Characterization of regulatory elements within the coat protein (CP) coding region of *Tobacco mosaic virus* affecting subgenomic transcription and green fluorescent protein expression from the CP subgenomic RNA promoter

Michal Man and Bernard L. Epel

A replicon based on *Tobacco mosaic virus* that was engineered to express the open reading frame (ORF) of the green fluorescent protein (GFP) gene in place of the native coat protein (CP) gene from a minimal CP subgenomic (sg) RNA promoter was found to accumulate very low levels of GFP. Regulatory regions within the CP ORF were identified that, when presented as untranslated regions flanking the GFP ORF, enhanced or inhibited sg transcription and GFP expression. Full GFP expression from the CP sgRNA promoter required more than the first 20 nt of the CP ORF but not beyond the first 56 nt. Further analysis indicated the presence of an enhancer element between nt +25 and +55 with respect to the CP translation start site. The inclusion of this enhancer sequence upstream of the GFP ORF led to elevated sg transcription and to a 50-fold increase in GFP accumulation in comparison with a minimal CP promoter in which the entire CP ORF was displaced by the GFP ORF. Inclusion of the 3′-terminal 22 nt had a minor positive effect on GFP accumulation, but the addition of extended untranslated sequences from the 3′ terminus of the CP ORF downstream of the GFP ORF was basically found to inhibit sg transcription. Secondary structure analysis programs predicted the CP sgRNA promoter to reside within two stable stem–loop structures, which are followed by an enhancer region.

INTRODUCTION

Over the past decade, plant RNA viruses have been exploited as vectors for foreign gene expression in plants. Generally, the gene of interest has been cloned into the viral genome and the large quantities of viral mRNA obtained in the cytoplasm during virus infection have served as the template for foreign gene expression leading to the accumulation of the desired protein in the infected tissue. This strategy has the potential virtue of economical large-scale production of proteins with commercial and therapeutic merits (Pogue et al., 2002; Yusibov et al., 1999). One of the most employed viruses for conducting this approach is the well-characterized single-stranded RNA virus *Tobacco mosaic virus* (TMV) (Dawson et al., 1989; Donson et al., 1991; Kumagai et al., 1993, 1995; Takamatsu et al., 1987; Turpen, 1999). The genome of TMV consists of 6395 nt encoding at least four characterized proteins (Dawson & Lehto, 1990; Okada, 1999; Mushegian & Shepherd, 1995): two proteins of 126 kDa and 183 kDa, termed the replicase, are directly translated from the genomic RNA and are necessary for virus replication. The two other proteins, the 30 kDa movement protein (MP), which functions in cell-to-cell spread of infection, and the 17 kDa coat protein (CP), which is essential for viral RNA encapsidation and systemic movement in the vascular system, are translated from subgenomic (sg) promoters. At the 5′ and 3′ termini of the virus are untranslated regions (UTRs) that function in regulation of replication and translation and contribute to RNA stability (Gallie et al., 1991; Boyer & Haenni, 1994; Mushegian & Shepherd, 1995; Buck, 1999). Both terminal regions are required for recognition by the replicase to initiate minus- and plus-sense RNA synthesis.

In preliminary studies involving a TMV-based expression vector, or replicon, in which the ORF for the green fluorescent protein (GFP) replaced the complete CP ORF, measurable GFP levels were not detected in inoculated *Nicotiana tabacum* leaves, although replication ability was conserved, as determined by the formation of necrotic lesions following inoculation of susceptible host plants with replicon-derived *in vitro* transcripts. This result led us to assume that, along with the deletion of the CP coding region, regulatory elements that are crucial to transcription and/or translation of the CP sgRNA were also eliminated. It was assumed that if regulatory elements do exist, they would most likely be situated in the 5′ region of the CP ORF due to...
its proximity to the CP promoter and/or within the 3’ region of the CP ORF due to its proximity to the 3’UTR, which is a regulatory sequence by itself. This hypothesis was supported by previous studies which showed that replacing the CP ORF with a foreign ORF resulted in diminished levels of the foreign protein in comparison with the native CP (Donson et al., 1991). Our hypothesis was also supported by the results of Shivprasad et al. (1999) and Grdzelishvili et al. (2000), which were published during the course of our work and presented qualitative evidence for the presence of regions within the CP ORF of TMV that function as part of the CP sgRNA promoter. In the present study, we have extended their findings and have presented a quantitative statistical analysis testing the contribution of different regulatory elements situated within the CP coding region to sg transcription and protein expression.

