Specific site selection in RNA resulting from a combination of nonspecific secondary structure and -CCR- boxes: Initiation of minus strand synthesis by turnip yellow mosaic virus RNA-dependent RNA polymerase

RAVINDRA N. SINGH^{1,3} and THEO W. DREHER²

¹ Department of Microbiology, Oregon State University, Corvallis, Oregon 97331-3804, USA

² Center for Gene Research and Biotechnology, Oregon State University, Corvallis, Oregon 97331-3804, USA

ABSTRACT

A turnip yellow mosaic virus RNA-dependent RNA polymerase activity was used to study the template requirements for *in vitro* minus strand synthesis, which is initiated specifically opposite the 3'-CCA that terminates the 3'-tRNA-like structure. A deletion survey confirmed earlier results suggesting the absence of minus strand promoter elements upstream of the pseudoknotted acceptor stem and 3'-terminus. Reiteration of this 27-nt domain provided two competing initiation sites. By varying the added downstream element, it was shown that the pseudoknotted domain could be functionally replaced by various simple stem/loops, although with some decrease in activity. The addition of varying numbers of consecutive -CCA- triplets to the 3' end of the tRNA-like structure resulted in accurate initiation from each added triplet. A similar spectrum of initiations occurred with an unstructured RNA consisting of 12 consecutive -CCA- triplets and no additional viral sequence. Substitution mutations revealed no influence on minus strand synthesis of the identity of the nucleotide immediately upstream of a -CC- initiation site, but a preference for a purine immediately downstream. The introduction of secondary structure into the linear template showed that the usage of potential -CCR- initiation sites is influenced by nonspecific secondary structure. We conclude that specificity arises from the requirement that a -CCR- sequence be sterically accessible. This mechanism is only applicable to interactions that do not involve RNA unwinding during site selection, but may be used commonly in positive strand RNA virus replication and be applicable to other RNA-protein interactions.

Keywords: *in vitro* transcription; positive strand RNA virus

INTRODUCTION

Specific RNA recognition by proteins underlies many biological processes that demand high-fidelity performance. In the best-described examples, highly specific outcomes are the result of the recognition of RNA elements that comprise short sequence-specific features placed in a defined structural framework. Thus, specific aminoacylation of transfer RNAs by the cognate aminoacyl-tRNA synthetase requires the accurate placement of a few identity nucleotides within the generic tRNA structure that is approximately 76 nt long (Pallanck et al., 1995). Shorter elements (ca. 20–30 nt long) rec-

ognized by combined structural and sequence information are the helix/loop combinations bound by the HIV Tat protein (Puglisi et al., 1992), phage R17/MS2 coat protein (Valegård et al., 1994), and U1A spliceosomal protein (Oubridge et al., 1994).

This is not the only way in which proteins recognize specific sites on RNA, however. For instance, evidence exists that the phage T4 translational repressor, regA, recognizes a consensus sequence of some 12–15 linear nucleotides in unstructured RNA (Szewczak et al., 1991; Brown et al., 1997). Here, as a result of studying the replication of a positive strand RNA virus, we report a further strategy for specific RNA recognition that requires a combination of a very short specific-sequence and adjacent non- or low-specificity secondary structure.

Positive strand RNA viruses replicate via two sequential transcriptional steps that synthesize full-length minus and plus sense genomes; additionally, in some

Reprint requests to: Theo W. Dreher, Department of Microbiology, Oregon State University, Corvallis, Oregon 97331-3804, USA; e-mail: dreher@bcc.orst.edu.

³ Present address: Institute of Cellular and Molecular Biology, University of Texas, Austin, Texas 78712, USA.

viruses, internal initiation on the minus strand (Miller et al., 1985) results in the synthesis of subgenomic RNAs that are collinear with the 3'-region of the genomic RNA and that serve as mRNAs expressing downstream cistrons (Buck, 1996). Specific initiation sites are used for each of these transcriptional events, and a breakdown of the fidelity of initiation site selection would lead to truncated genomes and viral proteins. The *cis*-required promoter elements controlling these specific strand initiations have been studied in a number of viruses by in vivo approaches using deleted genomes and by in vitro approaches using viral RNA-dependent RNA polymerase (RdRp) preparations (Buck, 1996). These studies have generally supported the view that accurate transcription is controlled by the detection of specific features of either sequence or a combination of sequence and structure (e.g., Miller et al., 1986; Dre-

her & Hall, 1988; Levis et al., 1990; Cui & Porter, 1995; Song & Simon, 1995; Miranda et al., 1997; Siegel et al., 1997).

Against this backdrop of apparently specific recognition of defined, discrete promoter elements, we have studied the features directing minus strand synthesis by the turnip yellow mosaic virus (TYMV) RdRp. Minus strand synthesis initiates specifically opposite the penultimate residue (C2; Fig. 1A) (Singh & Dreher, 1997) in the 3'-ACCA that terminates the 82-nt-long tRNA-like structure (TLS), the prominent feature at the 3' end of the genomic RNA. The TLS has an overall L conformation like that of canonical tRNAs, although the amino acid acceptor stem consists of a pseudoknot structure (nt 1–27, numbered from the 3' end; see Fig. 1A) not found in tRNAs (Rietveld et al., 1983). Previous in vitro experiments have shown that, although the TYMV RdRp

FIGURE 1. Search for promoter elements in the TYMV TLS upstream of the pseudoknotted acceptor stem. **A:** Linear representation of the 83-nt wild-type TLS RNA (top) indicating base pairing. Mutant derivatives are shown below, with unmodified segments shown as solid lines, deletions as dashed lines, and substitutions with the new nucleotides indicated. Note that nucleotide numbering is from the 3' end, because this RNA is a 3' fragment of the 6,318-nt-long viral RNA. **B,C:** Transcription products made by TYMV RdRp from the templates indicated above each lane. Seven picomoles of each template RNA were present in reactions labeled with [α - 32 P]UTP; 7 pmol of TLS RNA was additionally present in the reactions of lanes B9–14. Products were treated with ribonuclease in high salt to remove excess template RNAs, and the protected, double-stranded RNAs were separated by 8% sequencing PAGE and detected on a phosphorimager.

