



## Structure and promoter/leader deletion analysis of mirabilis mosaic virus (MMV) full-length transcript promoter in transgenic plants

Nrisingha Dey and Indu B. Maiti\*

Molecular Plant Virology and Plant Genetic Engineering Laboratory, Tobacco and Health Research Institute, University of Kentucky, Lexington, KY 40546-0236, USA (\*author for correspondence)

Received 19 October 1998; accepted in revised form 8 April 1999

**Key words:** caulimovirus, mirabilis mosaic virus (MMV), transcriptional enhancer, promoter deletion, transgenic plants

### Abstract

A full-length transcript (FLt) promoter fragment was isolated from a genomic clone of mirabilis mosaic virus (MMV), a double-stranded DNA plant pararetrovirus belonging to the caulimovirus family. The boundaries required for maximal promoter expression were defined by 5' and 3' deletion analysis of the MMV promoter fragments coupled to a GUS reporter gene. The expression patterns of these chimeric gene constructs were evaluated both in transgenic *Nicotiana tabacum* cv. Samsun NN plants and in protoplast transient expression experiments. A 360 bp FLt promoter fragment (sequence –297 to +63 from the transcription start site) was found sufficient for strong promoter activity. The transcription start site (TSS) of the MMV FLt promoter was determined by primer extension analysis using total RNA isolated from transgenic plants containing a MMV promoter:*uidA* fusion gene. Analysis of the 5' and 3' deletion constructs showed that an upstream region (sequence –248 to –193 from the transcription start site) is required for the MMV FLt promoter activity along with the as-1, TATA box regions. In addition, a 31 bp sequence (+33 to +63 from the transcription start site) located downstream of a TATA box is also essential for the maximum expression of the MMV FLt promoter. Analysis of transcripts (mRNA) from these chimeric constructs also indicated that the MMV FLt promoter fragment (–297 to +63 from the transcription start site) has the highest promoter activity. In a comparative analysis the MMV FLt promoter showed much greater activity than the CaMV 35S promoter.

### Introduction

Mirabilis mosaic virus (MMV), a member of the caulimovirus family, has a circular double-stranded DNA genome of about 8 kb with four single-stranded discontinuities in the DNA, one in the  $\alpha$ -strand and three in the complementary strand (Richins and Shepherd, 1983). It infects *Mirabilis* plant species (family Nyctaginaceae) generally found in warm parts of North America. The MMV was characterized as a member of this genus of plant pararetroviruses based upon morphology of its virions and inclusion bodies (Brunt and Kitajima, 1973). MMV is serologically distinct from cauliflower mosaic virus, the type member of the genus (Brunt and Kitajima, 1973). The restriction enzyme map of the MMV genome is also quite

different than that of other members of this genus (Richins and Shepherd, 1983).

The genomes of several caulimoviruses including cauliflower mosaic virus (CaMV) (Gardner *et al.*, 1981), carnation etched ring virus (CERV) (Hull *et al.*, 1986), figwort mosaic virus (FMV) (Richins *et al.*, 1987), soybean chlorotic mottle virus (SoCMV) (Hasegawa *et al.*, 1989), peanut chlorotic streak virus (PCISV) (Richins, 1993), cassava vein mosaic virus (CVMV) (Calvert *et al.*, 1995), strawberry vein banding virus (SVBV) (Petrzik, 1996), petunia vein clearing virus (PVCV) (Richert-Poggler and Shepherd, 1997), and MMV (Maiti, unpublished) have been fully sequenced. The caulimovirus genome generally contains two transcriptional promoters, one for the full-length transcript and the other for a subgenomic

transcript; these are equivalent to the CaMV 35S and 19S transcripts respectively (Odell *et al.*, 1981; Hasegawa *et al.*, 1989; Driesen *et al.*, 1993).

A number of transcriptional promoters have been derived from the genomes of pararetroviruses, for example rice tungro bacilliform virus (RTBV) (Bhattacharyya-Pakrasi *et al.*, 1993); commelina yellow mottle virus (CYMV) (Medberry *et al.*, 1992), CaMV (Odell *et al.*, 1985; Lawton *et al.*, 1987), soybean chlorotic mottle virus (SoyCMV) (Hasegawa *et al.*, 1989), figwort mosaic virus (FMV, strain DxS) (Gowda *et al.*, 1989; Maiti *et al.*, 1997), FMV strain M3 (Sanger *et al.*, 1990), cassava vein mosaic virus (CVMV) (Verdaguer *et al.*, 1996), and peanut chlorotic streak virus (PCISV) (Maiti and Shepherd, 1998). The CaMV 35S promoter has been well characterized (Odell *et al.*, 1985; Ow *et al.*, 1987; Benfey and Chua, 1989, 1990; Fang *et al.*, 1989; Benfey *et al.*, 1990a, 1990b; Lam, 1994) and it has been widely used in chimeric gene constructs for heterologous gene expression in transgenic plants (Holtorf *et al.*, 1995; Wilmink *et al.*, 1995; Mitsuhara *et al.*, 1996). The 35S promoter from CaMV is also active in bacteria (Assaad and Singer, 1990), in yeast (Pobjecky *et al.*, 1990), in HeLa cells (Zahm *et al.*, 1989), and in *Xenopus* oocytes (Ballas *et al.*, 1989).

The transcriptional activity of the CaMV 35S promoter is the result of the synergistic and combinatorial effects of the enhancer *cis*-elements residing in the upstream region of a TATA box (Fang *et al.*, 1989; Benfey and Chua, 1990; Benfey *et al.*, 1990a, 1990b) similar to other promoters such as SV40 in mammalian systems (Schirm *et al.*, 1987; Fromental *et al.*, 1988; Ondek *et al.*, 1988). Detailed deletion analysis of the different elements of the CaMV 35S promoter established that domains A1 (−90 to −46 relative to transcription start site) and domain B (−343 to −90) are essential for developmental and tissue-specific promoter activity (Benfey and Chua, 1989; Fang *et al.*, 1989).

In the present study we report on the identification and characterization of the full-length transcript (FLt) promoter from mirabilis mosaic virus (MMV), a newly described caulimovirus. The optimal boundaries required for maximal promoter activity were defined by 5'- and 3'-end deletion analysis of the promoter/leader region of the full-length transcript promoter of MMV both in transgenic tobacco plants and in protoplast transient expression experiments. A 360 bp MMV FLt promoter fragment (sequence −297 to +63 from the transcription start site) was found suf-

ficient for strong constitutive promoter activity. The strength of the MMV FLt promoter is greater than that of the CaMV 35S promoter.

## Materials and methods

### *Plants and enzymes*

Tobacco plants (*Nicotiana tabacum* cv. Samsun NN) were used for plant transformation. For screening transformants, transgenic tobacco seeds (R<sub>1</sub> progeny) were germinated on agar medium in the presence of kanamycin (200 µg/ml). Restriction enzymes, DNA modifying enzymes, and RNA isolation kits were purchased from commercial sources and used according to the manufacturers' specifications. Nitrocellulose and Nytran membranes were obtained from Schleicher & Schuell, USA.

