Mutational analysis of the core and modulator sequences of the BMV RNA3 subgenomic promoter

Loren E.Marsh, Theo W.Dreher and Timothy C.Hall*

Department of Biology, Texas A&M University, College Station, TX 77843-3258, USA

Received October 5, 1987; Revised and Accepted December 30, 1987

ABSTRACT

The subgenomic promoter of a (+)-stranded RNA virus, brome mosaic virus (BMV) controlling synthesis of subgenomic RNA4 has been defined *in vitro*. Truncated and mutant (-)-strand RNA templates were produced by *in vitro* transcription of cloned RNA3 cDNA. Subgenomic (+)-sense RNA was synthesized *in vitro* from these templates by a replicase (RNA-dependent RNA polymerase) preparation extracted from infected barley leaves. The activities of templates with truncations and deletions surrounding the RNA4 initiation site revealed a promoter of approximately 62 bases grouped into four functional domains. The core sequence consists of about twenty bases immediately upstream of, and including, the initiation nucleotide. In addition to the core sequence, a domain overlapping the 5' untranslated end of RNA4 apparently determines correct initiation. Two domains immediately upstream of the promoter core consist of the internal poly(A) tract of RNA3, which probably serves as a non base-paired spacer facilitating access of the replicase to the promoter, and a sequence, UUAUUAUU, that is required for high levels of promoter activity. Homologies to sequences surrounding the initiation sites of subgenomic RNAs from several plant RNA viruses, and from alphaviruses, have been detected.

INTRODUCTION

The majority of viruses infecting both plants and animals have RNA based genomes. Furthermore, most RNA viruses infecting eucaryotes are (+)-stranded viruses such as the alphaviruses, tobacco mosaic virus (TMV) and brome mosaic virus (BMV). These (+)-stranded viruses are responsible for many important diseases of both animals and plants. The generation of subgenomic RNAs from polycistronic genomic RNAs of (+)-stranded viruses represents a mechanism for expression of internal cistrons, and an effective amplification system for proteins needed in large quantities. In the case of brome mosaic virus (BMV), the coat protein is expressed *in vivo* (1) and *in vitro* from subgenomic RNA4 (2), which is colinear and completely homologous with the 3' end of the dicistronic genomic RNA3 (3, 4). The method by which subgenomic RNAs are generated has been a matter of conjecture (4-8), but we recently demonstrated (9) that BMV RNA4 production proceeds via polymerization initiated at an internal site on (-)RNA3. In this paper, the internal promoter region of BMV (-)RNA3 is characterized and is shown to be very different in structure and sequence from the 3' tRNA-like region which promotes replication of the (-)-strands from (+)-strand genomic RNAs.

This work comprises the first such study of an internal subgenomic promoter of an RNA virus which infects eucaryotic cells.

These experiments utilized a replicase (RNA-dependent RNA polymerase) extract isolated from infected barley plants which is dependent on exogenous template and is highly specific for BMV RNA (10-12). Minus-strand replicase templates corresponding to wild type and mutant sequences were produced by *in vitro* transcription of cloned BMV RNA3 cDNA with SP6 or T7 RNA polymerase. For clarity, the term transcript is reserved in this article for template RNAs whereas the term "repliscript" (13) is applied to the replicase products of the RNA transcripts. Also, the term "repliscription" is used to distinguish the process of production of subgenomic mRNAs from that of replication of genomic RNAs. The term "wild type transcript" is used for template RNAs which contain the entire subgenomic RNA promoter region in an unaltered state (Figure 1A). "Downstream" is used in the sense of being in a 3' direction with reference to the (+)-strand and "upstream" is used in the sense of being in a 5' direction.

MATERIALS AND METHODS

Transcriptional vectors and mutagenesis

Four transcriptional vectors (Figure 1) were used in the isolation of the mutants described here. Three were derived from pSP4HE5, described previously (9). Two of the derivative vectors contain the cDNA insert of pSP4HE5 as a wild type sequence: mp3SP6HE5 and pH2T7HE5. The former is an M13 phage construct having the relevant region (including the SP6 promoter) of pSP4HE5 cloned into mp18. The latter was constructed with pIBI76 (International Biotechnology Inc.) and differs in that (-)RNA3 is transcribed with T7 RNA polymerase. These two vectors facilitate sequencing of mutants in that they allow isolation of single stranded DNA (pH2T7HE5 requires a helper phage). A vector, mpSPEI (Figure 2), derived from mp3SP6HE5 was used in the isolation of deletion mutants in the upstream end of the promoter, an *SpeI* site being inserted immediately 5' of the upstream (A-U) tract (see Figure 2).