discriminates against other full-length viral RNAs, it is able to transcribe quite efficiently a range of unmodified tRNAs and tRNA-like structures (Singh & Dreher, 1997). The TLS and genomic-length templates were copied with similar activity, and 3'-TLS fragments as short as 21 nt could be transcribed from the usual initiation site with at least half the template activity of TLS RNA (Deiman et al., 1997; Singh & Dreher, 1997). Further, mutations within the acceptor stem pseudoknot again resulted in at least half maximal template activity (Deiman et al., 1997; Singh & Dreher, 1997). These results brought into question the existence of specific sequence or structural promoter elements directing minus strand synthesis, a notion further supported by the amplification in vivo of a number of chimeric TYMV genomes with heterologous 3' termini of varying sequence (Skuzeski et al., 1996; Goodwin et al., 1997).

Our further enquiries into the specificity of minus strand initiation reported here have led to the conclusion that the -CCA sequence immediately at the minus strand initiation site is the only identifiable specific requirement for initiation. Specificity is further guided by nonspecific secondary structure. We believe that this type of specificity scheme applies to promoter selection in other positive strand RNA viral systems, and may also be a strategy used by other RNA-binding proteins in specific interactions with RNA.

RESULTS

Confirmed absence of promoter elements upstream of the acceptor stem pseudoknot

Concerned that the truncation experiments used previously may be unable to detect the presence of promoter elements in the TLS that functioned in longer templates, we systematically scanned the TLS domains upstream of the acceptor stem for effects on minus strand synthesis with discrete structural deletions or sequence substitutions (Fig. 1A). Sequence changes that cumulatively introduced radical changes into the anticodon, D-, and T-domains of the TLS failed to significantly affect minus strand transcription (Fig. 1B,C; refer to wild-type TLS in lanes 2 as control). The mutations included replacement of the D- and anticodon loops with GAAA tetraloops (Fig. 1B, lanes 3 and 9; 5 and 11), deletion of the entire D- and anticodon arms (Fig. 1B, lanes 4 and 10; 6 and 12), sequence changes in the T-loop (Fig. 1B, lanes 7 and 13; Fig. 1C, lanes 6–9) and in the T-stem (Fig. 1C, lanes 3, 4, 5, and 10), and deletion of the entire T-arm (Fig. 1B, lanes 8 and 14). RNAs with deletions in the anticodon, D-, and T-domains were equivalent templates to the wild-type TLS RNA in competition experiments (Fig. 1B, lanes 9–14).

These mutations clearly indicate the absence of detectable promoter elements upstream of the acceptor

stem pseudoknot (nt 27), results that are consistent with the high template activity of 3' 28-nt-long RNA (Deiman et al., 1997; Singh & Dreher, 1997). This conclusion is also consistent with the ability of various tRNA-like RNAs to serve as templates despite widely differing primary sequence (Singh & Dreher, 1997). The absence of promoter elements from the T-loop and T-stem, conserved in sequence to some degree among all the tRNA-like templates tested previously, is demonstrated by the uniformly high transcription of RNAs with mutant T-loop and T-stem sequences (Fig. 1C), and even deletion of the entire T-arm (Fig. 1B, lanes 8 and 14).

Duplication of the acceptor stem results in two competing initiation sites

If specific promoter elements are absent upstream of nt 27, any specific information directing template selection by RdRp must be located in the acceptor stem pseudoknot (nt 5–27) and/or the 3'-ACCA. To further examine the previously identified weak influence of the pseudoknot in minus strand synthesis, we created derivatives of the TYMV TLS with an additional amino acid acceptor domain added to the normal 3'-A, connected through either an A₃ or an A₁₅ spacer (TLS-112 or TLS-127 RNAs; Fig. 2A). This provided a new potential initiation site at the 3' end in addition to the normal, now internal, site. Mutation of the downstream pseudoknot permitted a study of the effect of mutations in the 3'-terminal 27 nt on the competition for minus strand initiation at the two sites.

For both spacer lengths (A₃ or A₁₅), initiation occurred principally at the new 3'-terminal -ACCA, with less initiation at the internalized -ACCA (Fig. 2B, lanes 2 and 6). Examination of template RNAs by ethidium bromide staining at the conclusion of the polymerase reaction showed no evidence of RNA fragmentation (not shown), indicating that the TLS-sized products resulted from internal initiation. For both spacer lengths, the effects of deleting nucleotides in the upstream half of the added pseudoknot (Fig. 2A) were studied. Deletion of nt 22–27 (Δ S1 mutation) was designed to disrupt the tertiary base pairing that supports the pseudoknot, leaving a potential stem/loop, whereas deletion of nt 16–27 (Δ S1+2 mutation) disrupted both base paired segments of the pseudoknot. The Δ S1 mutation decreased the ratio of 3'-end to internal initiation by about fivefold with both spacer lengths (Fig. 2B, lanes 3 and 7), whereas the Δ S1+2 mutation decreased this ratio by 24-fold with an A₃ spacer and 13-fold with an A₁₅ spacer (Fig. 2B, lanes 2 versus 4 and 6 versus 8, respectively). These results underscore previous findings indicating a role for the pseudoknotted acceptor stem in minus strand synthesis, and, as found previously, that role appears to be secondary, evidenced by the significant 3'-end initiation observed for templates with the Δ S1 mutation (Fig. 2B, lanes 3 and 7), which have a stem/

