOCSΔ2, pEN1, pENR1, pEN2, pENR2, pENR22, pENΔ1, pEN4, pENR4, and pSPEN0.6 described in the text. (2), pocs promoter (~116 to ~1); ( ), pocs gene (~1 to +1273) that includes the 5' untranslated sequence, the *ocs* structural gene, and the 3' untranslated sequence. Blank space after arrow represents the 1153 bp downstream of the polyadenylation site; ( ), *ocs* activator (~333 to ~116). Thirty base pairs separates the *ocs* activator from the *ocs* promoter in the constructs pEN4 and pENR4, and 608 bp separates the activator and promoter in pSPEN0.6. Figure is not drawn to scale. Twenty calli were analyzed for *ocs* activity for each construction. The octopine in the spots shown is approximately 6000 pmol for pEN1, 6200 pmol for pENR1, 6500 pmol for pEN4, 6000 pmol for pENR4, and 0 pmol for all of the remaining constructions. Octopine standards used for quantitating the levels of octopine synthesized by extracts from calli incited by various constructions are shown below. The amount (pmol) of octopine present within each spot is indicated.

---

**OCS Activity**

<table>
<thead>
<tr>
<th>Construct</th>
<th>OCS Activator</th>
<th>OCS Promoter</th>
</tr>
</thead>
<tbody>
<tr>
<td>pOCSΔ2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pEN1</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>pENR1</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>pEN2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pENR2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pENR22</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pENΔ1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pEN4</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>pENR4</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>pSPEN0.6</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

---

**Gene Contains**

**Region**

**Activity**

**Arrow**

**OD**

**Scale**

**Fig. 2**. Schematic diagram showing the positions of the octopine synthase positive region and the upstream activator. (A) The minimal promoter (pOCSA2) with a 3' untranslated sequence of 294 bp. (B) The upstream activator (ocs+0 to +1273), which includes the whole 5' untranslated sequence, the structural gene, and the 3' untranslated sequence, separated by 30 bp and 608 bp from the promoter on the 3' side. (C) The upstream activator containing an additional 608 bp of the 5' untranslated sequence; (D) Schematic diagram showing the predicted positions of the octopine synthase and arginine promoters with respect to the octopine synthase coding region. (E) Schematic diagram showing the positions of the octopine synthase and arginine promoters with respect to the octopine synthase coding region. (F) The upstream activator containing an additional 608 bp of the 5' untranslated sequence; (G) Schematic diagram showing the predicted positions of the octopine synthase and arginine promoters with respect to the octopine synthase coding region. (H) The upstream activator containing an additional 608 bp of the 5' untranslated sequence; (I) Schematic diagram showing the predicted positions of the octopine synthase and arginine promoters with respect to the octopine synthase coding region.
octopine synthase activity was not detected. Therefore, the ocs upstream element does not greatly stimulate octopine synthase activity downstream of the ocs gene. Alternatively, the lack of stimulation of ocs gene activity could also be explained by a distance effect (see below). In our assay system, the limit of detection of octopine is 30 pmol (however, it is difficult to photograph spots of <300 pmol). By comparing the spot intensity from calli harboring constructions designated as octopine synthase positive with known standards, 2500–8100 pmol of octopine was synthesized during a 2-hr incubation.

The ocs activator may function in both orientations upstream of the ocs gene because it is a bidirectional promoter. To test this hypothesis, a plasmid was constructed (pENΔ1) containing the activator but lacking the ocs promoter (those sequences between –116 and –28, including the TATA element). Octopine synthase activity was not detected in calli harboring this construct (Fig. 3). Activator function is therefore dependent on promoter-proximal functions.

To test whether the ocs activator can function over a long distance from the ocs promoter, the plasmids shown in Fig. 3 were constructed. When the activator was moved from its normal position to a position 30 bp further upstream of the ocs promoter (pEN4 and pENR4), it still functioned in both orientations. However, when a 608-bp fragment was inserted between the activator and the promoter of the ocs gene, octopine synthase activity was abolished. These results agree with the findings of DeGrove et al. (9), who showed that an insertion of a 2500-bp fragment into the BamHI site (which lies between the ocs activator and promoter) resulted in the loss of octopine synthase activity. These data indicate that the ocs activator is sensitive to its distance from the ocs promoter, although it is possible that the spacer contains sequences that inhibit transcription.

**Structure of the Octopine Synthase Upstream Activator Sequence.** To determine which sequences are necessary for activator function, various restriction endonuclease fragments of the ocs activator were cloned into pOCSΔ2 (Fig. 4). When the Sau133 fragment (–249 to –116 from the ocs transcription initiation site) was cloned upstream of the ocs gene, octopine synthase activity was observed. This fragment does not contain the SV40 enhancer core sequence. Octopine synthase activity was also observed when the Alu106 fragment (–222 to –116) was cloned upstream of the ocs gene. No activity was observed, however, when the RsaI fragment (–177 to –116) was cloned onto the ocs gene. Thus, a 45-bp fragment (–222 to –177 from the transcription initiation site) contains an element necessary for the expression of octopine synthase activity in tobacco calli. Recent experiments indicate that this 45-bp fragment, when cloned into the BamHI site of pOCSΔ2, can activate the ocs promoter (data not shown). When the Sau84 fragment (–333 to –249) was cloned onto the ocs gene as a tandem dimer, octopine synthase activity was observed. The nuclease protection experiments indicated that the transcripts from this construct start from the correct initiation site (data not shown). However, a monomer of the Sau84 fragment failed to stimulate octopine synthase activity.