Deletion mutations were constructed by cleaving the target area of the DNA with the appropriate restriction endonucleases, followed by digestion with mung bean nuclease. Conditions for restriction enzyme digestion were those recommended by the commercial source, New England Biolabs; mung bean nuclease digestions contained 30 mM sodium acetate (pH 4.6), 50 mM NaCl and 1 mM ZnCl₂. The DNA concentration was 1 µg per 10 µl (5 to 10 µg total) with 1 unit of enzyme per µg of DNA for production of blunt ends. For the lengthier deletions, where ladders of nested deletions were desired, 4 units of enzyme per µg of DNA were used. For production of blunt ends the incubations were for 10 min at 37°, but for production of lengthier deletions the digestions were extended to 30 min and samples were taken at intervals. The digests were extracted with phenol/chloroform and the DNA was

precipitated with 2.5 volumes of ethanol. The DNA samples were resuspended in 0.25 volumes of TE [10 mM tris-HCl (pH 8.0), 1 mM EDTA]. The DNA was religated at 16° C overnight and then digested again, using an excess of the restriction enzyme corresponding to the site being eliminated, prior to transformation or transfection. The resulting clones were initially screened by digesting the DNAs with appropriate restriction enzymes. Those clones from which the site or sites had been eliminated were characterized by dideoxynucleotide termination sequencing (14,15).

Oligonucleotide-directed site specific mutagenesis was performed as described previously (16). Mutants were initially screened by plaque-hybridization. Where appropriate, the clones were subsequently screened with restriction enzymes, and further characterized by sequencing.

In vitro transcription and repliscription

BMV (-)RNA3 was transcribed from the wild type vectors and the mutants with SP6 or T7 RNA polymerases in 100 μ l reactions using the conditions of Melton *et al.* (17), except that they contained 1 mM of ATP, CTP, and GTP, 0.5 mM UTP and 5 μ Ci of [³H]UTP. The RNA was isolated by extraction with phenol/chloroform, precipitated with ethanol, and resuspended in TE and stored at -80° C. The yield of the individual transcripts from the reactions was determined by precipitation of either 5% or 10% of the yield with cold 10% trichloroacetic acid and subsequent scintillation counting.

Replicase extracts were prepared as described previously (10, 11). Prior to use in the replicase assays, the background from endogenous viral RNAs was eliminated or greatly reduced by treatment of the extract with micrococcal nuclease (12). Replicase assays (25 μ l) were performed as described previously (9) using 1-2 μ g of (-)RNA3 template (adjusted to keep the molar concentration of mutant template equivalent to 2 μ g of wild-type template). In order to remove single stranded RNA, the reaction products were treated with RNase in high salt as described previously (10). The resulting double-stranded replicase products were subsequently analyzed on 3.5 % polyacrylamide gels with the buffer of Peacock and Dingman (17). A ladder of DNA fragments (One KB Ladder: Bethesda Research Laboratories) labeled with polynucleotide kinase and [γ -³²P]ATP was routinely used as a molecular weight standard. The relative yield of the expected repliscript was determined by excising the labeled bands from dried gels and measuring the incorporated ³²P by scintillation counting. Alternatively, after autoradiography of the dried gels with preflashed film (18), the autoradiograms were scanned with a BioRad model 620 video densitometer.

RESULTS

Characterization of the core promoter region

A series of truncated (-)RNA3 templates was obtained by transcription of cDNA linearized at the *Cl*aI, *B*glII, *S*alI and *B*ssHII restriction sites (see Figure 1). Such truncated templates are functionally analogous to deletions upstream of the restriction site at which the