FIGURE 2. Transcription from two initiation sites on RNAs with a duplication of the acceptor arm. **A:** Expected structure of two RNAs with a duplicated acceptor arm, TLS-112, a 112-nt-long RNA that has an A_3 spacer, and TLS-127, a 127-nt-long RNA that has an A_{15} spacer. The tertiary conformation of the TLS shown toward the left has been determined by structure probing (Rietveld et al., 1983; Mans et al., 1991). Within the TLS, the anticodon arm (marked A; CAC valine anticodon is underlined), D-arm, and T-arm are labeled, and the folding of the pseudoknot in the acceptor stem is indicated. The deletion mutations $\Delta S1$ and $\Delta S2$ are boxed in the extra acceptor arm domain; the deletion $\Delta S1+2$ represents deletion of both sets of boxed nucleotides. **B:** Transcription products labeled with [α - 32 P]ATP from the indicated RNA templates, analyzed as in Figure 1. The arrowhead marks product initiated from the usual TLS site; slower-migrating products represent initiations from the 3'-CCA. The molar ratios of products initiating from the 3' relative to the internal site are indicated (average of two replicates).

loop in place of the pseudoknot. These results were reminiscent of the definition of the stem/loop promoter element at the 3' end of TCV satellite RNA C (Song & Simon, 1995). The overall higher rates of initiation at the internal site in the presence of an A_{15} compared to A_3 spacer (three to six times higher) suggests that accessibility of a potential initiation site plays a role in initiation site selection, although an influence of distance between initiation site and structural elements is also possible.

Influence of secondary structure adjacent to the initiation site

To investigate further whether TYMV RdRp responds specifically to a structural element adjacent to an -ACCA initiation site, the additional acceptor stem domain was redesigned structurally while conserving aspects of the TYMV sequence. Conversion of the 12-base pair acceptor/T arm composed of three stacked helical segments to a continuous 13-base pair stem and six-membered loop (T+S13L6 RNA; Fig. 3A) resulted in a low ratio (0.31) of 3'-end to internal initiation (Fig. 3B, lane 1). Similar RNAs with 9, 8, and 7 base pair stems (T+S9L6, T+S8L3, and T+S7L4 RNAs; Fig. 3A) resulted in increased initiation at the 3' end (Fig. 3B, lanes 2, 4, 3, respectively). The ratios of 3'-end to internal initiation observed for the RNAs with seven- and eight-base pair 3'-terminal stems (3.4 and 3.8; Fig. 3B,

lanes 3 and 4) were about threefold lower than for the parental RNA with a 3' pseudoknot (TLS-112 RNA; Fig. 2B). This difference could correspond to the previously determined small influence of the pseudoknotted conformation on minus strand synthesis (Deiman et al., 1997; Singh & Dreher, 1997), or it could be due to differences in accessibility to the internal initiation site.

The above results show that a conventional stem/loop can function almost as well as the wild-type pseudoknot in directing initiation to an adjacent -ACCA. Based on T+S8L3 RNA, further mutants were designed to test the influence of alterations of the secondary structure adjacent to a potential initiation site. In RNAs BOT Δ 1S, BOT Δ 2S, and BOT Δ 3S, one, two, or three base pairs at the bottom of the stem were disrupted, whereas in RNAs TOP Δ 1S, TOP Δ 2S, and TOP Δ 3S, one, two, or three base pairs at the top of the stem were disrupted (Fig. 3A); the solution conformations of these RNAs were not probed, so it is possible that structural changes other than the base pair disruptions just mentioned are present in these RNAs. The RNAs were designed to avoid -ACCA- sequences that may serve as initiation sites within the loops. The disruptions of the lower and upper base pairs (which may also be viewed as mutations of nucleotides near the 3' end) led to decreased relative initiation at the 3' end (Fig. 3B, lanes 5–10, cf. lane 4). Notably, progressive disruption of the upper base pairs (Fig. 3B, lanes 8–10)

FIGURE 3. Role of RNA structure, studied with variants of TLS-112 RNA. **A:** Plausible structures for the extra 3' domains in RNAs used as RdRp templates. Top row shows the pseudoknotted acceptor/T arm of the TYMV TLS and the additional pseudoknotted 3' domain of TLS-112 RNA for reference, together with derivatives in which the pseudoknot of TLS-112 has been replaced with simple stem/loops. Names are indicated for each RNA; note that the stem sequence of T+S13L6 RNA is derived from the wild-type acceptor/T arm shown at top left. The lower row shows two sets of three variants each derived from T+S8L3 RNA, in which lower base pairs (to the left) or upper base pairs (to the right) of the 3' stem have been disrupted. Note that the structures of these RNAs have not been probed experimentally; in one case, a possible isomeric form is shown. Arrows mark possible novel initiation sites in the loops of some of the RNAs. **B:** Transcription products labeled with [α - 32 P]ATP from the indicated RNA templates, analyzed as in Figure 1. TLS and 3' mark products initiated from the usual TLS site and from the 3'-CCA, respectively; products migrating at intermediate positions result from novel internal initiations. Note that the 3'-initiation product in lane 1 has migrated as a diffuse band due to incomplete denaturation of the highly stable hairpin. The migration of a radiolabeled tRNA^{Phe} transcript, added after terminating the RdRp reactions and used as in internal control, is indicated (RM). The molar ratios of products initiating from the 3' and novel sites relative to the internal site are indicated (average of at least two replicates); the ratio for the parental RNA, TLS-112, shown in Figure 2, is given at right.

resulted in increasing initiation from the internal site, as well as increasing initiation from novel sites apparently in the loop of the 3'-stem/loop. These initiations were tentatively mapped to loop -UCA- sequences (see arrow in Fig. 3A); another novel initiation site observed on disruption of two base pairs at the bottom of the 3'-stem/loop (Fig. 3B, lane 6) was tentatively assigned to a -CCG- loop sequence.

Addition of accessible -ACCA- sequences results in new initiation sites

The above results indicate that structured RNA immediately adjacent to a potential initiation site influences minus strand initiation. However, the range of active templates copied by TYMV RdRp *in vitro* (Singh & Dreher, 1997) and *in vivo* as chimeric infectious genomes with novel 3' ends (Skuzeski et al., 1996; Goodwin et al., 1997) makes it highly unlikely that the role of such secondary structure is sequence-dependent. Indeed, among RNAs that actively serve as templates, the only conserved sequences are the 3'-(A/C)CCA.