METHODS

Transcriptional replicon construction. TMV (Genebank accession no. V01408) sequence data and numbering are in accordance with Goelet et al. (1982). The cDNA of the TMV mutant pU3/1728

A

B

C

5’ 20-GFP-50 3’ (TMV 5712-5731)*-GFP-(TMV 6142-6191)
5’ 0-GFP-22 3’ (0)-GFP-(TMV 6170-6191)
5’ 20-GFP-0 3’ (TMV 5712-5731)*-GFP-0
5’ 20-GFP-22 3’ (TMV 5712-5731)*-GFP-(TMV 6170-6191)
5’ 56-GFP-22 3’ (TMV 5712-5767)*-GFP-(TMV 6170-6191)
5’ 81-GFP-22 3’ (TMV 5712-5792)*-GFP-(TMV 6170-6191)
5’ 56-GFP-50 3’; (TMV 5712-5767)*-GFP-(TMV 6142-6191)
5’ 56-GFP-424 3’ (TMV 5712-5767)*-GFP-(TMV 5768-6191)
5’ 56-Rev GFP-22 3’ (TMV 5712-5767)*-GFP in reverse orientation-(TMV 6170-6191)
5’ (55→1)-GFP-50 3’ (TMV 5766-5712)**-GFP-(TMV 6142-6191)
5’ 23/(55→25)-GFP-50 3’ (TMV 5712-5734**; 5766→5736)-GFP-(TMV 6142-6191)

* nt 5713, ** nt 5723; U mutated to C
D

<5’ 0 nts>
5’ GC

<5’ 20 nts>
5’ ac*(u)guuuacaguacacuacGC 3’

<5’ 56 nts>
5’ ac*(u)guuuacaguacacuaccuccaguguguuugucacaguacgugggcG 3’

<5’ 81 nts>
5’ac*(u)guuuacaguacacuaccuccaguguguuugucacaguacgugggcG 3’

<5’ 55 nts in reverse orientation>
5’ gggugacacucugcuugcuugacuacucaccacacacuac* (u)gacauucuguaGC 3’

<5’ 23 nts/ inverted 31 nts>
5’ ac*(u)guuuacaguacacuaccuccgguugacacucuguucuugugacugG 3’

<3’ 0 nts>
5’ UAAAGGCCGCGTCGACCTAGGCGATATCGGGACGTGA 3’

<3’ 22 nts>
5’ UAAAGGCCGCCCCAGGUGCaccucctggcugcaacuaga 3’

<3’ 50 nts>
5’ UAAAGGCCGCGCCAGGcuucagagccuuucugwuuggwuuggacucgccугугacuauga 3’

Fig. 1. (A, B) Schematic diagram of the genomic organization of TMVΔMP (A) and a TMV-based replicon (B). Only modified regions are presented. Numbering is extrapolated from the TMV U1 sequence (Goelet et al., 1982). Figures are not drawn to scale. (A) Nt 1 and 6395 denote the 5’- and 3’-terminal nucleotides of the TMV genome; nt 69 and 4919 denote the replicase start and stop codon positions, respectively; nt 4903 and 5709 denote the MP start and stop codon positions, respectively; nt 5712 and 6191 denote the CP start and stop codons, respectively. Δ4923–5402 denotes the nucleotide sequence deleted from the MP ORF. (B) Design of the TMV-based replicons. The example shown is 5’ 56-GFP-22 3’. The numbering indicates the number of nucleotides from the CP ORF presented as UTRs upstream and downstream of the GFP ORF. The position of the replicase is indicated by a triangle. (C) The replicons are designated according to the length of untranslated sequences derived from the 5’ and 3’ termini of the CP ORF, situated upstream and downstream of the GFP ORF, respectively. (D) Sequences of the CP 5’- and 3’UTRs flanking the GFP ORF, verified by restriction map analysis and sequencing. The CP sequences are shown in lower case. Inverted CP sequences are in bold. Non-CP sequences derived from the polyclonal site are shown in capital letters. In the 5’UTR, AUGs were mutated to ACGs, shown as * (u). Capital underlined GC are part of the Nol site where GFP was inserted. UAA in the 3’UTRs indicates the stop codon of GFP.
Plant lines. N. tabacum cv. Xanthi line 277 (kindly provided by Dr R. N. Beachy), a systemic host line transgenic for the MP gene of TMV strain U1, (Deom et al., 1987), referred to as N. tabacum MP*, was grown at a constant temperature 25°C with 16 h of light and 8 h of darkness.

**In vitro transcription and inoculation.** After linearization of the replicons with KpnI, in vitro transcripts were synthesized using T7 RNA polymerase, according to the protocol of the mMessage mMachine Transcription kit (manual version 0004; Ambion). Transcripts were incubated directly onto the adaxial surfaces of N. tabacum MP* leaves (leaves 3 and 4 from the bottom) in 6- to 7-week-old plants (inoculated leaves were not fully expanded). Immediately after inoculation, leaves were rinsed with water and the plants placed in a growth chamber.

**RNA inoculation and Northern blotting.** Total RNA was purified at 7 days post-infection (p.i.) from fluorescent areas of leaves inoculated with in vitro transcripts of GFP-expressing replicons. RNA (10 μg per lane) was analysed by Northern blotting using standard methods (Sambrook et al., 1989). Blots were probed with a PCR-synthesized digoxigenin-labelled probe complementary to the GFP sequence and exposed to X-ray film to record the chemiluminescent signal, following the protocol of the digoxigenin system user's guide for filter hybridization (Boehringer). Bands were quantified by scanning and densitometry with the ImageMaster 1D prime program, version 3.01 (Amersham Pharmacia Biotech).