Fig. 1. Map of the cDNA insert of the transcriptional vectors used to delineate the subgenomic promoter, the truncated transcripts and the expected repliscript. A proportional restriction map of the relevant region of the BMV RNA3 cDNA insert present in the transcriptional vectors is shown in A. The area of interest is expanded around the subgenomic promoter. The numbering of the bases given for the restriction sites corresponds to (+)RNA3 (3). The cDNA insert in all of the wild type transcriptional vectors is identical to that of pSP4HE5 (9). In addition to pSP4HE5, two other transcriptional vectors were also used in this work, mp3SP6HE5 and pH2T7HE5 (see Materials and Methods). The (▲) indicates the initiation site of the subgenomic repliscript and the (■■■■) indicates the approximate extent of the internal poly(A) tract of (+)RNA3. SpeI indicates the position of a restriction site introduced in mpSPEI to facilitate generation of truncated transcripts that included the upstream (A-U) tract of BMV (-)RNA3 (see Fig. 2). The SP6 RNA polymerase transcripts used as truncated replicase templates *in vitro* are shown in B, and the resulting repliscript is shown below in C. This repliscript is 200 bases shorter than RNA4 isolated from virus since the cDNA was cloned into the transcriptional vector at the HindIII site in the cDNA at position 1917, 200 bases upstream from the 3' terminus of RNA3.

transcriptional plasmids are linearized. These templates will be referred to by the enzyme used to linearize the cDNA for transcription e.g. the "ClaI transcript" was obtained from cDNA linearized at the ClaI site. These transcripts were then tested for replicase template activity *in vitro*. The expected repliscript was produced using the ClaI and BglIII transcripts (Figure 2A), but no correctly initiated repliscripts were obtained from templates shorter than the BglIII transcript. Increased levels of initiation at cryptic sites downstream of the correct site were

Fig. 2. Replicase products from truncated templates. Replicase products from truncated transcripts, transcribed from plasmid linearized at the indicated restriction sites, are shown in A.: (+)RNA4 (lane 1), template activity endogenous to the replicase extract (lane 2), *ClaI* product (lane 3), *BglIII* product (lane 4), *SalI* product (lane 5) and *BssHII* product (lane 6). Lane 7 was not loaded and lane 8 contained molecular weight standards. Full length copying of the templates is indicated by FL; E4 indicates the products of endogenous (+)RNA4 template; SG indicates the expected subgenomic repliscripts.

In B., RF DNA of mpSPEI (which contains an inserted *SpeI* site; see text) was linearized with *SpeI* and *ClaI* and RF DNA of the wild type vector, mp3SP6HE5, was linearized with *ClaI*. The desired (-)RNA3 sequences were transcribed using SP6 RNA polymerase and their template activity was measured in the repliscription assay. Replicase products of template activity endogenous to the replicase extract is shown in lane 4. Replicase products of *SpeI* and *ClaI* transcripts from mpSPEI are shown in lanes 2 and 3, respectively. The replicase products of the wild type *ClaI* transcript are shown in lane 1. Sequences of wild type and mutant RNAs [(+)-strands] are shown in panel C. The introduced *SpeI* site and the *BglIII* sites are underlined. The ^ precedes the G at which synthesis of the subgenomic RNA is initiated.

observed as the length of the supplied template was shortened. These data extended our previous observations (9) and provided strong evidence that the core sequence (the minimal sequence needed for detectable, correct, initiation) of the subgenomic promoter lies downstream from the *BglIII* site (Figures 1 and 2).

Although correctly initiated repliscripts were observed using the *BglIII* transcript as

Fig. 3. Replicase products from templates containing deletions downstream of the subgenomic initiation site. Both wild type and mutant transcriptional plasmids were linearized with *Cla*I prior to SP6 transcription. The replicase products from transcripts containing deletions constructed using downstream restriction sites are shown in A. Deletions in the supplied templates and their template activities relative to wild type (lane 4; 100%) were: lane 5 (*Sal*I-*Bss*HII; 73%), lane 6 (*Bss*HII-*Bal*I; 128%) and lane 7 (*Sal*I-*Bal*I; 62%). Lanes 2 and 3 show, respectively, the replicase products from endogenous RNA and (+)RNA4. Lane 1 contained molecular weight standards. The WT indicates the expected wild type repliscript; other products are denoted as in fig. 2.