To test the potential for added -ACCA- sequences to serve as initiation sites, we constructed RNAs with variable numbers of -CCA- triplets added to the 3' end of the normal TLS. Initiation was observed from each additional -ACCA- site, although the original -ACCA- adjacent to the TLS was the preferred initiation site (Fig. 4, lanes 3–8). This is consistent with a role for adjacent structured RNA in initiation site specification, although freely accessible -ACCA- sequences appear to have an intrinsic ability to support some initiation. After the natural initiation site, the next favored initiation occurred from the 3'-most -ACCA (Fig. 4). Whereas some initiation was observed from each of the seven 3'-ACCA- repeats in T+6CCA RNA (Fig. 4, lane 8), initiation was not evident from within the 3' A₁₅ or C₁₅ spacers of T+15A-CCA and T+15C-CCA RNAs (Fig. 4, lanes 9 and 10). The strong initiation from the 3'-terminal -CCCA of T+15C-CCA RNA (Fig. 4, lane 10) confirms our observation with TY-TMVPSK RNA (Singh & Dreher, 1997) that 3'-terminal -CCCA can provide as strong an initiation site as -ACCA.

The influence of nucleotide identity adjacent to the -CC-residues at initiation sites was studied by substituting the A residue normally at the 3' end of the TLS with U, C, and G in derivatives of TLS+2CCA RNA (Fig. 5), which has two additional -CCA triplets added to the normal 3' end. Substitution at the internal site simultaneously tests the influence of the identity of the nucleotide on the upstream and downstream side of a -CC-potential initiation site. The spectrum of initiation at the three available sites was nearly identical when A was replaced with G (Fig. 5, lanes 3 and 6); recall that a novel initiation site was tentatively mapped to the -CCG- sequence in the loop of BOTΔS2 (Fig. 3B, lane 6). Replacement of A with U resulted in decreased

relative initiation at the internal (-CCU-) site (a 2.5-fold effect), leaving the relative initiation at the two 3'-most (-CCA) sites unchanged. Replacement of A with C also left initiation from the 3'-most sites unaffected, but initiation from the internal site became heterogeneous with a shift by one or two nucleotides toward the 3' end, although the total amount of initiation was unchanged (Fig. 5, lane 5). These results indicate a preference for a purine residue to the 3' of a -CC- initiation site, but a lack of preference on the 5' side.

Unstructured RNAs can serve as templates, but secondary structure influences initiation site selection

As an extension of studying the RNAs with long 3'-CCA- repeats used in Figure 5, we tested the ability of TYMV RdRp to transcribe minus strand from linear templates. L40 RNA has a 5'-GCCC tetranucleotide followed by 12 -CCA triplets (Fig. 6A). Initiation was observed within each of the 11 3'-most -ACCA-repeats, producing a regular three-nucleotide spacing between the longer transcription products (Fig. 6B, lane 3). Interestingly, the preferred initiation from the 3'-terminal -ACCA over some internal sites that was observed in Figure 4 was not observed for L40 RNA, which lacks anticipated secondary structure. Rather, initiation was more frequent from internal sites. The regular pattern of the longer products was absent among the multiple short products seen at the bottom of the gel; these short products do not arise from premature termination, because a similar pattern of products is seen when labeling with [α -³²P]CTP, which is incorporated exclusively at the 3' end of the RdRp product (not shown). It appears that, close to the 5' end of the template, initiation site selection is under relaxed constraints.

An experiment analogous to that of Figure 5—variation of the nucleotides adjacent to a -CC- initiation site—was performed in the context of the unstructured L40 RNA by substitution of the eighth -CCA- from the 3' end with -CCN- (Fig. 6A). As observed in Figure 5, substitution of A with G did not change the spectrum of initiation sites used (Fig. 6B, lanes 3 and 6), whereas replacement of A with U or C decreased initiation from the eighth -CC- and induced a subtle alteration in the precise initiation position (Fig. 6B, lanes 4 and 5; see arrowheads). Replacement of the eighth and seventh -CCA- repeats (counting from the 3' end) with -AAA (L40-8AAA and L40-7AAA RNAs; Fig. 6A) suppressed initiation as expected from those sites (Fig. 6B, lanes 10 and 11; see arrowheads).

The introduction of complementary sequences into L40 RNA (A:U base pairs, unrelated in sequence to the native TYMV acceptor stem) resulted in an altered spectrum of initiation sites (Fig. 6B, lanes 7–9). The introduction adjacent to the 3'-ACCA of a six-base pair stem subtending a nine-membered loop (L-40S6L9 RNA)

FIGURE 4. Multiple initiations from 3'-CCA repeats added to the end of the 83-nt-long TLS RNA. **A:** Sequences at the 3' ends of RNAs used as templates, with the relative usage of each -CCA- initiation site indicated (%). These data were derived from the gel in B, and similar results were obtained in replicate experiments. Relevant lanes in the gel are indicated at left. **B:** Transcription products labeled with [α - 32 P]ATP from the indicated RNA templates, analyzed as in Figure 1. As in Figure 3B, TLS marks product initiated from the usual TLS site and RM product from radiolabeled tRNA^{Phe} internal control. It is not known why the homopolymer-containing products in lanes 9 and 10 run faster than the longest product in lane 8, but their initiation sites are presumed to be within the 3'-most CCA.

or a six-base pair stem subtending a six-membered loop (L-40S6L6 RNA; Fig. 6A) resulted in preferred initiation from -ACCA- sequences within the loops, and in increased initiation from the 3' end (Fig. 6B, lanes 5 and 6). There were changes in the usage of initiation sites on the 5' side of the stem/loop.