**Protein extraction and analysis.** Total soluble proteins were extracted 7 days p.i. from fluorescent areas of leaves inoculated with in vitro transcripts produced from GFP-expressing replicons: infected leaf tissue (60 mg) was frozen in liquid nitrogen, pulverized while frozen and then ground with a pestle and pre-chilled mortar in 4 ml ice-cold extraction buffer (final concentrations: 20 mM Tris/HCl, pH 8.5, 0.25 M sucrose, 0.02% sodium azide, 4 mM EDTA, 10 mM EGTA plus the protease inhibitors 1 mM aprotinin, 1 mM leupeptin, 2 mM PMSF and 14 μM pepstatin). Extract was cleared of particulate matter by centrifugation at 34,000 g for 15 min at 4°C and relative GFP levels were quantified in the clear supernatant using a Perkin-Elmer spectrofluorometer model LS50B with excitation at 485 nm (5 nm slit) and emission at 509 nm, as described by Epel et al. (1996). Relative GFP fluorescence values were calculated by subtracting background autofluorescence. Fluorescence intensity was measured in the linear zone where fluorescence is directly correlated to protein level.

**Construction of expression constructs for in vitro translation.** To investigate a possible effect of the regulatory elements on translation independent of their effect on transcription, expression constructs (ECs) for in vitro translation were synthesized to produce similar transcription levels for all constructs tested from a T7 promoter rather than from the CP sgRNA promoter. Downstream of the T7 promoter, the 17 nt region upstream of the CP transcription start codon was inserted followed by (i) the native CP transcript leader, defined as the first 9 nt downstream of the transcription start site (Guilley et al., 1979); (ii) the putative regulatory elements from the CP cistron, created as a UTR by mutating the ATG start codon to ACG; (iii) the GFP ORF; or (iv) the 3′UTR. Each EC was based on its equivalent expression replicon (Fig. 1C), resulting in a series of ECs in which the GFP ORF was flanked by different putative regulatory elements from the CP cistron (see description of ECs below); each EC fragment was amplified by PCR using the corresponding transcriptional replicon as template, a forward primer containing a SacI site followed by the nucleotides corresponding to TMV nt 5686–5703 and a T3 (Stratagene) reverse primer corresponding to the T3 promoter flanking the KpnI site at the 3′ terminus of the replicon (see Fig. 1A). The obtained PCR product was digested with SacI and KpnI and ligated into the SacI/KpnI-digested KS plasmid (Stratagene) downstream of the T7 promoter. Equal samples of transcripts synthesized from the T7 promoter in each EC were subsequently tested for their translational capacity using an in vitro translation system.

The ECs were delineated as follows: EC 5′ 56-GFP–22 3′: T7 promoter::(TMV 5686–5767)–GFP–(TMV 6142–6395); EC 5′ 20–GFP–50 3′: T7 promoter::(TMV 5686–5731)–GFP–(TMV 6142–6395); EC 5′ 20–GFP–22 3′: T7 promoter::(TMV 5686–5731)–GFP–(TMV 6170–6395); EC 5′ 56-GFP–50 3′: T7 promoter::(TMV 5686–5767)–GFP–(TMV 6142–6395); and EC 5′ 0–GFP–22 3′: T7 promoter::(5686–5711)–GFP–(TMV 6170–6395). The asterisk indicates that nt 5713 has been mutated from U to C.

**In vitro translation.** Translation of equal amounts (600 ng) of in vitro transcripts from ECs was performed using a rabbit reticulocyte lysate in vitro translation system, according to the manufacturer’s protocol (Promega). To allow non-radioactive detection of proteins synthesized in vitro, a transcend non-radioactive translation detection tRNA composed of a biotinylated lysine (Promega) was added to the translation reaction as a precharged ε-labelled biotinylated lysine–RNA complex (transcend tRNA) rather than a free amino acid. After translation, lysate samples (1 μl) were subjected to SDS-PAGE followed by transfer to a PVDF membrane. The biotinylated proteins were visualized by binding with streptavidin–alkaline phosphatase, followed by colorimetric (BCIP/NBT) detection, according to the protocol for the transcend non-radioactive translation detection system (Promega).

**Fluorescence microscopy.** Infected leaf tissues were analysed using a Leica DMBRE fluorescence microscope equipped with a BP 450–490 (FITC) excitation filter, an RKP dichromatic mirror and a BP 515–560 suppression filter. Images were recorded using a Sensicam 12 bit cooled imaging camera and processed using the PMIS software package (Photometrics).

**Computer prediction of the RNA secondary structure.** The MFold program followed by the PlotFold program from the GCG (Genetics Computer Group, University of Wisconsin) package was used for secondary structure prediction of the sgRNA promoter region followed by the CP ORF in TMVAMP and of the modified CP sgRNA promoter regions followed by the GFP ORF in various GFP-expressing replicons.