Panel B shows the replicase products from a series of deletion mutations (constructed using the *Sal*I site which extend 5' into the downstream (A-U) tract (see text). Lanes 3-7 show the replicase products from the mutant transcripts, lane 2 shows the replicase products from the wild type transcript and lane 1 the replicase products from the endogenous template. The relative template activities (wild type = 100%) of the mutant transcripts in generating the correct subgenomic repliscripts (SG) were: 63% (lane 3), 28% (lane 4), 27% (lane 5), 15% (lane 6) and 13% (lane 7). FL indicates the full length end to end copying of the templates. The sequences of the wild type (lane 2) and mutant constructs are shown in C. Data represent averages from two separate experiments. The ^ precedes the initial G of the subgenomic RNA.

template, products of cryptic initiation events were also evident. Cryptic products were only faintly seen among the repliscripts from the longer, *Cla*I, transcript. The more precise initiation, and greater yield, of the correct repliscript observed with the *Cla*I template

compared with the *Bgl*II template suggested that sequences upstream from the *Bgl*II site modulate the accuracy and strength of the promoter core region. It appeared likely that such modulating sequences should lie close to the *Bgl*II site, but no convenient endonuclease restriction site existed 5' proximal to the *Bgl*II site in the wild type cDNA. In order to circumvent this difficulty, an *Spe*I site was created by the insertion of three bases immediately upstream of the (A-U) rich tract which includes the intercistronic poly(A) sequence of RNA3 (Figure 2C). The *Spe*I transcript was observed to be as efficient a template for repliscription as was the *Cla*I template, yielding approximately wild type levels of the correctly initiated subgenomic repliscript and similar trace amounts of the cryptic products (Figure 2B). These results implied that the (A-U) tract between the *Bgl*II and *Spe*I sites was involved in modulating the yield and correct initiation of repliscript by the subgenomic promoter core sequence.

Mapping the 3' extremity of the subgenomic promoter

The preceding experiments did not define the downstream limit of the core promoter region. Therefore, a series of mutants was constructed using convenient restriction sites in the cDNA (Figure 1A). These included deletion of the *Sal*I-*Bss*HII region, the *Bss*HII-*Bal*I region, and the entire *Sal*I-*Bal*I region. Transcripts containing these deletions were efficient templates for repliscription of correctly initiated products of the expected lengths (Figure 3A). Several other deletion mutants were constructed using the restriction sites shown in Figure 1. In contrast with templates having deletions downstream from the *Sal*I site, all transcripts in which the *Bgl*II-*Sal*I region was deleted were inactive templates for repliscription (data not shown). These experiments provide further evidence that the *Bgl*II-*Sal*I region includes the core sequence of this RNA promoter and that no obligatory promoter elements are located downstream of the *Sal*I site.

In order to further delineate the 3' extent of the promoter core sequence, a nested series of deletions was constructed (Figure 3B). These extended 5' of the *Sal*I site into a short (A-U) tract that lies between the subgenomic initiation site and the *Sal*I site. Provided the base at which the subgenomic RNA initiates was not eliminated, some correct initiation was observed. However, as the 5' extent of the deletion increased, the yield of the correct subgenomic repliscript decreased. Conversely, increasing amounts of end to end copying of the exogenous template were observed as the deletion length increased, resulting in approximately equal amounts of total products. Therefore, it appears that deletion into this (A-U) tract separates the process of correct initiation from that of promotion, with the (A-U) tract downstream from the subgenomic initiation site being involved in determining correct initiation of the subgenomic RNA4 rather than having a primary function in modulating promoter strength. Indeed, in the case of the largest deletion (Figure 3B, lane 7) replicase products were observed that originated at cryptic sites both upstream and downstream from the correct initiation site.

Fig. 4. Replicase template activity of deletions in the upstream (A-U) tract. Repliscripts from transcripts with localized deletions extending 3' from the *SpeI* site into the upstream (A-U) tract are shown in A. The vector DNAs were linearized with *ClaI* prior to transcription. Lane 1 contained molecular weight standards; lane 2 contained the replicase products of template from mpSPEI; lanes 3-8 contained replicase products from deletion mutants; and lane 9 contains template activity endogenous to the replicase extract. The

relative template activities of the mutants in generating the subgenomic repliscript, SG, (mpSPEI, lane 2 = 100%) were; lane 3 = 27%, lane 4 = 17%, lane 5 = 14%, lane 6 = 19%, lane 7 = 12% and lane 8 = 10%. The data represent the averages of two separate experiments. The sequences of mpSPEI and the derived mutant are shown in C. The *SpeI* site and bases in the mutant (A-U) tracts which differ from wild type are underlined.