**DISCUSSION**

The upstream region of the octopine synthase gene contains a transcriptional activator element that activates the ocs promoter independent of orientation. However, the ocs element does not stimulate the ocs promoter when placed downstream of the ocs gene. If this is not simply due to a distance effect, then the ocs activator is unlikely to be classical animal viral and cellular enhancers and functions more like the upstream activating sequences of yeast (4, 5).

Ellis et al. (12, 24) showed that sequences upstream of the ocs gene can activate a heterologous promoter. The chloramphenicol acetyltransferase (CAT) gene under control of the maize alcohol dehydrogenase (ADH1) promoter is poorly expressed in tobacco. When the ocs activator was cloned, in either orientation or at either end of the CAT gene, CAT activity was expressed at high levels in tobacco. The ocs activator therefore could stimulate the ADH1 promoter. The discrepancy between the data of Ellis et al. (12) and our data concerning the effectiveness of the ocs activator downstream of a gene may be explained in at least three ways. First, our octopine synthase assay is less sensitive than the CAT assay used by Ellis et al. (12, 24). Second, it is possible that different factors affect transcription in the transient expression system used by Ellis et al. (12) and our stable expression system. Finally, the ocs upstream element may activate the ocs and ADH1 promoters in different ways. It is known that the SV40 enhancer can activate various promoters to different extents (25).

Only a few transcriptional activator sequences have been identified in genes expressed in plants, including the ribulose-1,5-bisphosphate carboxylase small subunit (rbcS) genes (26, 27), the chlorophyll a/b binding protein (cab)
genes (28, 29), the napaline synthase (nos) gene (30), the cauliflower mosaic virus 35S enhancer (31, 32), and now the ocs gene. All of these viruses stimulate homologous and heterologous promoters in an orientation-independent manner. The activator present within the upstream region of the nos gene (30) contains a stretch of alternating purines and pyrimidines homologous to that found within the ocs activator. The ocs element also resembles the cab and rbcS activators in its apparent inability to function when placed downstream of the ocs promoter (26–29). When placed downstream of reporter genes, the rbcS and cab activators function weakly, if at all. This is interesting in light of the fact that the ocs, rbcS, and cab activators all contain SV40 enhancer core sequences.

Unlike the cab (28, 29), rbcS (26, 27), and cauliflower mosaic virus 35S (31, 32) activators, the ocs activator is distance dependent (Fig. 3). These findings agree with the work of Ellis et al. (12), in which the ability of the ocs activator to stimulate the ADHI promoter decreases rapidly with increasing distance. Thus, the ocs element is not a classical transcriptional enhancer.

Our results suggest that an activator of the ocs gene is localized between −222 and −177. This region contains a 16-bp palindrome described by Ellis et al. (12) that is essential and sufficient for activating the ADHI promoter. It has not been determined whether the entire palindrome and truncated copies can stimulate the ocs promoter. A dimer of the sequence from −333 to −249 that is upstream of the 16-bp palindrome can activate the ocs gene. Because this upstream region is only functional as a dimer, it may or may not contribute to ocs expression in the normal situation where there is only one copy. It is possible that another, although artificial, activator is produced by dimerization of this sequence. S1 nuclease analysis of transcripts from calli harboring this construction indicates that the normal transcription start site is utilized. The Sau84 dimer, therefore, does not function as a promoter.

The presence of several potentially active domains within the ocs activator may, however, be similar to the situation in the SV40 enhancer (2). Mutations in any two of the three domains greatly impair the ability of the enhancer to activate the SV40 early promoter. However, activity can be restored when the remaining domain is duplicated. The SV40 enhancer core sequence, present within the promoter-distal portion of the ocs activator, may be important for activator function.

Because the ocs gene is expressed in a wide variety of plant tissues and species (8, 33), one could speculate that the upstream region (−333 to −249) that functions as a dimer may modulate the activity of the activator sequence (−222 to −177) in different plant tissues or species. It is also possible that this region may be required for ocs expression under conditions different from those studied in this paper. Future experiments should address the potential complexity of the ocs activator and its relationship to other plant activating sequences.

The authors thank Brad Goodner, Dr. Scott Boyer, and Dr. K. Veluthambi for critical reading of the manuscript; Mar Wyncott for help with plant tissue culture; and Wilma Fouseth for help in preparing the manuscript. This work was funded by Agrigenetics Research Associates and the National Science Foundation (DMB-8545586).