**RESULTS**

**Regulatory elements derived from the CP coding region affect the accumulation of GFP in N. tabacum MP* leaves inoculated with in vitro transcripts of replicons**

A series of TMV-based replicons for transient gene expression was created. To ensure the safety of the environment should these replicons be employed in the future as a tool for expressing heterologous proteins in plants, large portions of the MP and CP ORFs were excised to prevent their expression, leaving only the replicase gene retained intact. The GFP ORF was then inserted downstream of the CP sgRNA promoter, in place of the completely or partially deleted CP ORF (Fig. 1B). To elucidate possible regulatory roles of specific sequences from the CP ORF, each replicon contained different sequences derived from the 5′ and 3′
termini of the CP ORF, positioned as UTRs upstream and downstream of the GFP ORF, respectively (Fig. 1C and D). Leaves of *N. tabacum*, which are transgenic for the TMV MP gene (*N. tabacum MP*+), enable local cell-to-cell spread of replicons following inoculation with replicon transcripts. The accumulation of GFP was followed by fluorescence microscopy and at 7 days p.i., infected sites expressing GFP were harvested, homogenized and relative GFP fluorescence in the cleared supernatants was measured. The results are summarized in Fig. 2 and shown in Fig. 3. It has been shown previously that the minimal CP promoter extends into the first three nucleotides of the CP ORF, i.e. the start codon (Grdzelishvili et al., 2000). In our system, replicon 5′-0-GFP-0 3′ was designed so that the start codon of the GFP ORF would replace the start codon of the CP ORF. However, in *N. tabacum MP*+ leaves infected with replicon 5′-0-GFP-0 3′, GFP fluorescence was not visible, although its replication capability was retained as indicated by local lesion assays. Only with the addition of the 3′-terminal 22 nt of the CP ORF as a UTR downstream of the GFP ORF in replicon 5′-0-GFP-22 3′ was a minor accumulation of GFP detected (Fig. 2). This result suggested a positive effect of the 3′-terminal 22 nt of the CP ORF on GFP accumulation. The inclusion of the first 20 nt of the CP ORF as a UTR upstream of the GFP ORF in replicon 5′-20-GFP-22 3′ resulted in a ninefold increase in GFP accumulation (Fig. 2) with respect to replicon 5′-0-GFP-22 3′. However, in replicon 5′-20-GFP-50 3′, which contained the last 50 nt of the CP ORF as a UTR downstream of the GFP ORF, GFP accumulation decreased by 70% (Fig. 2) with respect to replicon 5′-20-GFP-22 3′. Replicon 5′-56-GFP-22 3′ yielded the highest levels of GFP (Fig. 2), which was 5.4-fold greater than replicon 5′-20-GFP-22 3′ and 50-fold greater than replicon 5′-0-GFP-20 3′. Extending the 5′ region of the CP ORF to include the first 81 nt as a UTR did not further increase the GFP accumulation level (Fig. 2) compared with

![Figure 2](http://vir.sgmjournals.org)

**Fig. 2.** Fluorescence analysis of GFP accumulation in *N. tabacum MP*+ leaves inoculated with replicon-derived transcripts. Leaf tissue expressing GFP 7 days p.i. from *in vitro* replicon transcripts was detected by fluorescence microscopy. Total soluble proteins were extracted from equal quantities of fluorescent tissue (60 mg) and relative fluorescence was determined. Values presented were calculated by subtracting background autofluorescence and represent the mean, derived from at least three independent experiments (in arbitrary units) ± SD, as indicated by error bars, measured using a one-way ANOVA test with *P* < 0.001.

![Figure 3](http://vir.sgmjournals.org)

**Fig. 3.** Visualization of infected sites produced by various replicons on *N. tabacum MP*+ leaves. All fluoromicrographs were taken at 7 days p.i. under equal exposure conditions. (A) 5′-22-GFP-0 3′; (B) 5′-22-GFP-20 3′; (C) 5′-56-GFP-50 3′; (D) 5′-23/(55→25)-GFP-50 3′; (E) 5′-55-GFP-20 3′.
replicon 5’ 56-GFP-22 3’. Thus, the presence of the first 56 nt of the CP OR as a UTR upstream of the GFP ORF gave maximal expression, while the inclusion of the first 20 nt was suboptimal. In comparison with replicon 5’ 56-GFP-22 3’, replicon 5’ 56-GFP-50 3’ accumulated 40 % less GFP (Fig. 2) and replicon 5’ 56-GFP-424 3’ accumulated 80 % less GFP (Fig. 2). These results confirmed the inhibitory effect of extended 3’-untranslated CP sequences situated downstream of the GFP ORF; the more extended the 3’ UTR, the more GFP accumulation was reduced.