### *Isolation of MMV FLt promoter fragments and construction of plant expression vectors*

The plasmid pMMV-B10, a full-length genomic clone of MMV (Richins and Shepherd, 1983) has been fully sequenced (Maiti, 1996, unpublished data). A 590 bp segment (coordinates 6621 to 7210 of the MMV genome) was selected for MMV promoter deletion analysis. A series of promoter fragments included in constructing the plant transformation vector with the MMV FLt promoter were designed to study the influence of the upstream and downstream repeat sequences with respect to the TATA box. The defined MMV FLt promoter sequence, of length as indicated (see Figure 3A), was amplified by PCR using appropriately designed oligonucleotides to tailor *EcoRI* at the 5'-end and *HindIII* sites at the 3'-end positions. The plasmid pMMV-B10, a genomic DNA clone of MMV (Richins and Shepherd, 1983), was used as template for PCR reaction. PCR amplification was carried out for 30 cycles under the following standard conditions; denaturation (92 °C for 1 min), annealing (55 °C for 1 min), synthesis (72 °C for 2 min) using recombinant *Taq* DNA polymerase (Gibco-BRL). Each of these appropriately sized PCR-generated MMV promoter fragments 1 to 16 was restricted with *EcoRI* and *HindIII*; the restricted fragment was gel-purified and cloned into the corresponding sites of pUC119 for DNA sequencing. Before use, all PCR products cloned into pUC119 were sequenced by the dideoxy chain terminator method (Sanger *et al.*, 1997) using synthetic primers. Subsequently the promoter fragments

were isolated after restriction from pUC119 plasmids. Each of these promoter fragments was gel-purified and cloned into the plant expression vector pKYLX 71 (Schardl *et al.*, 1987) at its unique *EcoRI* and *HindIII* sites that flank the promoter. The following deletion plasmids were developed (the 5' and 3' ends of the promoter fragments are given in parenthesis): pKM1 (-457 to +133), pKM2 (-418 to +133), pKM3 (-378 to +133), pKM4 (-340 to +133), pKM5 (-297 to +133), pKM6 (-248 to +133), pKM7 (-193 to +133), pKM8 (-133 to +133), pKM9 (-78 to +133), pKM10 (-297 to -38), pKM11 (-457 to +63), pKM12 (-297 to +63), pKM13 (-248 to +63), pKM14 (-193 to +63), pKM15 (-297 to +2) and pKM16 (-297 to +33). These plant expression vectors have multiple cloning sites (MCS): 5'-*HindIII*-*BamHI*-*XhoI*-*SstI*-*XbaI*-3' with the following unique sites: *HindIII*, *XhoI*, *SstI* and *XbaI*. The reporter GUS gene was tailored by PCR to include just the coding sequence with the initiation and termination codons, flanked by a *XhoI* site at the 5' end and a *SstI* site at the 3' end. The PCR-isolated fragment for the GUS reporter gene was digested with *XhoI* and *SstI*, gel-purified and cloned into the corresponding sites of the pBS(KS+); the resulting plasmid was named pBSGUS. The GUS gene in pBSGUS was fully sequenced before use. The GUS gene from pBSGUS as *XhoI*-*SstI* fragment was inserted into the above-mentioned pKYLX-based plant expression vectors and the resulting deletion plasmids (constructs 1 to 16 in Figure 3A) were designated as pKM1GUS, pKM2GUS, pKM3GUS, pKM4GUS, pKM5GUS, pKM6GUS, pKM7GUS, pKM8GUS, pKM9GUS, pKM10GUS, pKM11GUS, pKM12GUS, pKM13GUS, pKM14GUS, pKM15GUS and pKM16GUS. The *XhoI*-GUS-*SstI* fragment from pBSGUS was inserted into the corresponding sites of pKYLX71 (Schardl *et al.*, 1987). The resulting plasmid pKYLX71GUS contains the GUS gene directed by the CaMV 35S promoter.

#### *Construction of plasmids for transient expression in protoplasts*

The fragment 5'-*EcoRI*-CaMV 35S.2 promoter-*HindIII*-*BamHI*-*XhoI*-*PstI*-*SstI*-*XbaI* *rbcS* 3' terminator-*ClaI*-3' was isolated from pKYLX7135S.2 (Schardl *et al.*, 1987) and cloned into the *EcoRI* and *AccI* sites of the pUC119 $\Delta$ H (a modified pUC119 in which the unique *HindIII* site was destroyed by digesting with *HindIII*, end-filled with Klenow fol-

lowed by ligation). The resulting plasmid was named pUCPMA. The GUS reporter gene from pBSGUS as *XhoI* and *SstI* fragment was inserted into the corresponding sites of pUCPMA to generate the plasmid pUCPMAGUS. The MMV FLt promoter fragments were cloned into the unique *EcoRI* and *HindIII* sites of transient expression vector pUCPMAGUS that flanks the promoter. The resulting PUC-based deletion plasmids (corresponding constructs 1 to 16 in Figure 3A) were designated as pPM1GUS, pPM2GUS, and so forth.

#### *DNA sequencing*

The nucleotide sequence of PCR products cloned into pUC119 (MMV FLt promoter fragments) or into pBS(KS+) (GUS gene) was determined from both direction by the dideoxy chain terminator method (Sanger *et al.*, 1990). Automated DNA sequencing was performed with an Applied Biosystem ABI Prism 310 Genetic Analyzer (Perkin Elmer) with ABI Prism Dye terminator cycle sequencing ready reaction kit containing Ampli *Taq* DNA polymerase. Primer extension was carried out by PCR (denaturation at 96 °C for 30 s; annealing at 50 °C for 30 s; extension at 60 °C for 4 min; 26 cycles using 5 nmol template DNA and 50 pmol specific primer in 20  $\mu$ l of reaction volume. Excess terminators were removed by 95% ethanol precipitation. Pellet containing the purified extension product was dried under vacuum and finally suspended in 25  $\mu$ l of template suppression reagent (P/N 401674; Perkin Elmer), vortexed, heat-denatured at 95 °C for 2 min before loading.

#### *Protoplast isolation and electroporation*

Isolation of protoplast from established tobacco cell suspension cultures (Xanthi 'Brad') and electroporation of protoplasts with supercoiled DNA containing GUS gene were done essentially as described earlier (Maiti *et al.*, 1998). The tobacco cell suspension culture was obtained from Dr S. Gowda, University of Florida, CREC, FL. Electroporation was performed by using the GenePulser II Apparatus (BioRad) with the Capacitance Extender II (Model 165-2107). An aliquot of 800  $\mu$ l containing  $2 \times 10^6$  protoplasts in an electroporation cuvette (0.4 cm electrode gap) was electroporated (200 V used for charging 960  $\mu$ F capacitance for 40 ms) with 5  $\mu$ g of supercoiled plasmid DNA containing GUS gene. After 20 h,  $2 \times 10^5$  protoplasts were harvested for each GUS assay.

### *Plant transformation and analysis of transgenic plants*

MMV FLt promoter constructs were introduced into *Agrobacterium tumefaciens* strain C58C1:pGV3850 by triparental mating and tobacco (*Nicotiana tabacum* cv. Samsun NN) was transformed with the engineered *Agrobacterium* as described earlier (Maiti *et al.*, 1993). Regenerated kanamycin-resistant plants were transferred to soil and grown in greenhouse conditions. On average, eight to twelve independent transgenic plants lines were generated for each construct.