To test whether a 3'-terminal -ACCA sequence is available for initiation when base paired, L-40S6L6 RNA

was modified by substituting the eighth -CCA- with -UGG- that was capable of base pairing with the 3'-CCA (L-40S9L6 RNA; Fig. 6A). The spectrum of initiation sites used on this template was much reduced, with no detectable initiation at the 3' end (Fig. 6B, lane 9); most of the short products observed in other lanes of Figure 6B were also absent in reaction products made from this RNA. These results demonstrate

FIGURE 5. Influence of nucleotide substitutions adjacent to -CC- initiation sites. **A:** Sequences at the 3' ends of RNAs derived from the 83-nt-long TLS RNA. Relative molar yield from each initiation site for a given RNA is indicated above arrowheads marking the deduced initiation sites. The reference RNA, T+2CCA (see Fig. 4, lane 4), is labeled as T/A+2CCA to emphasize the mutated nucleotide. Relevant lanes in the gel are indicated at left. **B:** Transcription products labeled with [α - 32 P]ATP from the indicated RNA templates, analyzed as in Figure 1.

that nonspecific secondary structure can strongly influence TYMV RdRp in directing initiation to a specific site that is accessible and non-base paired.

DISCUSSION

Absence of specific promoter elements outside the -CCR- initiation box

Previous studies showing that truncated 3' fragments of the TYMV genome can be transcribed efficiently by TYMV RdRp *in vitro* using the normal initiation site suggested that there are no strong promoter elements upstream of the pseudoknotted acceptor stem (Deiman et al., 1997; Singh & Dreher, 1997). We have now verified this conclusion by showing with deletion and substitution mutations that the three domains of the

TLS upstream of the acceptor stem—the D-, anticodon, and T-arms—contain no detectable specific information influencing minus strand initiation (Fig. 1). Various mutations in these domains have affected neither the efficiency of transcription nor the specificity of initiation at the 3' terminus. The ability of the 27-nt-long acceptor arm to act independently in directing complementary strand initiation was again demonstrated by the constructs shown in Figure 2, which support two independent initiation sites, each associated with an acceptor arm domain.

Taken together with the observed amplification *in vivo* (Skuzeski et al., 1996; Goodwin et al., 1997) and transcription *in vitro* (Singh & Dreher, 1997) of RNAs with various acceptor arm sequences, the above results indicated an absence of sequence-specific promoter requirements upstream of the 3'-terminal three or four nucleotides. We therefore explored the influence on transcription of RNA structure immediately upstream of -NCCA initiation sites, using RNAs in which the efficiency of initiation from the normal site can be compared with initiation from an additional 3' element (Figs. 2, 3). We have observed that different secondary structures can support 3'-initiation to differing degrees: 3' initiation varies 13-fold for the RNAs used as templates in Figure 3. Simple stem/loops or structures that may well have internal or bulge loops within a stem can substitute for the normal pseudoknotted structure present in the TYMV acceptor stem. However, the non-pseudoknotted RNAs we have used as templates are not as effective as the wild-type acceptor stem on two counts: first, the pseudoknotted additional domain (with an A₃ spacer) supported some threefold higher initiation from the 3' end relative to the internal site than the best-performing stem/loop RNA (Fig. 2, lane 2 versus Fig. 3B, lane 4); second, whereas the pseudoknotted acceptor stem supported specific initiation from the 3'-end, some weakly base paired stem/loop structures supported initiation from novel sites (Fig. 3B, lanes 6, 8, 9, and 10). These results are consistent with a two-fold decrease in transcription observed previously after disruption of the acceptor stem pseudoknot with point or multiple substitutions (Deiman et al., 1997; Singh & Dreher, 1997). We conclude that various conformations of structured RNA, including pseudoknots, can effectively direct initiation to -NCCA initiation sites.

With specific promoter elements absent from the acceptor stem and upstream parts of the TYMV TLS, the only remaining site for specificity is the -ACCA initiation region. We have previously shown that deletion of the 3'-CCA terminal triplet completely abrogates transcription from the 83-nt-long TYMV TLS (Singh & Dreher, 1997). We have now assessed the role of the nucleotides 5'- and 3'-adjacent to a -CC- dinucleotide supporting initiation (Fig. 5). There seems to be no effect of substituting the A residue on the upstream side, as evidenced by a constant ratio of initiation at the pen-

FIGURE 6. Linear RNAs as templates and the influence of structure. **A:** Sequences of RNAs used as templates are shown, with sequence changes relative to L40 RNA highlighted. Expected base pairing is indicated for RNAs used in lanes 7, 8, and 9. The CCA repeats are numbered at the top of the figure, starting from the 3' end. **B:** Transcription products labeled with [α - 32 P]UTP from the indicated RNA templates, analyzed as in Figure 1, but by 12% sequencing PAGE. TY-41 (lane 2) is a 41-nt-long fragment from the 3' end of TYMV RNA (Singh & Dreher, 1997). Positions of mutations are marked with arrowheads in lanes 4, 5, 6, 10, and 11, and nucleotides expected to be base paired are marked with brackets in lanes 7–9. Lane 1 contained a labeled 10-bp DNA ladder.

ultimate versus 3' -CCA in lanes 3–6 of Figure 5. This result is consistent with the strong terminal initiation at the 3'-CCCA at the end of a C₁₅ spacer added to the TYMV TLS (Fig. 4, lane 10) and the *in vivo* amplification of chimeric TYMV RNAs with 3'-CCCA termini (Goodwin et al., 1997). On the downstream side, adenylate and guanylate appear to be interchangeable, whereas a transversion results in about twofold lower initiation (substitution with U) or the introduction of initiation site microheterogeneity (substitution with C) (Fig. 5, lanes 4 and 5). Long C-tracts do not support initiation at internal C residues (Fig. 4, lane 10). Novel initiation sites observed with RNAs possessing modified 3'-secondary structure (Fig. 3B, lanes 6, 8, 9, and 10) were tentatively assigned to -UCA- and -CCG- sites. In summary, it appears that TYMV RdRp has at least some specific sequence requirements at the initiation site. Although a systematic mutational analysis of this "initiation box" remains to be done, the specific target for initiation appears to be very small, and at present can be described as -CCR (R = purine).

A concurrent study (Deiman et al., 1998) has reached similar conclusions regarding the ability of TYMV RdRp to initiate complementary strand synthesis *in vitro*. Short stem/loop RNAs with 3'-CCA termini were efficient templates despite the absence of upstream TLS features, and substitution of the A residue upstream of the 3'-CCA with U or C had little effect on product yield.