Panel B shows the repliscripts produced from transcripts of mutants containing deletions obtained by selection for loss of the *BstXI* and *SpeI* sites. These constructs had extensive sequences deleted 5' of the subgenomic promoter core sequence, but the 3' extent of the deletions nested within the UUAUUAUU block. The vector DNAs were linearized with *EcoRI* prior to transcription. Lane 1 contains replicase products from template endogenous to the replicase extract; Lane 2 contains the replicase products of template from mpSPEI and lanes 3-7 contain replicase products of deletion mutants. The bases deleted and relative template activities of the mutants (mpSPEI, lane 2 = 100%) were: lane 3 (880-1191) = 43%, lane 4 (1157-1193) = 69%, lane 5 (167-1195) = 124%, lane 6 (1157-1197) = 29%. Lane 7 (12%) contained the same mutant as that shown in lane 5 of panel A. The data represent the averages of two separate experiments. The corresponding sequences of the upstream (A-U) tracts of mpSPEI and the derived mutants are shown in D. Symbols are as denoted in fig. 2. The ^ precedes the initial G of the subgenomic mRNA.

Further investigation of the 5' modulating region

The modulation effects revealed by the downstream deletion series shown in Figure 3 suggested that further analysis of the upstream modulating region was appropriate. Consequently, a series of localized deletions which extended 3' into the upstream (A-U) tract [that includes the intercistronic poly(A)] was constructed by digestion with mung bean nuclease after linearization with *SpeI*. Transcripts containing these deletions generally yielded only 10 to 20% of wild type activity (Figure 4A), suggesting that this region does contribute to promoter strength. Only very low levels of cryptic products were present, indicating that this upstream (A-U) tract makes only a minor contribution to the enhancement of correct initiation seen with the *SpeI* transcript (Figure 2). In light of the low activity of the above mutants, transcripts containing deletions surrounding the *SpeI* site were tested for replicase template activity (Figure 4B). In accord with the preceding results, deletion of all, or most, of the UUAUUAUU sequence between the *SpeI* site and the poly(A) tract produced templates with only a fraction of wild type activities (lanes 5 and 6 of Figure 4C). In contrast, those mutants with most of the UUAUUAUU sequence intact had template activities approaching or equaling that of wild type templates. These mutants (except that shown in lane 6) delete the region between the *BstXI* site and the *SpeI* site; two of the mutants (lanes 3 and 5) deleted much further upstream. Therefore, these deletions delimited the 5' extent of the subgenomic promoter to a position immediately downstream of the *SpeI* site.

Taken together, the data of Figures 2B, 4A, and 4B accentuate the importance of the UUAUUAUU block immediately upstream of the poly(A) tract. Comparison with sequences upstream of the subgenomic initiation sites of other plant viruses reveals blocks that are similar to, and probably evolutionarily and functionally homologous with, this UUAUUAUU block (Table I). However, this block is not separated from the other homologous sequence

Table I. Sequence homologies between the intercistronic regions of BMV RNA3 and other plant viral RNAs.

BMV	- UUAUUAUUaaaaaaaaaaaaAAGAUCUA	(3) bases 1192-1223
AlMV	- UU.UUAUU.....AACA	(20) bases 1109-1120, (21) bases 1127-1138
TRV	- UUAUUAU.....AAGA	(22) bases 518-525
TRV	- UUAUUGUc.....aagugAAGA	(23) bases 1551-1567
TMV	- UUCaU.....AAGA	(24, 25) bases 4813-4821
BSMV	- UUAUUUAUUUU.....guuaUCUA	(26) bases 657-677
BSMV	- UUAUUUGAUcuuuucaaaacaaacauUCUA	(26) bases 2437-2467

Spaces introduced for alignment are indicated by (...) and conserved sequences are underlined. The UCUA sequence is similar to the conserved block in the alphavirus subgenomic sequences (see Table II). Abbreviations used here and in the text are: brome mosaic virus (BMV); alfalfa mosaic virus (AlMV); tobacco rattle virus (TRV); barley stripped mosaic virus (BSMV); and cucumber mosaic virus (CMV).

elements of the promoter by an extensive internal poly(A) tract in the putative subgenomic promoters of other plant viruses shown in Table I.

In order to test the role of the internal poly(A) tract in the BMV promoter, mutants with poly(A) tracts of only (A)₅ and (A)₁₀ were constructed by oligonucleotide site-directed mutagenesis. The mutant with a internal tract of (A)₁₀ yielded approximately wild type levels of replicase product, but the (A)₅ mutant showed only a low level of replicase activity (< 20% of wild type; data not shown). The latter observation seems to indicate that, for BMV replicase, there is a spacing requirement between the UUAUUAUU block and the core sequence of the promoter. The poly(A) tract may serve as a non base-paired spacer to facilitate access of the replicase to the promoter. Computer modeling of the secondary structure of the subgenomic promoter region (Figure 5) indicates that the upstream (A-U) tract is probably not base-paired. This contrasts with the genome of this virus as a whole, which generally shows considerable secondary structure.