### *Fluorometric GUS assay and histochemical localization of $\beta$ -glucuronidase activity*

Fluorometric assays to measure GUS activity of plant tissue extract and histochemical analysis of GUS activity in plant organs from primary transformants and whole seedlings (R1 progeny) were performed essentially according to the published procedure (Jefferson *et al.*, 1987) as described earlier (Maiti *et al.*, 1997). Protein in plant extract was estimated according to the published method (Bradford, 1976) using BSA as a standard.

### *RNA isolation, RNA dot blot and northern analysis*

Total RNA was prepared from transgenic tobacco (*Nicotiana tabacum* cv. Samsun NN) seedlings (R1 progeny) and untransformed control seedlings by extracting with guanidine thiocyanate solution (Chomczynski and Sacchi, 1987) using an Ambion RNA isolation kit (RNAqueous). The clarified lysate (supernatant) was diluted with ethanol and allowed to pass through a filter cartridge to enable RNA to bind to the glass filter fiber (Boom *et al.*, 1990). The filter was washed three times with wash buffer (Ambion Kit, catalog No. 1911) and bound total RNA was eluted by warm (65 °C) TE. Total RNA was purified by LiCl precipitation before analysis. For northern blot analysis, total RNA (10  $\mu$ g) was separated on 1.2% denaturing agarose gel containing 2.2 M formaldehyde in MOPS buffer (20 mM MOPS, 8 mM sodium acetate, 1 mM EDTA pH 8.0). After electrophoresis the gel was rinsed with water and 6 $\times$  SSC. RNA from the gel was transferred onto Nytran membrane (Schleicher & Schuell) by vacuum filtration using a vacuum-blotting apparatus (Stratagene). The GUS probe was made using the random priming kit (Stratagene). RNA on the membrane was fixed by UV cross-linker (Statalinker,

Stratagen) at auto-setting. The membrane was prehybridized for 2 h at 42 °C in a solution containing 6 $\times$  SSC (1 $\times$  SSC is 0.15 M NaCl, 0.015 M sodium citrate), 1% SDS, 5 $\times$  Denhardt's solution, 100 ng/ml salmon sperm DNA and 10% dextran sulfate. Hybridization was carried out under the same conditions using <sup>32</sup>P-labeled GUS probe for 20 h. After hybridization the membrane was washed four times at 65 °C for 30 min each with 0.5 $\times$  SSC, 0.1% SDS and 0.2 $\times$  SSC, 0.1% SDS, and finally exposed to X-ray film for autoradiography.

### *RNA dot blot*

The RNA (10  $\mu$ g) sample isolated from leaves of transgenic tobacco developed for each construct was denatured with 50% deionized formamide and 2.2 M formaldehyde at 55 °C for one hour, chilled on ice. Denatured RNA samples were diluted to 500  $\mu$ l, and transferred onto nitrocellulose membrane using a BRL hybrid-dot manifold. After RNA transfer the membrane was baked at 80 °C for 2 h under vacuum. The RNA blot was prehybridized, hybridized with <sup>32</sup>P-labeled GUS-probe, washed, and detected by autoradiography as described above for northern analysis.

### *Determination of transcriptional start site by primer extension analysis*

A 36 nucleotide long primer with the sequence 5'-TTTCGCGATCCAGACTGAATGCCCCACAGGCCGTCGA-3' which is complementary to a region 36 nucleotides downstream of the ATG start codon in the *uidA* gene was used for primer extension. The oligonucleotide was 5'-end-labeled with 6 U of T<sub>4</sub> polynucleotide kinase (BRL) and 10  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P]-ATP (3000  $\mu$ Ci/mmol). End-labeled primer was annealed with 50  $\mu$ g of total RNA for 18 h at 42 °C. The primer extension reaction was carried out at 42 °C for 1 h with 20 U of Superscript reverse transcriptase (Gibco-BRL). The extension product was separated on a 7.5% polyacrylamide gel containing 7 M urea. The reaction was carried out according to the published procedure (Sambrook *et al.*, 1989). Sequencing reactions were carried out according to Sanger *et al.* (1977) using Sequenase Version 2, USB. Sequencing reactions and the primer extended product was run side by side on the same sequencing gel.

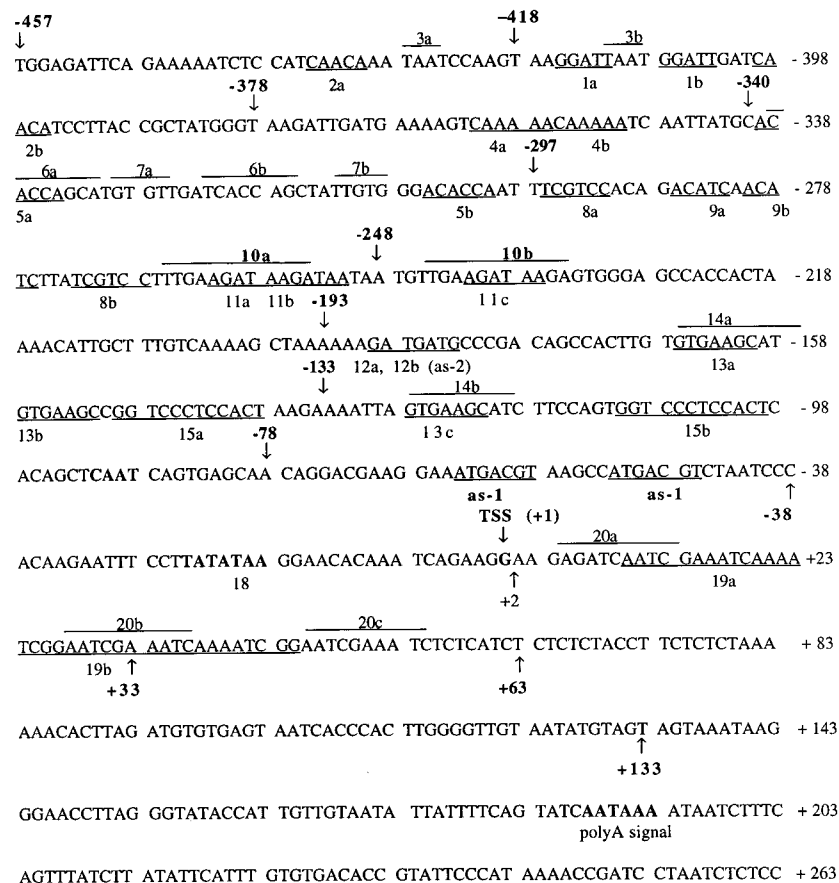


Figure 1. The DNA sequence of the MMV FLt promoter in the 5' to 3' direction. The tentative regulatory elements: the TATA box (TATATAA), the CAT box (CAAT), the poly(A) signal (AATAAA) are shown in bold; all the repeat sequence domains including *as-1* and *as-2* motifs (designated domains: 1a, 1b; 2a, 2b; 3a, 3b; 4a, 4b; 5a, 5b; 6a, 6b; 7a, 7b; 8a, 8b; 9a, 9b; 10a, 10b; 11a, 11b, 11c; 12a, 12b; 13a, 13b, 13c; 14a, 14b; 15a, 15b; 16; 17; 18; 19a, 19b; 20a, 20b & 20c) are underlined or overlined. The end points for the 5' or the 3' deletion plasmids are indicated above or below the sequence respectively. The transcription start site (TSS) is indicated as +1.