How can specific initiation on the 6,318-nt long TYMV RNA be achieved by the TYMV RdRp?

The same preparations of TYMV RdRp used in this paper are able to transcribe full-length TYMV RNA to yield a discrete full-length product (Fig. 3B, lane 7 in Singh & Dreher, 1997). However, the sequence -CCR- appears 415 times in the 6,318-nt long virion RNA. The observed specific transcription into full-length minus strands is a remarkable achievement for an enzyme that responds to such a small initiation box and no specific RNA structure.

Experiments with small RNA templates unrelated to the normal viral templates have given a view into how specificity can arise. A short RNA, L40, designed to be devoid of secondary structure and that contains arrays of -CCA- repeats, can support initiation from virtually all -CC- dinucleotides (Fig. 6B). Interestingly, Deiman et al. (1998) recently have also observed initiation events from an unstructured short template derived from an internal region of the TYMV genome rich in -CC- dinucleotides. Mutation of the nucleotides immediately adjacent to a -CC- couple in L40 RNA affects initiation (Fig. 6B, lanes 3–6) in a manner consistent with that discussed above (see Fig. 5), but many initiation sites are observed on the one RNA. By introducing base pairing unrelated in sequence to that of the TYMV TLS into the unstructured L40 RNA, the spectrum of initia-

tion sites was strongly altered and simplified (Fig. 6B, lanes 7–9): initiation was not only suppressed or lost from mutated sites that no longer had -CCR- initiation boxes, but initiation was strongly favored at one of the two CCA boxes in the internal loop, whereas the spectrum of short products (seen in Fig. 6B, lanes 3–6) was greatly suppressed. Significantly, inclusion of the 3'-terminal -CCA within the base pairing of the stem entirely suppressed initiation at this site (Fig. 6B, lane 9).

Our experiments have identified two ways in which nonspecific RNA structure can influence specificity. First, potential initiation sites that are base paired are unavailable to the polymerase. Second, RNA structure has the ability to preferentially direct the RdRp to initiate at certain potential initiation sites among those not directly involved in base pairing (Fig. 6B, lanes 7–9). The mechanism of this second influence is not understood, but our experiments suggest a twofold influence: on the one hand, structure may have a negative influence suppressing initiation at sites that are sterically inaccessible to the polymerase active site; on the other hand, the TYMV RdRp may have some preference for binding structured or condensed RNA, favoring initiation sites adjacent to structured RNA, as observed in Figure 5. If a selection criterion in the evolution of the TYMV genome is the avoidance of -CCR- sequences in a placement that is sterically accessible to the TYMV RdRp, it is conceivable that a genomic RNA with a specific 3' site for minus strand synthesis can be achieved. A useful adaptation would presumably also be to place the appropriate initiation site in an exceptionally accessible site; this has certainly been done in the case of TYMV RNA, as demonstrated by the efficient interaction of the virion RNA with valyl-tRNA synthetase and with elongation factor EF-1 α (Mans et al., 1991; Dreher & Goodwin, *forthcoming*), both of which interact with the 3' terminus.

The selection of alternate codons for a given amino acid is an obvious source of evolutionary diversity to satisfy the above criteria. Two features of the TYMV genome present a challenge to even this opportunity for selection, however. The genome is exceptionally C-rich (39.1% C) and a significant proportion of the genome (36.1%) simultaneously codes in two reading frames (Bransom et al., 1995). The specificity mechanism we have demonstrated here may thus need to be augmented by some additional mechanism. Note, however, that the most obvious challenge to the specificity of TYMV RdRp in the cell—the avoidance of being titrated away by the abundant tRNAs, all of which have an accessible 3'-CCA and can serve as *in vitro* templates—is probably not the problem it initially appears: hypermodified, mature tRNAs are copied less actively than unmodified transcripts (Singh & Dreher, 1997); tRNAs probably are associated predominantly with translational elongation factor or other components of the translational apparatus; the translational

components including tRNA are organized in a closed, metabolically channelled system (Negruksii et al., 1994); the replication of TYMV RNAs is thought to occur in vesicles separated from ribosomes (and presumably tRNAs: Matthews, 1991); and TYMV replication is characterized by a strong preference for the replication proteins to amplify the same RNA molecule from which they were translated (Weiland & Dreher, 1993).

Insight into the mechanism of TYMV RdRp initiation site selection

From the strong sensitivity of initiation site selection to RNA structure, it follows that the TYMV RdRp does not unfold a potential template prior to initiation, but rather scans the folded RNA for an accessible initiation box. Thus, although a helicase-like protein is an essential protein encoded by TYMV (Weiland & Dreher, 1993) and may well be part of the viral RdRp complex (Buck, 1996), no helicase activity precedes strand initiation. In finding a potential initiation site, TYMV RdRp does not bind and then scan the linear sequence of the RNA, as for instance, occurs for a eukaryotic 40S ribosomal subunit seeking an initiation codon on an mRNA. There is no apparent preference per se for initiation at the 3' end, as evidenced by several different experiments in Figures 2, 3, 4, 5, and 6. Nor is there an obligatory preference for initiation immediately downstream of an element of RNA structure, as shown by the strong internal loop initiation in lanes 8 and 9 of Figure 6. Rather, it appears that TYMV RdRp scans at random the surface of a folded RNA for accessible -CCR- potential initiation sites that can fit into the enzyme's active site without steric constraints. There appears to be no required interaction with folded RNA, because unstructured RNAs support high levels of initiation (Fig. 6) that are not out-competed by initiation on folded TYMV TLS competitor RNA (not shown).