Comparison with the putative alphavirus subgenomic promoters

Ou *et al.* (28) noted that a conserved sequence of 19 bases is present upstream of the subgenomic initiation site of several alphaviruses. Alignment of the BMV subgenomic core sequence (approximately 20 bases) defined above with this alphavirus sequence shows close homology (Table II). Indeed, additional homology is revealed between the GCGUA sequence surrounding the BMV RNA4 initiation site and a region downstream from the presumed initiation site for the alphavirus subgenomic RNAs. Three regions (totalling 16 bases) having complete homology with the alphavirus consensus promoter can be identified within the 20 bases of the BMV RNA3 promoter core sequence (Table IIB), although the spacing is not strictly conserved. The UCUA motif common to BMV RNA3 and the alphavirus RNAs can also

Fig. 5. A secondary structure for (-)RNA3 in the region of the subgenomic promoter generated with a program based on the algorithm of Zuker and Steigler (27). The (→) indicates the direction of RNA4 repliscription and (*) indicates the base opposite the first base of RNA4. The brackets indicate the bases corresponding to restriction sites on the cDNA.

be found upstream of the subgenomic initiation site in several plant virus RNAs, and is often preceded by, or replaced by, an AAGA sequence in these viruses (Table I). Similarly, although the AAGA block is not found in the putative promoters of the alphaviruses (see Table II), this sequence is found somewhat upstream. The AAGA sequence is found at position -54 (54 bases upstream of the subgenomic initiation site) in Sindbis virus, and at -53 in Ross River virus.

Table II. Comparison of the postulated alphavirus promoters to that of BMV

A. Postulated alphavirus promoters (28).

SIN AUCUCUACGGUGGUCCUAA^AUAGUCAGCAUA
 MBV ACCUCUACGGCGGUCCUAA^AUAGUUGCGUGA
 SFV ACCUCUACGGCGGUCCUAA^AUUGGUGCGUUA
 RRV ACCUCUACGGCGGUCCUAA^AUAGAUGCAGAG
ACCUCUACGGCGGUCCUAA^AUAGUUGCGUUA -consensus

B. Postulated BMV subgenomic promoter sequence.

Bgl II Sal I

UUAUUAUAAAAAAAAAAAAAAAAAAGAUCUAUGUCCUAAUUCAGC^GUAUUAUAAUGUCGAC

C. Comparison of homologous sequence blocks

Alphaviruses -	UCUA	GUCCUAA	GC GUUA
BMV -	UCUA	GUCCUAA	GC^GUA

See text for details. Blocks of homologous sequence between the BMV subgenomic promoter core sequence and the consensus sequence of the postulated alphavirus promoters are underlined. The ^ precedes the initial base of the subgenomics.

Semliki Forest virus has the sequence AAGG at position -53, but an AAGA sequence is present at -44. In addition to the AAGA sequence, Ross River virus has AAAGC at position -41. Middleburg virus has an AAGA sequence located at position -36. A more thorough comparison of the subgenomic promoter regions of these viruses is in preparation (L. E. Marsh and T. C. Hall).