## Results and discussion

### Structure analysis of the MMV FLt promoter

Mirabilis mosaic virus (MMV) has a double-stranded circular DNA genome of 8 kb (Richins and Shepherd, 1983). The DNA sequence of a portion of the MMV genome had been earlier determined (R. Kormelink and R.J. Shepherd 1987, unpublished data). The sequence of the MMV FLt promoter in the large intergenic region is shown in Figure 1.

This MMV promoter sequence contains several putative regulatory domains found in other caulimovirus promoters: the TATA box sequence TATATAA, a CAAT sequence 63 bp upstream of the TATA box, and a poly(A) signal consisting of an AATAAA sequence 204 bp downstream of the TATA

box. In caulimoviruses both subgenomic and pregenomic (full-length) transcripts share the same 3' ends using the same poly(A) signal. A pentanucleotide repeat sequence motif (TGACG) shown as a transcriptional enhancer element as found in an *as-1* motif (Lam, 1994) is located 23–41 5' of the TATA box. An overlapping GATG repeat sequence, a core sequence for the enhancer element *as-2* motif (Lam, 1994), occurs at 159–166 5' of the TATA box. Additionally there are several repetitive sequences (shown in Figure 1, underlined or overlined) upstream of the TATA box, and a direct repeat of 18 bp 26–63 3' of the TATA box as well as a 10 nucleotide repeat 44–69 3' of the TATA box. These repeat sequence domains are not found in other caulimoviruses but they may have some regulatory function.

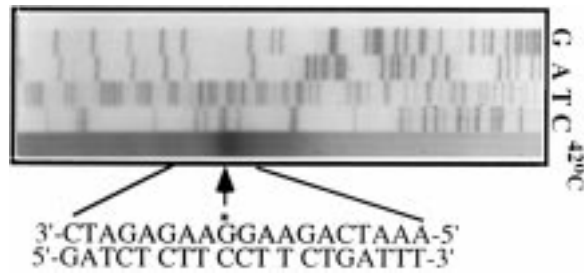


Figure 2. Determination of the MMV FLt promoter transcription start site (TSS) by primer extension analysis of GUS transcript isolated from transgenic tobacco. Primer extension reaction was carried out as described in Materials and methods. Primer extension products were subjected to electrophoresis alongside sequence reaction of GUS gene construct (lanes G, A, T and C) performed with the same labeled primer. The minus strand DNA sequence read on the gel is shown, and the transcription start site (G\*) in the corresponding plus strand DNA is indicated by an arrow.

#### Determination of the transcriptional start site of the MMV FLt promoter

The transcription start site (TSS) of the MMV FLt promoter was determined by primer extension analysis with total RNA isolated from transgenic tobacco seedlings ( $R_1$  progeny) developed with the construct pKM12GUS (Figure 2). A major extension product was detected and mapped to a guanidine residue located 24 nucleotides downstream of the TATA box and most likely it represents the 5' end of MMV FLt transcript. The location of TSS reported for other caulimoviruses including CaMV 35S (Guiley *et al.*, 1982), FMV34S (Sanfacon, 1994), FMVFLt (Maiti *et al.*, 1997), PCISV FLt (Richins, 1993) is at 32, 37, 45 and 29 nucleotides downstream of the respective TATA box respectively. Sequence comparison of the TATA box in the TSS region of the MMV promoter with that of CaMV, PCISV, FMV showed limited sequence homology.

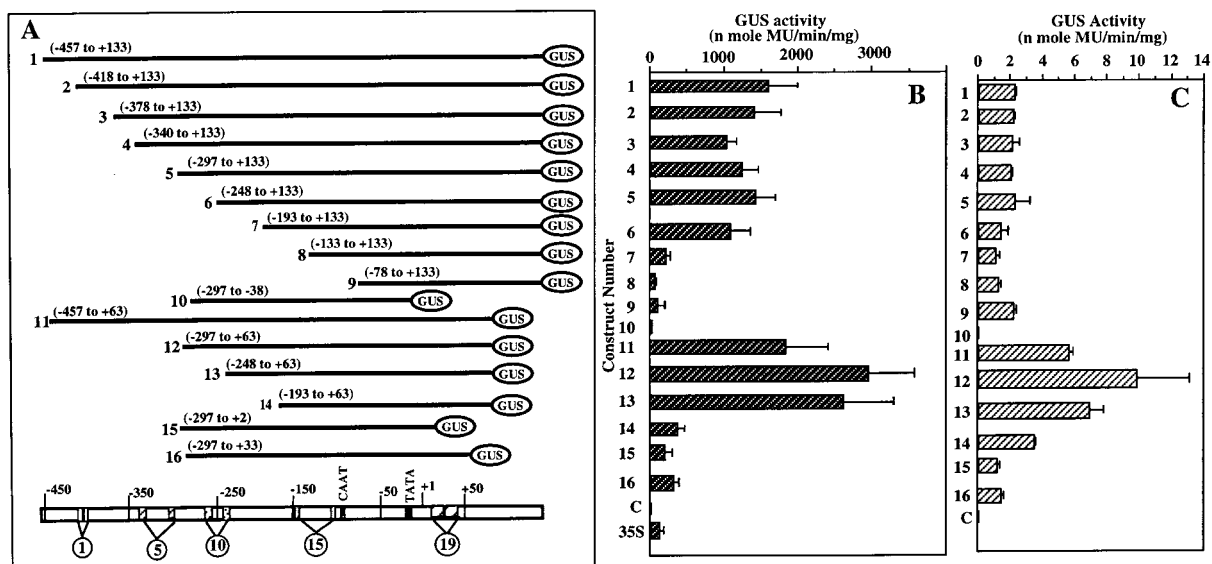
#### Expression analysis of the MMV FLt promoter deletion constructs in transgenic plants and in protoplast transient expression experiments

A deletion analysis scheme is shown in Figure 3A. The series of 5'- and 3'-end-deleted promoter fragments were included in constructing the plant transformation vector with the MMV FLt promoter in order to define the boundaries required for maximal expression from the promoter/leader region and also to study the influence of the upstream and downstream repeat sequences with respect to the TATA box. The defined MMV promoter fragments 1 to 16 (as indicated in

Figure 3A) were amplified by PCR and cloned into plant expression vector as described in Materials and methods. The resulting plasmids contain the defined promoter sequences 1 to 16 as described in Figure 3A. The upstream and downstream deletion end points of the promoter fragment in each plasmid are indicated in parenthesis (Figure 3A). The various constructs with a reporter gene inserted into the multiple cloning clusters (GUS gene in this case) were tested in transgenic plants ( $R_0$  and  $R_1$  progeny) and also in protoplast transient expression assay.