Applicability of this specificity mechanism to other protein-RNA interactions

We believe that the specificity mechanism we have described may be used rather commonly in the replication of positive strand RNA viruses. It fits observations made in mapping the genomic promoters recognized by turnip crinkle virus (TCV) RdRp in both the positive and negative sense strands of TCV-associated satellite RNA C (Song & Simon, 1995; Guan et al., 1997), and may also apply to subgenomic RNA synthesis from this satellite RNA (Wang & Simon, 1997). It may apply to the RNAs replicated by Q β replicase: the well-known requirement for secondary structure might not only be in order to ensure product release from the template strands (Arora et al., 1996), but perhaps also an expression of the specificity mechanism we have described. Interestingly, although potential pro-

moter elements have been described on Q β RNA, Barrera et al. (1993) have suggested that "template selectivity may be provided by the tertiary structure and topology of the RNA, and in addition by the C-rich 3'-terminal sequence." Further, the observation that poliovirus and human rhinovirus genomes entirely lacking the 3'-noncoding region upstream of the poly(A) tail (deletions of 68 and 44 nt, respectively) were capable of replication (albeit debilitated) led to the conclusion that the basic mechanism of picornavirus replication initiation may not be strictly template specific (Todd et al., 1997). The observed replication of the truncated genomes might alternatively be explained in terms of a specificity mechanism like that used by TYMV RdRp. Finally, the template preferences we have observed for TYMV RdRp are remarkably similar to the template requirements of the Mauriceville retroplasmid reverse transcriptase (Chen & Lambowitz, 1997), an enzyme that has been considered to be evolutionarily intermediate between TYMV-like RdRp's and retroviral-like reverse transcriptases (Maizels & Weiner, 1994).

These observations suggest that it will be useful to consider the specificity principle we have described here in future searches for *cis*-active elements controlling RNA transcription in positive strand viruses. We believe that this mechanism may also be found to operate in other RNA-protein interactions, such as transcript binding by the bacterial rho termination protein, for which RNA recognition criteria have proven elusive (Platt, 1994), or in the selection of splice sites in precursor mRNAs (Goguel & Rosbash, 1993; Black, 1995).

MATERIALS AND METHODS

Preparation of template RNAs

Template RNAs were transcribed with T7 RNA polymerase from PCR-amplified DNAs as described (Singh & Dreher, 1997). RNA variants of the TYMV tRNA-like structure were PCR-amplified (30 cycles) with *Taq* DNA polymerase from plasmid DNA templates using terminal deoxyoligonucleotides ranging from 24 to 66 residues in length. Shorter RNAs not based on TYMV sequences (those used in Fig. 6) were subjected to 30 PCR cycles that served to fill out the overhangs of two deoxyoligonucleotides annealed at their 3'-ends (Singh & Dreher, 1997). The list of oligos used in this study is available on request.

PCR products were purified by PAGE prior to transcription, and T7 transcripts were also purified by 7 M urea-PAGE (6 or 8%) and recovered by elution. RNA concentrations were estimated from ethidium bromide-stained gels.

RdRp transcription assays and analysis of products

TYMV RdRp was solubilized from membranes and collected from a glycerol gradient and dialyzed as described (Singh & Dreher, 1997). Endogenous RNAs were removed by treat-

ment with 10 $\mu\text{g/mL}$ micrococcal nuclease (Sigma) in the presence of 1 mM CaCl_2 at 30°C for 30 min. The reaction was stopped by adding EGTA to 10 mM.

The standard 25- μL reactions contained 20 μL of the RdRp preparation in the presence of 10 mM DTT, 10 mM MgCl_2 , 200 $\mu\text{g/mL}$ actinomycin D, 0.5 mM of each unlabeled NTP, 1 μM of labeled nucleotide $\{[\alpha\text{-}^{32}\text{P}]\text{UTP}$ or $[\alpha\text{-}^{32}\text{P}]\text{ATP}$ as indicated (400 Ci/mmol; NEN)}, and about 7 pmol template. Reactions were incubated at 30°C for 1 h, and stopped by adding 25 μL of TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) containing 2% SDS and 50 mM EDTA followed by two phenol/chloroform extractions. In some experiments (e.g., those shown in Figs. 3, 4), recovery was monitored by adding a labeled marker RNA as internal control to the reaction mixture immediately after deproteinization. Nucleic acids were precipitated with 10 μg glycogen as carrier in ammonium acetate/ethanol, washed in 70% ethanol, redissolved in 10 μL of water, and finally analyzed by 7 M urea/PAGE (8 or 12%). Template quality and recovery were monitored in all gels by ethidium bromide staining, and RdRp products were detected and quantitated using a phosphorimager (Molecular Dynamics). Reaction products were also treated with a mixture of ribonucleases T1 and A in high salt (0.3 M NaCl; Singh & Dreher, 1997) to remove the excess of template RNAs prior to gel analysis.

ACKNOWLEDGMENTS

We thank Drs. Alan Lambowitz and Dale Mosbaugh for critical reading of the manuscript, and the Central Services Facility of the Oregon State University Center for Gene Research and Biotechnology for synthesizing deoxyoligonucleotides. The work was supported by grants from NIH (AI-33907 and GM-54610). This is technical report no. 11316 of the Oregon Agricultural Experiment Station.

Received May 1, 1998; returned for revision June 12, 1998; revised manuscript received June 19, 1998