To define the putative enhancer region within the promoter, we examined the hypothesis that nucleotides within the first 55 nt of the CP ORF function as an enhancer, whereas the core promoter is in the UTR upstream of the AUG. Enhancer elements elevate transcription level by several fold, irrespective of orientation (Blackwood & Kadonaga, 1998). To test our hypothesis, the replicon 5’ (55→1)-GFP-50 3’, was created in which the first 55 nt of the CP in the positive-strand were cloned in reverse orientation. Replicon 5’ (55→1)-GFP-50 3’ yielded almost undetectable levels of GFP (Fig. 2), leading to the conclusion that not all 55 nt comprised an enhancer element. To identify smaller segments that might encompass an enhancer element, we tested the hypothesis that the enhancer was situated between positions +25 and +55 with respect to the CP translation start site. In support of this assumption was our finding that the presence of this sequence as a UTR upstream of GFP increased GFP accumulation 5-4-fold (Fig. 2, compare replicons 5’ 20-GFP-22 3’ and 5’ 56-GFP-22 3’). Hence, replicon 5’ 23/ (55→25)-GFP-50 3’ was constructed in which the first 23 nt of the CP ORF were followed by the reverse orientation of nt +25 to +55 (Fig. 1C and D). It was found that replicon 5’ 23/(55→25)-GFP-50 3’ produced the same levels of protein as replicon 5’ 56-GFP-50 3’ (Fig. 2). The similar function of CP nt +25 to +55, irrespective of orientation, thus suggested that they encompassed an enhancer motif. The similar GFP accumulation levels of replicons 5’ 23/(55→25)-GFP-50 3’ and 5’ 56-GFP-50 3’ also implied that the deletion of two nucleotides at positions +24 and +56 with respect to the translation start site, which were inadvertently deleted from the former replicon, did not significantly alter promoter function.

Fig. 2. shows the mean fluorescence ± SD measured using a one-way ANOVA test (P<0.001). A Tukey test (P<0.05) was used to divide replicons further into statistically significant homogeneous groups according to mean fluorescence values. Three subsets were obtained: replicons 5’ 56-GFP-22 3’ and 5’ 81-GFP-22 3’ were grouped into a subset producing high GFP expression levels; replicons 5’ 56-GFP-50 3’ and 5’ 23/(55→25)-GFP-50 3’ were grouped into a subset producing moderate GFP expression levels; and the rest of the replicons were grouped into a subset producing low GFP expression levels. Subsequently, representative replicons were selected from each of these three distinct subsets.

---

**Regulatory elements derived from the CP coding region affect the accumulation of subgenomic RNA containing the GFP ORF in N. tabacum MP+ leaves inoculated with in vitro transcripts of replicons**

Leaves of N. tabacum MP+ were inoculated with in vitro transcripts of each replicon. Total RNA was extracted from tissue expressing GFP at 7 days p.i. and analysed by Northern blot hybridization. Fig. 4 shows a representative Northern blot analysis. Genomic and sgRNA levels were quantified with scanning and densitometry, and results were normalized against the constitutive expression of the rRNA gene (Fig. 4). The absorbance values (in arbitrary units) of the bands containing the genomic RNAs of replicons 5’ 20-GFP-22 3’, 5’ 56-GFP-22 3’, 5’ 56-GFP-50 3’ and 5’ 23/(55→25)-GFP-50 3’ were 2860, 2434, 2473 and 2720, respectively, suggesting similar levels of genomic RNA accumulation. The absorbance values of the bands containing the sgRNAs derived from the CP sgRNA promoter for the replicons indicated above were 389, 2713, 1689 and 1597, respectively. The level of sgRNA accumulation of replicon 5’ 56-GFP-22 3’ containing the first 56 nt of CP ORF was higher than that of replicon 5’ 20-GFP-22 3’ containing only the first 20 nt of the CP ORF (Fig. 4, lanes 3 and 2, respectively), indicating that the
complete CP sgRNA promoter extends further than the first 20 nt of the CP ORF (but not beyond the first 56 nt, as the addition of the sequence between positions +56 and +81, upstream of the GFP ORF, did not further increase GFP accumulation, as described above). This elevated sgRNA transcription level was also consistent with higher protein levels (Fig. 2). Our previous assumption that the sgRNA promoter activity is stimulated by an enhancer element located between nt +25 and +55 was additionally confirmed by Northern blot analysis, which revealed similar transcript accumulation levels for replicons 5' 56-GFP-50 3' and 5' 20/(55→25)-GFP-50 3' (Fig. 4, lanes 4 and 5, respectively); i.e. nt +25 to +55 were shown to function to the same degree in stimulating transcription, irrespective of orientation. Inclusion of the last 50 nt of the CP ORF downstream of the GFP ORF reduced sgRNA accumulation of replicon 5' 56-GFP-50 3' in comparison with replicon 5' 56-GFP-22 3', as shown in Fig. 4, lanes 4 and 3, respectively. Thus, it is reasonable to assume that the reduced GFP accumulation level produced by replicon 5' 56-GFP-50 3' in comparison with replicon 5' 56-GFP-20 3' (Fig. 2) is a reflection of its lower transcription level.