At least 10 to 12 independent primary transgenic tobacco (*Nicotiana tabacum* cv. Samsun NN) lines ( $R_0$  progeny) were generated with each of these constructs 1 to 16, and construct pKYLX71GUS (Figure 3A). Seeds were collected from self-fertilized independent lines. Segregation analysis of the marker gene ( $Kan^R$ ) was performed. Individual transgenic tobacco lines showing expected segregation ratio ( $Kan^R/Kan^S = 3:1$ ) for the marker  $Kan^R$  gene were further analyzed. Eight to ten independent lines for each construct showed the expected segregation ratio. The expression of the GUS reporter gene in seedlings ( $R_1$  progeny/second generation) with these constructs (Figure 3A) was examined by fluorometric assay of tissue extracts (Figure 3B) and in protoplast transient expression (Figure 3C). Histochemical GUS staining (Figure 4) examined the tissue distribution of GUS activity in seedlings ( $R_1$  progeny/second generation) developed with these constructs.

The MMV FLt promoter deletion analysis in transgenic plants (Figure 3B) and in protoplast transient expression assay (Figure 3C) showed overall a very similar profile. However the GUS activity in protoplast assay is about 200 to 800 times less than in stably transformed transgenic lines ( $R_1$  progeny). Analysis of transgenic lines ( $R_1$  progeny) and transient expression assay showed more GUS activity with construct 12 which contains the promoter fragment (coordinate -297 to +63 from TSS, Figure 1). In transgenic plants the 5'-deletion analysis of MMV FLt promoter fragments 1 to 9 (Figure 3A) gave 54%, 47%, 36%, 42%, 47%, 37%, 7.5%, 2.2%, and 4%, respectively, of full activity (compared with construct 12 giving the highest activity, Figure 3B). Similarly in protoplast transient expression experiments the 5'-deletion fragments 1 to 9 showed 23%, 22%, 21%, 20%, 24%, 14.5%, 11%, 13.6%, and 21.8% of full activity respectively (compared with construct 12 giving the highest activity, Figure 3C).



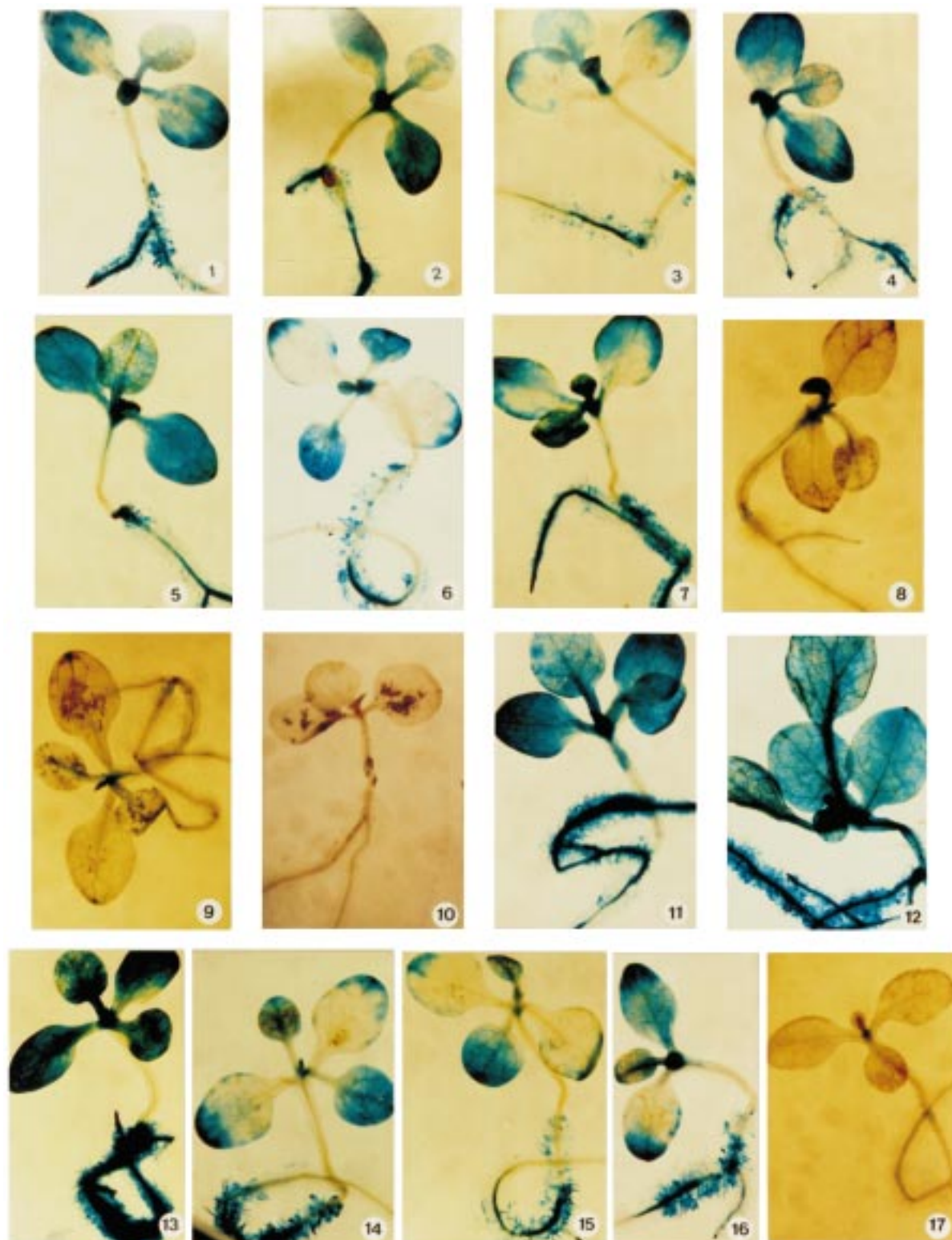
**Figure 3.** A. A schematic map of the GUS constructs (numbers 1 to 16) developed for 5' and 3' deletion analysis of MMV FLt promoter. The coordinates of the relative deletion fragment are indicated in parenthesis. At the bottom, the relative positions of the TATA box and the CAAT box, transcription starts site (+1) and direct repeat sequences (denoted in Figure 1, numbered in circle) are shown. B. MMV FLt promoter deletion analysis in transgenic plants expressing the GUS reporter gene. Expression analysis of 5'- and 3'-end deletion constructs (1 to 16) of the MMV FLt promoter in transgenic *N. tabacum* cv. Samsun NN (R<sub>1</sub> progeny, 24-day-old seedlings) was conducted. The 5'- and 3'-deletion end points for each construct are as indicated in A. The MMV promoter activity was monitored in 24-day old seedlings (R<sub>1</sub> generation, Kan<sup>R</sup>) grown aseptically on an MS-agar medium in the presence of kanamycin (200 mg/ml) and 3% sucrose. Soluble protein extracts (5  $\mu$ g) from whole seedlings were used for GUS assay. The data are means of four independent experiments for each construct (1 to 16); eight to ten independent lines for each construct were assayed. The average GUS activity is presented for each chimeric construct in the histogram with the standard deviation from the mean indicated by an error bar. Error bars show the 95% confidence intervals on the means. The statistical (one-way analysis of variance, ANOVA) analysis showed a *P* value of <0.0001, considered extremely significant. c, untransformed control, tissue extract from wild-type *N. tabacum* cv. Samsun NN. 35S, GUS gene directed by the CaMV 35S promoter. C. MMV FLt promoter deletion analysis in protoplast transient expression experiments. The average GUS activity from three independent experiments is presented for each construct in the histogram with the standard deviation from the mean indicated by an error bar. Error bars show the 95% confidence intervals on the means. The statistical ANOVA analysis showed a *P* value <0.0001, considered extremely significant.