REFERENCES

- Arora R, Priano C, Jacobsen AB, Mills DR. 1996. cis-Acting elements within an RNA coliphage genome: Fold as you please, but fold you must!! *J Mol Biol* 258:433–446.
- Barrera I, Schuppli D, Sogo JM, Weber H. 1993. Different mechanisms of recognition of bacteriophage Q β plus and minus strand RNAs by Q β replicase. *J Mol Biol* 232:512–521.
- Black DL. 1995. Finding splice sites within a wilderness of RNA. *RNA* 1:763–771.
- Bransom KL, Weiland JJ, Tsai CH, Dreher TW. 1995. Coding density of the turnip yellow mosaic virus genome: Roles of the overlapping coat protein and p206-readthrough coding regions. *Virology* 206:403–412.
- Brown D, Brown J, Kang C, Gold L, Allen P. 1997. Single-stranded RNA recognition by the bacteriophage T4 translational repressor, regA. *J Biol Chem* 272:14969–14974.
- Buck KW. 1996. Comparison of the replication of positive-stranded RNA viruses of plants and animals. *Adv Virus Res* 47:159–251.
- Chen B, Lambowitz AM. 1997. De novo and DNA primer-mediated initiation of cDNA synthesis by the Mauriceville retroplasmid reverse transcriptase involve recognition of a 3' CCA sequence. *J Mol Biol* 271:311–332.
- Cui T, Porter AG. 1995. Localization of binding site for encephalomyocarditis virus RNA polymerase in the 3'-noncoding region of the viral RNA. *Nucleic Acids Res* 23:377–382.
- Deiman BALM, Kortlever RM, Pleij CWA. 1997. The role of the pseudoknot at the 3' end of turnip yellow mosaic virus RNA in minus-strand synthesis by the viral RNA-dependent RNA polymerase. *J Virol* 71:5990–5996.
- Deiman BALM, Koenen AK, Verlaan PWG, Pleij CWA. 1998. Minimal template requirements for initiation of minus-strand synthesis in vitro by the RNA-dependent RNA polymerase of turnip yellow mosaic virus. *J Virol* 72:3965–3972.
- Dreher TW, Hall TC. 1988. Mutational analysis of the sequence and structural requirements in brome mosaic virus RNA for minus strand promoter activity. *J Mol Biol* 201:31–40.
- Dreher TW, Goodwin JB. 1998. Transfer RNA mimicry among tymoviral genomic RNAs ranges from highly efficient to vestigial. Submitted.
- Goguel V, Rosbash M. 1993. Splice site choice and slicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. *Cell* 72:893–901.
- Goodwin JB, Skuzeski JM, Dreher TW. 1997. Characterization of chimeric turnip yellow mosaic virus genomes that are infectious in the absence of aminoacylation. *Virology* 230:113–124.
- Guan H, Song C, Simon AE. 1997. RNA promoters located ion (–)-strands of a subviral RNA associated with turnip crinkle virus. *RNA* 3:1401–1412.
- Levis R, Schlesinger S, Huang HV. 1990. Promoter for Sindbis virus RNA-dependent subgenomic RNA transcription. *J Virol* 64:1726–1733.
- Maizels N, Weiner AM. 1994. Phylogeny from function: Evidence from the molecular fossil record that tRNA originated in replication, not translation. *Proc Natl Acad Sci USA* 91:6729–6734.
- Mans RM, Pleij CW, Bosch L. 1991. tRNA-like structures: Structure, function and evolutionary significance. *Eur J Biochem* 201:303–324.
- Matthews REF. 1991. *Plant virology, third edition*. San Diego, California: Academic Press. pp 232–237.
- Miller WA, Bujarski JJ, Dreher TW, Hall TC. 1986. Minus-strand initiation by brome mosaic virus replicase within the 3' tRNA-like structure of native and modified RNA templates. *J Mol Biol* 187:537–546.
- Miller WA, Dreher TW, Hall TC. 1985. Synthesis of brome mosaic virus subgenomic RNA in vitro by internal initiation on (–) sense genomic RNA. *Nature (Lond)* 313:68–70.
- Miranda G, Schuppli D, Barrera I, Hausherr C, Sogo JM, Weber H. 1997. Recognition of bacteriophage Q β plus strand RNA as a template by Q β replicase: Role of RNA interactions mediated by ribosomal proteins S1 and host factor. *J Mol Biol* 267:1089–1103.
- Negrutskii BS, Stapulionis R, Deutscher MP. 1994. Supramolecular organization of the mammalian translation system. *Proc Natl Acad Sci USA* 91:964–968.
- Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. 1994. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* 372:432–438.
- Pallanck L, Pak M, Schulman LH. 1995. tRNA discrimination in aminoacylation. In: Söll D, RajBhandary UL, eds. *tRNA: Structure, biosynthesis and function*. Washington, DC: ASM Press. pp 371–394.
- Platt T. 1994. Rho and RNA: Models for recognition and response. *Mol Microbiol* 11:983–990.
- Puglisi JD, Tan R, Calnan BJ, Frankel AD, Williamson JR. 1992. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* 257:76–80.
- Rietveld K, Pleij CWA, Bosch L. 1983. Three-dimensional models of the tRNA-like 3' termini of some plant viral RNAs. *EMBO J* 2:1079–1085.
- Siegel RW, Adkins S, Kao CC. 1997. Sequence-specific recognition of a subgenomic RNA promoter by a viral RNA polymerase. *Proc Natl Acad Sci USA* 94:11238–11243.
- Singh R, Dreher TW. 1997. Turnip yellow mosaic virus RNA-dependent RNA polymerase: Initiation of negative strand synthesis in vitro. *Virology* 233:430–439.
- Skuzeski JM, Bozarth CS, Dreher TW. 1996. The turnip yellow mosaic virus tRNA-like structure cannot be replaced by generic tRNA-like elements or by heterologous 3' untranslated regions known to enhance mRNA expression and stability. *J Virol* 70:2107–2115.

- Song C, Simon AE. 1995. Requirement of a 3'-terminal stem-loop in in vitro transcription by an RNA-dependent RNA polymerase. *J Mol Biol* 254:6-14.
- Szewczak AA, Webster KR, Spicer EK, Moore PB. 1991. An NMR characterization of the regA protein-binding site of bacteriophage T4 gene 44 mRNA. *J Biol Chem* 266:17832-17837.
- Todd S, Towner JS, Brown DM, Semler BL. 1997. Replication-competent picornaviruses with complete genomic RNA 3' non-coding region deletions. *J Virol* 71:8868-8874.
- Valegård K, Murray JB, Stockley PG, Stonehouse NJ, Liljas L. 1994. Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* 371:623-626.
- Wang J, Simon AE. 1997. Analysis of the two subgenomic promoters for turnip crinkle virus in vivo and in vitro. *Virology* 232:174-186.
- Weiland JJ, Dreher TW. 1993. Cis-preferential replication of the turnip yellow mosaic virus RNA genome. *Proc Natl Acad Sci USA* 90:6095-6099.