DISCUSSION

Our results indicate that the complete subgenomic promoter of BMV RNA3 encompasses approximately 62 bases [1192-1253, (3)] and includes four functional domains: the core sequence, the poly(A) tract, the upstream UUAUUAUU block and the downstream (A-U) tract. The promoter core sequence comprises some twenty bases immediately upstream of the initiation site, and contains four sequence blocks with homology to either the postulated alphavirus promoter sequences or the intercistronic regions of other plant viruses: AAGA, UCUA, GUCCUAA and GC[^]GUA (the [^] symbol precedes the first nucleotide of the subgenomic repliscript). Our evidence that the core sequence of the promoter lies upstream of the AUG is in agreement with results described by French *et al.* (29) in which chloramphenicol acetyltransferase gene was expressed when inserted at the *SalI* site. Similarly, analysis of the sequences required for replication of genomic RNA3 (30) indicated a location for the core sequence of the promoter that is in agreement with the data presented here. The poly(A) tract [present as poly(U) in the (-)RNA] may serve primarily as an unstructured spacer, a possible function being to facilitate access of the replicase to the core sequences of the promoter. This possibility is also consistent with the natural heterogeneity in the length of the poly(A) tract (31), which can be partially deleted without affecting repliscript levels. The lack of apparent base-pairing in the upstream (A-U) tract contrasts with the viral genome as a whole which is generally base-paired. Evidence that the downstream (A-U) tract is involved primarily in either determining or allowing correct initiation is derived from the finding that deletions extending into the downstream (A-U) tract seemingly can separate promotion of repliscription from correct initiation of repliscription. The abnormal initiation events observed upon deletion into this downstream (A-U) tract (Figure 3B) are apparently the result of cryptic sequences which function as initiation sites (probably due to a fortuitous resemblance to the correct site) only upon alteration of the normal initiation site. The presence of products initiated at cryptic sites implies that the formation of an active replicase complex and initiation of replication are separate steps in promotion. Presumably, the active replicase complex assembles at the core sequence of the promoter and subsequently searches both up and downstream for an initiation site. Normally, the correct initiation site immediately downstream (and possibly overlapping with the core sequence of the promoter) is used. End-to-end copies of the (-)-strand template occur when the 5' end of the template serves as a

cryptic initiation site. Such events may mimic initiation of the 5' ends of the (+)-strand genomic RNAs and, although the subgenomic promoter core is not essential for replication of genomic RNA3, the poly(A) tract of the subgenomic promoter does appear to contribute actively (30). Indeed, homologies recognized between sequences at the 5' termini of the genomic RNAs and the subgenomic promoter (32) are likely to be significant in the overall replication strategy of this virus.

The upstream UUAUUAUU block apparently has a specific function in that deletion of it alone reduces repliscription to a minimal level and deletions further into the poly(A) tract do not cause significantly greater reduction in the repliscript yield. The UUAUUAUU block and a complete poly(A) tract are not required for correct initiation of repliscription, but do contribute strongly to the yield of correct repliscript. Consequently this block may be thought to be analogous to the enhancers of orthodox eucaryotic promoters, but we have no evidence that this sequence element will function in different positions or orientations with respect to the rest of the promoter.

The AAGA and UCUA sequence blocks (Table I) are contiguous in BMV, but when the putative subgenomic promoter sequences of other plant (+)RNA viruses are examined, these sequence blocks are often spaced somewhat differently. Often, the AAGA block is not part of the core sequences but is located somewhat upstream, usually near the homolog of the UUAUUAUU block (Table I). Although the subgenomic promoters of the plant viruses typically possess both up and downstream (A-U) rich tracts, their relative length and spacing is quite variable. For example, that of alfalfa mosaic virus has only a short (A-U) rich region upstream, but possesses a rather extensive (A-U) rich region downstream. In contrast, cowpea chlorotic mottle virus possesses a short downstream (A-U) tract, GUAAUUUAUCAUG (4), which resembles the GUAUUAAUAUG sequence of BMV. Most of the other viruses examined had a block homologous to the UUAUUAUU block of BMV located immediately upstream of the AAGA block (Table I), whereas BMV has these blocks separated by the internal poly(A) tract. That the conserved sequence blocks of the subgenomic promoters show this variable spacing is in contrast to more familiar eucaryotic and procaryotic transcriptional promoters. The freedom apparently allowed the (+)-stranded RNA viruses in the evolution of the spacing of the sequence elements of their subgenomic promoters may in part reflect the nature of the template, single stranded RNA is inherently more flexible than double stranded DNA. Furthermore, the replicases may recognize and interact with a tertiary structure. This is seen in the case of the tRNA-like structures which are the promoters at the 3' ends of the genomic (+)RNAs of BMV (33, 34).

French *et al.*, (29) have already demonstrated that BMV can be used as an expression vector; an intimate understanding of subgenomic promoters is likely to be of use in optimizing the design and construction of such vectors from BMV and other (+)-stranded RNA viruses.

Indeed, the detailed deletion analysis and delineation of the subgenomic promoter of BMV presented here may guide elucidation of subgenomic promoters of other (+)-stranded RNA viruses, including animal viruses such as the alphaviruses.

ACKNOWLEDGEMENTS

Din Pow Ma is thanked for technical advice and support with respect to sequencing of the mutants and for synthesis of the mutagenic oligonucleotides. We gratefully acknowledge the excellent technical support of Jillaine Maes, Victoria Vastine and Roy Stevens. This work was supported by NIH grant AI22354 and the Texas A&M Office of University Research.