Fig. 5. In vitro translation of GFP. Equal amounts of EC in vitro transcripts were translated in vitro using a rabbit reticulocyte expression system followed by SDS-PAGE and detection using the transcend non-radioactive system. (A) Lane 1, EC 5' 56-GFP-22 3'; lane 2, EC 5' 20-GFP-50 3'; lane 3, EC 5' 20-GFP-22 3'; lane 4, EC 5' 56-GFP-50 3'; lane 5, negative control; lane 6, molecular mass markers. (B) Lane 1, positive control (luciferase); lane 2, negative control; lane 3, EC 5' 0-GFP-22 3'; lane 4, EC 5' 20-GFP-22 3'; lane 5, EC 5' 56-GFP-50 3'; lane 6, EC 5' 56-GFP-22 3'. Note that comparison of bands between two gels was not permissible, or necessary. Replicon 5' 56-GFP-22 3' was used as the basis for comparison between the two gels.

Fig. 6. Analysis of total soluble protein from N. tabacum MP+ leaves infected with replicon 5' 56-GFP-22 3' or with TMVΔMP by SDS-PAGE followed by Coomassie brilliant blue staining. Lane 1, molecular mass markers; lane 2, protein derived from non-infected leaves; lane 3, protein derived from leaves infected with TMVΔMP-derived transcript; lane 4, protein derived from leaves inoculated with replicon 5' 56-GFP-22 3'-derived transcript.

Fig. 7. Accumulation of RNA transcripts produced by replicon 5' 56-GFP-22 3' and by TMVΔMP in N. tabacum MP+ leaves. Northern blot hybridization of total RNA extract was probed with a digoxigenin-labelled probe complementary to the 3'UTR of the GFP sequence. Lane 1, RNA from control non-transfected leaves; lane 2, RNA from leaves infected with TMVΔMP-derived transcript; lane 3, RNA from leaves inoculated with replicon 5' 56-GFP-22 3'-derived transcript. Bands of constitutive rRNA detected with a specific probe are also shown.
Regulatory elements derived from the CP coding region do not affect GFP translation from the CP sgRNA promoter in replicons as indicated by in vitro translation

To investigate a putative role for the regulatory elements in translational control, in addition to their primary effect on transcription, in vitro sg transcripts of each EC containing the GFP ORF flanked by different untranslated CP ORF sequences were produced from a T7 promoter rather than the native CP sgRNA promoter (see Methods). An in vitro translation assay, using a rabbit reticulocyte expression system, was employed to compare translational efficiency of equal amounts of sgRNAs produced from the T7 promoter. Upon loading of equal volumes of lysate from the translation assays on SDS-PAGE, all ECs tested yielded similar levels of translated GFP (Fig. 5). These data suggested that the regulatory elements from the CP coding region do not differentially affect translation.

The GFP ORF in a TMV-based replicon is poorly expressed compared with the CP gene in TMV\text{\textcopyright}M

Finally, in an attempt to assess the effectiveness of the most productive replicon, 5' 56-GFP-22 3', its productivity was compared with that of replicon TMV\text{\textcopyright}M (Fig. 1A) expressing the CP. N. tabacum MP\textsuperscript{+} leaves were manually inoculated with in vitro RNA transcripts of one or other of the vectors. Total protein or RNA was extracted 7 days p.i. from fluorescent areas expressing GFP in inoculated leaves infected with replicon 5' 56-GFP-22 3' and 14 days p.i.

---

**Fig. 8.** Computer-predicted secondary structures generated by MFold (Zuker, 1989) followed by PlotFold analysis for the sequence of the CP sgRNA promoter (initiated at \(-157\) with respect to the transcription start site) plus the CP ORF (A) and for the corresponding sequences in GFP-expressing replicons, in which the CP ORF was completely or partially replaced with the GFP ORF: 5' 0-GFP-0 3' (B), 5' 20-GFP-22 3' (C), 5' 56-GFP-22 3' (D) and 5' 23/(55\rightarrow25)-GFP-50 3' (E). For all GFP-expressing replicons, the sequence from position \(-157\) with respect to the CP transcription start site through to the end of the GFP ORF was used for computer analysis. The minus strand is shown. Numbering is according to the positive sense. Several nucleotide positions with respect to the transcription start site are indicated with an arrow. The thick arrow denotes the GFP ORF initiation. The dashed line indicates the enhancer region. Calculated thermal stabilities for the optimal configuration predicted by a PlotFold analysis for each predicted structure are presented. Note that replicon terminology is in accordance with the translation start site, whereas in this figure positions indicated are with respect to the transcription start site.
from upper leaves infected with TMVAMP showing a mosaic crinkled-leaf pattern. The difference in the incubation period prior to tissue harvest was due to the fact that replicon 5′-56-GFP-22 3′, which was designed to be environmentally safe, lacked functional CP and MP and thus could only spread locally within the borders of an inoculated transgenic MP+ leaf. The peak for local GFP accumulation produced by replicon 5′-56-GFP-22 3′ was found to be at 7 days p.i (data not shown). However, TMVAMP, which expresses CP, could not be detected on the leaf that was initially infected, yet upon systemic spread symptoms were evident at 14 days p.i. in upper leaves. Analysis of protein extracts from infected leaves by SDS-PAGE revealed much higher levels of CP, in contrast to GFP, which was almost undetectable (Fig. 6). In accordance with this difference in protein level, Northern blot analysis of total RNA extracts detected genomic and sgRNAs for TMVAMP, while levels of RNA for 5′-56-GFP-22 3′ were below the detection level under the conditions used (Fig. 7).