Construct 10, which is devoid of the TATA box, showed very little or no GUS activity indicating the importance of a TATA element in the MMV promoter sequence. A 360 bp MMV FLt promoter/leader fragment, sequence  $-297$  to  $+63$  from TSS (in construct 12), was found to be sufficient for maximal GUS expression in transgenic tobacco. The MMV FLt promoter with a longer upstream sequence ( $-457$  to  $+63$ , construct 1) is about 54% less active than the promoter fragment ( $-297$  to  $+63$ , construct 12).

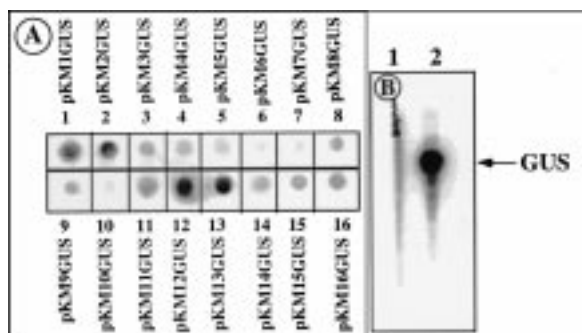
The strength of the MMV FLt promoter with a GUS reporter gene was evaluated by hybridization analysis of total RNA. Total RNA was prepared from independent transformed tobacco seedlings (R<sub>1</sub> progeny) developed for each construct 1 to 16. Total RNA was hybridized with a <sup>32</sup>P-labeled DNA fragment containing the GUS coding sequence (Figure 5A). The transcript level was maximal with construct 12 and was in good agreement with the GUS ex-

pression analysis performed by fluorometric assay of total tissue extract (Figure 3B) and also by histochemical GUS staining of transgenic tobacco seedlings (Figure 4). The tissue-specific expression of GUS in different parts of plants developed for these deletion constructs was not quantitatively evaluated. The northern analysis of total RNA isolated from transgenic tobacco seedlings (R<sub>1</sub> progeny/ second generation), developed with construct 12, showed the expected size (2100 nt) of GUS transcript (Figure 5B).

The deletion plasmids containing MMV FLt promoter fragments 11 and 13 showed 62% and 88% of full activity respectively in transgenic plants (Figure 3B); and the corresponding values in protoplast transient assay were 57% and 70% of full activity (Figure 3C; compare with construct 12). In MMV FLt promoter deletion analysis in transgenic plants, the 5' deletion plasmid pKM14GUS (construct 14, promoter coordinate  $-193$  to  $+63$ ) containing 'as-1' domains



**Figure 4.** Histochemical analysis of GUS expression in tobacco seedlings (transgenic *N. tabacum* cv. Samsun NN, R<sub>1</sub> progeny, 24-day old seedlings) from the best expressing independent line representing each construct (1 to 16) containing the GUS gene driven by the designated MMV FLt promoter fragment. These data were derived from pools of transformed lines in each construct. The upstream and downstream deletion end points for each construct are as indicated in Figure 3A. 1 to 16. Histochemical localization of GUS activity in transgenic tobacco seedlings (magnification  $\times 8.6$ ) representing each construct 1 to 16. GUS staining was most intense with construct 12 followed by constructs 13 and 11. Construct 10, devoid of TATA box, showed no GUS activity and construct 8, devoid of upstream sequence (–133 to –193), showed very little GUS activity. 17. Untransformed control, wild-type *N. tabacum* cv. Samsun NN 24-day old seedling (magnification  $\times 8.6$ ): no GUS activity was detected.



**Figure 5.** A. RNA-dot hybridization analysis of total RNA (10  $\mu$ g) isolated from 24-day old seedlings developed for each construct (1 to 16). Total RNA was isolated from the best GUS-expressing independent line. B. Northern analysis of total RNA isolated from untransformed Samsun NN (lane 1) and transgenic tobacco seedlings (R<sub>1</sub> progeny) developed with construct 12 (lane 2). Total RNA was hybridized with <sup>32</sup>P-labeled GUS coding sequence.

and TATA box sequence, and the 3'-deletion plasmid pKM16GUS (construct 16, promoter coordinate –297 to +33) showed about 13% and 11% of full activity, respectively (compare with plasmid pKM12GUS, construct 12, giving the highest activity, Figure 3B). The 5'-deletion plasmid pKM14GUS (construct 14) containing 'as-1' domains and TATA box sequence showed less activity suggesting dependence on additional upstream sequence elements. These results of MMV FLt promoter deletion analysis suggest that a TATA-upstream sequence (coordinate –297 to –193, in Figure 1) and a TATA-downstream sequence (coordinate +2 to +63, in Figure 1) may be involved in regulating function of the full-length transcript promoter. Several direct repeat sequences (8a & 8b; 9a & 9b; 10a & 10b; 11a & 11b; as denoted in Figure 1) are present in the TATA-upstream sequence (coordinate –297 to –193, in Figure 1).

A 19 bp motif having strong homology with the activation sequence 1 (*as-1*) of the CaMV 35S promoter (Lam *et al.*, 1989) was identified in the MMV promoter at coordinate –64 to –46 relative to TSS. This motif is also found in FMV (Sanger *et al.*, 1990) and CVMV (Verdaguer *et al.*, 1996). This *as-1* mo-

tif is able to confer root-specific expression in the CaMV 35S promoter (Lam *et al.*, 1989). However the MMV FLt deletion constructs 8 and 9 containing putative *as-1* domain showed no appreciable GUS staining in root tips of seedlings (Figure 4). More studies will be needed to fully evaluate the function of *as-1* domain in the MMV FLt promoter. This *as-1* domain plays a more complex role in the regulation of the promoter activity by acting synergistically with other *cis* elements (Fang *et al.*, 1989; Lam *et al.*, 1989). It is also reported as a salicylic acid-responsive element (Qin *et al.*, 1994).

The promoter fragment –193 to +133 in construct 7 confers much less promoter activity than construct 6 with promoter coordinates –248 to +133 and construct 5 with promoter coordinates –297 to +133 (Figure 3B). This implies that the upstream sequence –297 to –193 is essential for promoter activity and two repeat sequences in this region (TTGAA-GATAAGA, Nos. 10a and 10b in Figure 1) at positions –265 to –254 and –245 to –234 may be important for promoter activity. More studies including internal deletion mapping will be required to fully analyze the importance of these sequences.

The MMV FLt promoter gives better activity with a longer leader sequence, extending to +63 compared with +2 and +33 (compare constructs 15 and 16 respectively with construct 12, Figure 4 and Figure 5A). Increases in GUS transcript level and also in GUS activity suggest that the 3' leader of MMV FLt promoter have some sequence important for initiation of transcription and concurrent translation. The MMV leader sequence may enhance transcription. It has been shown that the untranslated viral leader sequences from CaMV and tobacco mosaic virus (TMV) (Day Dowson *et al.*, 1993) stimulate the expression of a downstream reporter gene by enhancing translation. The untranslated leader sequence of FMV (Maiti *et al.*, 1997) also stimulates the expression of a downstream gene; it needs to be evaluated whether it is due to enhanced transcription or translation.