*To whom correspondence should be addressed

REFERENCES

1. Lane, L.C., Kaesberg, P. (1971) *Nature New Biology* **232**, 40-43.
2. Shih, D.S., and Kaesberg, P. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 1799-1803.
3. Ahlquist, P., Luckow, V., and Kaesberg, P., (1981) *J. Mol. Biology*, **153**, 23-38.
4. Dasgupta, R., and Kaesberg, P. (1982) *Nucleic Acids Res.* **10**, 703-713.
5. Gonda, T.J. and Symons, R.H. (1979) *J. Gen. Virol.* **45**, 723-736.
6. Keranen, S., and Kaariainen, L., (1979) *J. Virol.* **32**, 19-29.
7. Nassuth, A., Alblas, F., and Bol, J.F. (1981) *J. Gen. Virol.* **53**, 207-214.
8. Goelet P, and Karn, J. (1982) *J. Mol. Biol.* **154**, 541-550.
9. Miller, W.A., Dreher, T.W. and Hall, T.C. (1985) *Nature* **313**, 68-70.
10. Hardy, S.F., German, T.L., Loesch-Fries, L.S. and Hall, T.C. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4956-4960.
11. Bujarski, J.J., Hardy, S.F., Miller, W.A. and Hall, T.C. (1982) *Virology* **119**, 465- 473.
12. Miller, W.A. and Hall, T.C. (1984) *J. Virol.* **125**, 236-241.
13. Hall, T.C., Dreher, T.W. and Marsh, L.E. (1987) In Arntzen, C.J. and Ryan, C. (ed.), Molecular Strategies in Crop Protection. UCLA Symposium Series, Alan R. Liss Inc., New York, pp. 295-305.
14. Sanger, F. Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
15. Messing, J. (1983) *Methods in Enzymology* **101**, 20-78.
16. Zoller, M.J. and Smith, M. (1984) *DNA* **3**, 479-488.
17. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) *Nucleic Acids Res.* **12**, 7035-7056.
18. Peacock, A.C. and Dingman, C.W. (1968) *Biochemistry* **7**, 668-674.
19. Laskey, R.A. and Mills, A.D. (1977) *FEBS Lett.* **82**, 314-316.
20. Barker, R.F., Jarvis, N.P., Thompson, D.V., Loesch-Fries, L.S. and Hall, T.C. (1983) *Nucleic Acids Res.*, **11**, 2881-2891.
21. Ravelonandro, M., Pinck, M., and Pinck, L. (1984) *Biochimie* **66**, 395-402.
22. Bergh, S.T., Koziel, M.G., Huang, S.C., Thomas, R.A., Gilley, D.P. and Siegel, A. (1985) *Nucleic Acids Res.* **13**, 8507-8518.
23. Boccardo, M., Hamilton, W.D.O. and Baulcombe, D.C. (1986) *EMBO J.* **5**, 223-229.
24. Goelet, P., Lomonosoff, G.P., Butler, P.J.G., Akam, M.E., Gait, M.J. and Karn, J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5818 -5822.
25. Watanabe, Y., Meshi, T. and Okada, Y. (1984) *FEBS Lett.* **173**, 247-250.
26. Gustafson, G. and Armour, S.L. (1986) *Nucleic Acids Res.* **14**, 3895-3909.

27. Zuker, M. and Steigler, P. (1981) *Nucleic Acids Res.* **9**, 133-148.
28. Ou, J.H., Rice, C.M., Dalgarno, L., Strauss, L. and Strauss, J.H. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5235-5239.
29. French, R., Janda, M. and Ahlquist, P. (1986) *Science* **231**, 1294-1297.
30. French, R. and Ahlquist, P. (1987) *J. Virol.* **61**, 1457-1465.
31. Ahlquist, P. and Kaesberg, P. (1979) *Nucleic Acids Res.* **7**, 1195-1204.
32. Marsh, L. and Hall, T.C. (1988) *Cold Spring Harbor Symposia on Quantitative Biology* **52**, in press.
33. Ahlquist, P., Bujarski, J.J., Kaesberg, P., and Hall, T.C. (1984) *Plant Mol. Biol.* **3**, 37-44.
34. Bujarski, J.J., Dreher, T.W., and Hall, T.C. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5636-5640.