**DISCUSSION**

Modification of viruses to express foreign proteins has often met with unexpectedly low yields. In TMV, as well as in many other RNA plant viruses, sgRNA promoters extend downstream of the transcription initiation site (French & Ahlquist, 1998; Marsh et al., 1988; van der Kuyl et al., 1991; Balmori et al., 1993; Boccard & Baucombe, 1993; Johnston & Rochon, 1995; van der Vossen et al., 1995; Maia et al., 1996; Wang & Simon, 1997; Shivprasad et al., 1999; Wang et al., 1999; Grdzelishvili et al., 2000; Koey & Miller, 2000). Thus, replacement of a virus ORF with a foreign ORF must be designed with caution to avoid loss of regulatory elements due to a deletion or modification of relevant tertiary structural elements. Replication and transcription promoters have been characterized for many viruses (reviewed by De Graff & Jaspars, 1994; Buck, 1996). The importance of secondary structures such as stem-loops (SLs) in sg promoters as well as in genomic replication promoters has been revealed for several positive-sense RNA viruses (Song & Simon, 1995; Zavriev et al., 1996; Stupina & Simon, 1997; Wang & Simon, 1997; Carpenter & Simon, 1998; Koey et al., 1999; Wang et al., 1999; Grdzelishvili et al., 2000; Haasnoot et al., 2000; Koey & Miller, 2000; Nagy et al., 2001). Insertion of a foreign ORF that might interfere with such structures or the deletion of sequences essential for secondary structure folding might also lead to a diminished foreign protein yield. It has been reported that several different ORFs inserted into TMV-based vectors under the control of a partially active CP sgRNA promoter produced low product accumulation in tobacco in similar transient expression systems. For example, the insertion of the chloramphenicol acetyltransferase (CAT) gene in place of the CP gene resulted in low CAT activity in comparison with the CP produced by TMV infection (Takamatsu et al., 1987). Likewise, the insertion of a foreign ORF under the control of an additional viral sg promoter (Dawson et al., 1989), or under a heterologous CP sg promoter derived from *Odontoglossum ringspot virus* (ORSV) (Donson et al., 1991; Kumagai et al., 1993), also resulted in low levels of the foreign gene expression; it was reported that the insertion of neomycin phosphotransferase II (NPTII) downstream of the TMV CP sg promoter yielded considerably lower levels of extractable NPTII in comparison with the level of TMV CP (Donson et al., 1991). The insertion of the α-trichosanthin ORF led to the accumulation of the foreign protein up to levels of at least 2% of total protein upon systemic infection (Kumagai et al., 1993); still, this yield is considerably less than that of the TMV CP, which can encompass up to 70% of protein synthesis in infected leaves (Dawson, 1992).