Two 18-nucleotide direct repeat sequences (Nos. 19a and 19b in Figure 1) and three 10-nucleotide repeat sequences (Nos. 20a, 20b and 20c in Figure 1) are located in the TATA-downstream sequence, coordinates +2 to +63 (in Figure 1) of the MMV FLt promoter. These repeat sequence domains may have some regulatory functions. As the construct pKM15GUS (construct 15, promoter coordinate -297 to +2) showed much less activity than the pKM12GUS (construct 12, promoter coordinate -297 to +63), this indicates that the repeat sequences AATCGAAATCAAAA at positions +10 to +23 and +28 to +44 are also playing important roles in promoter activity. More work will be needed to fully define the *cis* elements of the MMV FLt promoter.

The comparative analysis of CaMV 35S-promoter and MMV FLt promoter in transgenic plants (R<sub>1</sub> progeny) developed for construct pKYLX71GUS and pKM12GUS, respectively, showed that the MMV FLt promoter is about 25 times stronger than the CaMV 35S promoter (Figure 3B).

Our studies indicate that the MMV promoter is a strong constitutive promoter able to direct foreign gene expression in transgenic plants. There is limited sequence homology between the MMV FLt promoter and those of CaMV and other caulimovirus promoters although they are functionally analogous, which may imply differences in the mechanisms of promoter regulation. Analysis of the MMV promoter sequence shows the presence of several motifs that resemble previously identified *cis* elements that are implicated in transcriptional regulation. The presence of such motifs in the MMV promoter sequence could explain the pattern and relative strength of different promoter deletion fragments in transgenic plants. We are interested in studying the regulatory region of the MMV FLt promoter and also interacting nuclear factors involved in tissue-specific and constitutive expression of genes in plants. The CaMV 35S promoter is composed of several discrete organ-specific *cis* elements in the upstream region of the promoter (Lam, 1994). The molecular analysis of the functional interactions between specific *cis* elements and cognate *trans* factor will be essential in developing more active 'super promoter' and modified tissue-specific promoter.

### Acknowledgements

We are greatly indebted to Kentucky State THRI for facilities and support. This work was supported by

Kentucky State THRI grants 5-41086 and 5-41132. We thank Dr Thomas Pirone and two anonymous reviewers for critical reading and valuable comments in improving the manuscript, Ms Ruth Babbit for assisting in color photography and Dr Pulak Tripathi for helping in primer extension analysis, Dr S. Gowda for making Xanthi suspension culture available to us, and Mr Willy Witt for technical assistance.

### References

- Assaad, F.F. and Signer, E.R. 1990. Cauliflower mosaic virus P35S promoter activity in *Escherichia coli*. *Mol. Gen. Genet.* 223: 517-520.
- Ballas, N., Shimshon, B., Hermona, S. and Abraham, L. 1989. Efficient functioning of plant promoters and polyadenylated sites in *Xenopus* oocytes. *Nucl. Acids Res.* 17: 7891-7904.
- Benfey, P.N. and Chua, N.H. 1989. The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue specific expression patterns. *EMBO J.* 8: 2195-2202.
- Benfey, P.N. and Chua, N.H. 1990. The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. *Science* 250: 959-966.
- Benfey, P.N., Ren, L. and Chua, N.H. 1990a. Combinatorial and synergistic properties of CaMV 35S enhancer subdomains. *EMBO J.* 9: 1685-1696.
- Benfey, P.N., Ren, L. and Chua, N.H. 1990b. Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. *EMBO J.* 9: 1677-1684.
- Bhattacharyya-Pakrasi, M., Pen, J., Elmer, J.S., Laco, G., Shen, P., Kaniewska, M.B., Kononowicz, H., Wen, F., Hodges, T.K. and Beachy, R.N. 1993. Specificity of a promoter from the rice tungro bacilliform virus for expression in phloem tissues. *Plant J.* 4: 71-79.
- Boom, R., Sol, C.J.A., Salimans, M.M.M., Jansen, C.L., Wertheim-van Dillen, P.M.E. and Noorda J.V.D. 1990. Rapid and simple method for purification of nucleic acids. *J. Clin. Microbiol.* 28: 495-503.
- Bradford, M.M. 1976. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein dye-binding. *Anal. Biochem.* 72: 248-254.
- Brunt, A.A. and Kitajima, E.W. 1973. Intracellular location and some properties of mirabilis mosaic virus, a new member of the cauliflower mosaic group of viruses. *Phytopath. Z.* 76: 265-275.
- Calvert, L., Ospina, M. and Shepherd, R.J. 1995. Characterization of cassava vein mosaic virus: a distinct plant pararetrovirus. *J. Gen. Virol.* 76: 1271-1276.
- Chomczynski, P. and Sacchi, N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162: 156-159.
- Day Dowson, M.J., Ashurst, J.L., Mathias, S.F., Watts, J.W., Wilson, T.M.A. and Dixon R.A. 1993. Plant viral leaders influence expression of a reporter gene in tobacco. *Plant Mol. Biol.* 23: 97-109.
- Driesen, M., Benito-Moreno, R.M., Hohn, T. and Futterer, J. 1993. Transcription from the CaMV 19S promoter and autocatalysis of translation from CaMV RNA. *Virology* 195: 203-210.
- Fang, R.X., Nagy, F., Sivasubramanian, S. and Chua, N.H. 1989. Multiple *cis* regulatory elements for maximal expression of the