To analyse the correlation between the TMV CP sgRNA promoter activity and its putative secondary structure, we used the Mfold program (Zuker, 1989). Secondary structure models were generated for the native CP sgRNA promoter followed by the CP ORF, as in replicon TMVAMP (Fig. 8A), as well as for altered sgRNA promoter regions followed by the GFP ORF in GFP-expressing replicons (Fig. 8B–E). Computer analysis for the minus strand sequence of the native CP sgRNA promoter starting at position −157 with respect to the transcription start site (TSS) and extending into the CP ORF predicted multiple SLs (Fig. 8A). The promoter region folds into two SL structures: SL1 encompasses the sequence between −157 and −74 and SL2 encompasses the sequence between −70 and +34, with respect to the TSS (position +34 corresponds to nt +25 of the CP ORF with respect to the translation start site). The sequence between nt +34 and +64 with respect to the TSS containing the putative enhancer element is outside SL2 (indicated by the dashed line in Fig. 8A). The resulting secondary structure presented here differs from that previously suggested by Grdzelishvili et al. (2000), which predicted that the sgRNA promoter folded into one long SL structure. Their analysis was based on a short sequence between −100 and +52 with respect to the TSS, although they mapped the beginning of the fully active promoter to position −157 with respect to the TSS. In contrast, our computer-generated model was based on a much longer sequence, starting at position −157 and extending to the end of the CP ORF. By testing the biological activity of differently altered CP sgRNA promoters situated within various GFP-expressing replicons, we were able to predict a direct correlation between the native SL2 sequence, structure and promoter activity. The predicted secondary structure for the modified sgRNA promoter in replicon 5′-0-GFP-0 3′, which produced undetectable levels of GFP, revealed that the stem in SL2 was shortened at its terminus (Fig. 8B). Inclusion of the 20 nt from the 5′ terminus of the CP ORF, which had previously been shown to be part of the extended CP sgRNA promoter (Grdzelishvili et al., 2000), elevated to some extent promoter activity in replicons 5′-20-GFP-22 3′ and 5′-20-GFP-0 3′. Folding analysis for the modified sgRNA promoter followed by the GFP ORF in these replicons predicted a partial restoration...
of the stem in SL2 (Fig. 8C). Full sgRNA promoter activity in our system was gained with the addition of the first 56 nt from the CP ORF upstream of the GFP ORF. In replicon 5’-56-GFP-22 3’, SL1 and SL2 were predicted to fold as for the native pattern (Fig. 8D). The GFP accumulation level for replicon 5’ (55->1)-GFP-50 3’ in which the first 55 nt of the CP were positioned in an inverted-sense orientation upstream of the GFP ORF was negligible. This loss of promoter activity may have derived from a total disruption of SL2 structure, as predicted by computer analysis (data not shown). The extended 3’UTR containing the terminal 50 nt from the CP ORF apparently had no effect on the secondary structure of the CP sgRNA promoter according to computer modelling (data not shown). The SL2 structure could be important for the recognition of the viral replicase with its cognate promoter template of specific sequence and/or structure. Destabilization of SL2 could abolish this process, leading to inhibition of sgRNA synthesis. In contrast, restricted inversion of the sequence between nt +34 and +64 with respect to the TSS (+25 to +55 with respect to the CP translation start site), which encompasses the putative enhancer sequence in replicon 5’-23/(55->25)-GFP-50 3’, retained the same promoter activity as that produced by replicon 5’-56-GFP-50 3’ and the same putative folding pattern of SL1 and SL2 (Fig. 8E and D).

Our findings elaborate those of Grdzelishvili et al. (2000) who demonstrated by Northern blot analysis that the boundaries of a fully active CP promoter included CP sequences within 42 nt downstream of the CP ATG start codon. They also showed that the sequence between positions +55 and +351 of the CP ORF relative to the TSS was inhibitory to sg transcription from the CP sgRNA promoter. We can add that the last 22 nt of the CP ORF (+468 to +489 with respect to the TSS) have a minor positive effect on GFP accumulation, for as yet unknown reasons, whereas the inclusion of an additional 28 nt (+440 to +467) upstream of these 22 nt inhibited sg transcription (Fig. 4) and decreased GFP accumulation 1-6-fold (Fig. 2). Including the 424 nt of the 3’-terminal region of the CP cistron in the GFP UTR further decreased GFP accumulation up to 5-8-fold as compared with the most optimal construct (Fig. 2). The inhibitory effect of long 3’ CP sequences downstream of GFP ORF could derive from the increased distance between the CP sgRNA promoter and the 3’UTR. Culver et al. (1993) found that the vicinity of the sg promoter to the 3’UTR is a crucial factor in enhancing expression. These findings have been further supported by Szecsi et al. (1999) and by Shivprasad et al. (1999) who located the positive effect of the 3’UTR to the three pseudoknots.

Replicon 5’-56-GFP-22 3’ was the most effective of all those tested in our system, although it performed poorly in comparison with the control replicon, TMVAMP (Figs 6 and 7). The impaired genomic RNA accumulation of replicon 5’-56-GFP-22 3’ in comparison with replicon TMVAMP (Fig. 7) could explain the diminished GFP accumulation level produced by this replicon relative to the much higher CP accumulation level produced by TMVAMP. In the absence of the CP, levels of accumulation of positive-strand RNA are sometimes reduced and this has been ascribed to degradation of the RNA in the absence of the protective capsid (e.g. Beet western yellow vein virus and Cucumber necrosis virus). In other cases, there appears to be little or no effect (e.g. Brome mosaic virus) (Buck, 1996). However, in our system the lack of CP is probably not accountable for the poor functioning of the TMV-based replicon. In a concurrent study, we have assessed different attributes of the GFP-expressing replicon that affect genomic replication. It was found that the cause of the impaired genomic replication of GFP-expressing replicons derived from the presence of the GFP coding sequence inserted into the viral genome, whereas the foreign protein product and the lack of CP had no effect on genomic replication capabilities (unpublished results).

Understanding the factors that inhibit the productivity of replicons could lay the foundations for the design of an efficient TMV-based vector as a tool for foreign gene expression in the future.

ACKNOWLEDGEMENTS

This research was partially supported by Research Grant number IS-2636-95 from BARD, the United States–Israel Binational Agricultural Research and Development Fund. M. Man was partially supported by a Manna Institute doctoral fellowship. The authors would also like to thank Dr R. N. Beachy, who kindly provided pU3/12Am-RV and the seeds of N. tabacum cv. Xanthi line 277.

REFERENCES