- cauliflower mosaic virus 35S promoter in transgenic plants. *Plant Cell* 1: 141–150.
- Fromental, C., Kanno, M., Nomiyama, H. and Chambon, P. 1988. Cooperativity and hierarchical levels of functional organization in the SV 40 enhancer. *Cell* 1: 141–150.
- Gowda, S., Wu, F.C., Herman, H.B. and Shepherd, R.J. 1989. Gene VI of figwort mosaic virus (caulimovirus group) functions in posttranscriptional expression of genes on the full-length RNA transcript. *Proc. Natl. Acad. Sci. USA* 86: 9203–9207.
- Gardner, R.C., Howarth, A.J., Hahn, P., Brown-Leudi, M., Shepherd, R.J. and Messing, J. 1981. The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing. *Nucl. Acids Res.* 9: 2871–2888.
- Guiley, H., Dudley, R.K., Jonard, G., Balasze, E., and Richards, K.E. 1982. Transcription of cauliflower mosaic virus DNA: detection of promoter sequences and characterization of transcripts. *Cell* 30: 763–773.
- Hasegawa A, Verver J, Shimada A, Saito M, Goldbach R, van Kammen, A., Miki, K., Kameya-Iwaki, M. and Hibi, T. 1989. The complete sequence of soybean chlorotic mottle virus DNA and the identification of a novel promoter. *Nucl. Acids Res.* 17: 9993–10013.
- Holtorf, S. Apel, K. and Bohlmann, H. 1995. Comparison of different constitutive and inducible promoters for the overexpression of transgenes in *Arabidopsis thaliana*. *Plant Mol. Biol.* 29: 637–646.
- Hull, R., Sadler, J. and Longstaff, M. 1986. The sequence of carnation etched ring viral DNA: comparison with cauliflower mosaic virus and retroviruses. *EMBO J.* 5: 3083–3090.
- Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. 1987. GUS fusion:  $\beta$ -glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* 6: 3901–3907.
- Lam, E. 1994. Analysis of tissue-specific elements in the CaMV 35S promoter. In: L. Nover (Ed.), *Results and Problems in Cell Differentiation*, Vol. 20, Plant Promoters and Transcription Factors, Springer-Verlag, Berlin/Heidelberg, pp. 181–196.
- Lam, E., Benfey, P.N., Gilmartin, P.M., Fang, R.X. and Chua N.H. 1989. Site specific mutation alter *in vitro* factor binding and change the promoter expression pattern in plants. *Proc. Natl. Acad. Sci. USA* 86: 7890–7894.
- Lawton, M.A., Tierney, M.A., Nakamura, I., Anderson, E., Komeda, Y., Dube, P., Hoffman, N., Fraley, R.T., and Beachy, R.N. 1987. Expression of a soybean  $\beta$ -conglycinin gene under the control of the cauliflower mosaic virus 35S and 19S promoters in transformed petunia tissues. *Plant Mol. Biol.* 9: 315–324.
- Maiti, I.B. and Shepherd, R.J. 1998. Isolation and expression analysis of peanut chlorotic streak caulimovirus (PCISV) full-length transcript (FLt) promoter in transgenic plants. *Biochem. Biophys. Res. Commun.* 244: 440–444.
- Maiti, I.B., Murphy, J.F., Shaw, J.G. and Hunt, A.G. 1993. Plants that express a potyvirus proteinase genes are resistant to virus infection. *Proc. Natl. Acad. Sci. USA* 90: 6110–6114.
- Maiti, I.B., Gowda, S., Kiernan, J., Ghosh, S.K. and Shepherd, R.J. 1997. Promoter/leader deletion analysis and plant expression vectors with the figwort mosaic virus (FMV) full length transcript (FLt) promoter containing single and double enhancer domains. *Transgen. Res.* 6: 143–156.
- Maiti, I.B., Richins, R.D. and Shepherd, R.J. 1998. Gene expression regulated by gene VI of caulimovirus: transactivation of downstream genes of transcripts by gene VI of peanut chlorotic streak virus in transgenic tobacco. *Virus Res.* 57: 113–124.
- Medberry, S.L., Lockhart, B.E.L. and Olszewski, N.E. 1992. The Commelina yellow mottle virus promoter is a strong promoter in vascular and reproductive tissues. *Plant Cell* 4: 185–192.
- Mitsuhara, I., Ugaki, M., Hirochika, H., Ohshima, M., Murakami, T., Gotoh, Y., Katayose, Y., Nakamura, S., Honkura, R., Nishimiya, S., Uneo, K., Mochizuki, A., Tanimoto, H., Tsugawa, H., Otsuki, Y. and Ohashi, Y. 1996. Efficient promoter cassettes for enhanced expression of foreign genes in dicotyledonous and monocotyledonous plants. *Plant Cell Physiol.* 37: 49–59.
- Odell, J.T., Dudley, R.K. and Howell, S.H. 1981. Structure of the 19S RNA transcripts encoded by the cauliflower mosaic virus genome. *Virology* 111: 377–385.
- Odell, J.T., Nagy, F. and Chua, N.-H. 1985. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* 313: 810–812.
- Ondek, B., Gloss, L. and Herr, W. 1988. The SV 40 enhancer contains two distinct levels of organization. *Nature* 333: 40–45.
- Ow, D.W., Jacobs, J.D. and Howell, S.H. 1987. Functional regions of the cauliflower mosaic virus 35S RNA promoter determined by use of the firely luciferase gene as a reporter of promoter activity. *Proc. Natl. Acad. Sci. USA* 84: 4870–4874.
- Petrzik, K. 1996. Strawberry vein banding virus complete genome sequence (Document ID 1360608), GenBank accession Number X97304.
- Pobjecky, N., Rosenberg, G.H., Dinter-Gottlieb, G. and Kaufer, N.F. 1990. Expression of the  $\beta$ -glucuronidase gene under the control of the CaMV 35S promoter in *Schizosaccharomyces pombe*. *Mol. Gen. Genet.* 220: 314–316.
- Qin, X.-F., Holuigue, L., Horvath, D.M. and Chua, N.-H. 1994. Immediate early transcription activation by salicylic acid via the cauliflower mosaic virus *as-1* element. *Plant Cell* 6: 863–874.
- Richert-Poggeler, K.R. and Shepherd R.J. 1997. Petunia vein-clearing virus: a plant pararetrovirus with the core sequences for an integrase function. *Virology* 236: 137–146.
- Richins, R.D. 1993. Organization and expression of the peanut chlorotic streak virus genome. Ph. D. Dissertation, University of Kentucky, Lexington, KY (for the PCISV genomic sequence: DNA EMBL Data Library GenBank accession Number U13988).
- Richins, R.D. and Shepherd, R.J. 1983. Physical maps of the genome of dahlia mosaic virus and mirabilis mosaic virus – two members of the caulimovirus group. *Virology* 124: 208–214.
- Richins, R.D., Scholthof, H.B. and Shepherd R.J. 1987. Sequence of figwort mosaic virus DNA (caulimovirus group). *Nucl. Acids Res.* 15: 8451–8466.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sanfacon, H. 1994. Analysis of figwort mosaic virus (plant pararetrovirus) polyadenylation signal. *Virology* 198: 39–49.
- Sanger, F., Nicklen, S. and Coulson, A.R. 1977. DNA sequencing with chain terminating inhibitor. *Proc. Natl. Acad. Sci. USA* 74: 5463–5467.
- Sanger, M., Daubert, S., Goodman, R.M. 1990. Characteristics of a strong promoter from figwort mosaic virus: comparison with the analogous 35S promoter from cauliflower mosaic virus and the regulated mannopine synthase promoter. *Plant Mol. Biol.* 14: 433–443.
- Schardl, C.L., Byrd, A.D., Benzion, G., Altschuler, M.A., Hildebrand, D.F. and Hunt, A.G. 1987. Design and construction of a versatile system for the expression of foreign genes in plants. *Gene* 61: 1–11.

- Schirm, S., Jiricny, J., and Schaffner, W. 1987. The SV 40 enhancer can be dissected into multiple segments each with a different cell type specificity. *Genes Dev.* 1: 65–74.
- Verdaguer, B., de Kochko, A., Beachy, R.N. and Fauquet, C. 1996. Isolation and expression in transgenic tobacco and rice plants, of the casava vein mosaic virus (CVMV) promoter. *Plant Mol. Biol.* 31: 1129–1139.
- Wilmink, A., van de Ven, B.C.E., Dons, J.J.M. 1995. Activity of constitutive promoter in various species from the Liliaceae. *Plant Mol. Biol.* 28: 949–955.
- Zahm, P., Seong-Iyul, R., Klaus, G. 1989. Promoter activity and expression of sequences from Ti-plasmid stably maintained in mammalian calls. *Mol. Cell. Biochem.* 90: 9–18.