Autonomous Role of 3'-Terminal CCCA in Directing Transcription of RNAs by Q β Replicase†

DAVID M. TRETHEWAY,¹ SHIGEO YOSHINARI,¹ AND THEO W. DREHER^{1,2*}

*Department of Microbiology¹ and Center for Gene Research and Biotechnology,²
Oregon State University, Corvallis, Oregon 97331-3804*

Received 15 June 2001/Accepted 23 August 2001

We have studied transcription in vitro by Q β replicase to deduce the minimal features needed for efficient end-to-end copying of an RNA template. Our studies have used templates ca. 30 nucleotides long that are expected to be free of secondary structure, permitting unambiguous analysis of the role of template sequence in directing transcription. A 3'-terminal CCCA (3'-CCCA) directs transcriptional initiation to opposite the underlined C; the amount of transcription is comparable between RNAs possessing upstream (CCA)_n tracts, A-rich sequences, or a highly folded domain and is also comparable in single-round transcription assays to transcription of two amplifiable RNAs. Predominant initiation occurs within the 3'-CCCA initiation box when a wide variety of sequences is present immediately upstream, but CCA or a closely similar sequence in that position results in significant internal initiation. Removal of the 3'-A from the 3'-CCCA results in 5- to 10-fold-lower transcription, emphasizing the importance of the nontemplated addition of 3'-A by Q β replicase during termination. In considering whether 3'-CCCA could provide sufficient specificity for viral transcription, and consequently amplification, in vivo, we note that tRNA^{His} is the only stable *Escherichia coli* RNA with 3'-CCCA. In vitro-generated transcripts corresponding to tRNA^{His} served as poor templates for Q β replicase; this was shown to be due to the inaccessibility of the partially base-paired CCCA. These studies demonstrate that 3'-CCCA plays a major role in the control of transcription by Q β replicase and that the abundant RNAs present in the host cell should not be efficient templates.

Genome replication among the positive-strand RNA viruses is accomplished by sequential end-to-end transcriptions, first of the encapsidated positive sense RNA and subsequently of the newly synthesized negative-sense antigenome. Except for viruses whose genomes are covalently linked at the 5' end to a specialized protein, these transcriptions occur by de novo initiation (9). For successful replication by this pathway, transcriptional initiation must occur predominantly or exclusively at the 3' ends of genome and antigenome RNAs, with minimal initiation occurring internally or on other RNAs present within the cell. Q β replicase provides a convenient means to study the template properties underlying these required specificities for a representative positive-strand RNA virus.

Q β replicase is the 4-subunit RNA-dependent RNA polymerase (RdRp) enzyme complex that amplifies the 4.2 kb positive-strand genome of bacteriophage Q β , a phage infecting *Escherichia coli* (6, 34). Unlike the RdRp of any eukaryotic positive-strand RNA virus, Q β replicase has been purified to homogeneity and shown to be capable of supporting the full viral genome amplification cycle in vitro (6). This enzyme catalyzes de novo strand initiation with GTP opposite a short cluster of C residues in the CCCA 3' termini that are a feature of both positive and negative strands of almost all amplifiable templates described in the literature (24, 26, 40). After full-length transcription, termination is accompanied by the addi-

tion of a nontemplated A residue (6), thereby restoring the 3' A that was not copied into the complementary strand.

While internal sequences whose removal decreases the transcription of Q β positive- and negative-sense genomic RNAs have been mapped (27, 29), the precise features required for directing the transcription of the genomic RNAs are unclear. Further, these elements are not universally present in the wide variety of short RNAs amplifiable by Q β replicase (40). Indeed, it was recognized several years ago (26) that the only feature common to replicatable RNAs appears to be the CCCA 3' terminus; this observation has held true with two exceptions, one being a variant Q β positive-sense RNA with a UCCA 3' end (28), the other being a 6S RNA amplified by Q β replicase with a GCCA terminus (33). Nevertheless, it has not been demonstrated experimentally whether the presence of a CCCA-terminal sequence is all that is required for an RNA to be transcribed by Q β replicase. Our recent studies have shown that Q β replicase can direct initiation from every C₂₋₄A repeat present in short linear RNAs comprised of multiple C₂₋₄A repeats (38), a property shared by the RdRps from turnip yellow mosaic and turnip crinkle viruses (39). While these studies suggested that a C₂₋₄A element could act as an independent initiation site, they did not resolve whether overall transcription in these RNAs was supported by the reiterated C-rich sequence motifs; Q β replicase is well known for its ability to transcribe poly(C) (6). Further, RNAs such as (C₂₋₄A)_n are not practical templates for amplification because of the large amounts of internal initiation.

We set out in the present studies to investigate the ability of a C₂₋₄A sequence to act independently to direct transcriptional initiation and to determine what sequences are necessary to ensure that initiation occurs at the 3'-most C residue and not

* Corresponding author. Mailing address: Department of Microbiology, 220 Nash Hall, Oregon State University, Corvallis, OR 97331-3804. Phone: (541) 737-1795. Fax: (541) 737-0496. E-mail: theo.dreher@orst.edu.

† Technical report 11771 of the Oregon Agricultural Experiment Station.

FIG. 1. Inactivation of the 3'-penultimate CCA initiation box does not increase initiation from the 3'-CCA initiation box. RNA variants derived from CCA9 RNA, bearing the mutations of the C residues in initiation box #8 shown in panel A, were tested as templates with Q β replicase. RNAs were incubated in the presence of Q β replicase and [α -³²P]CTP for 10 min at 37°C as described in Materials and Methods, except that MgCl₂ was present at 21 mM. Products made from templates identified by their initiation box #8 sequence were separated by 12.5% denaturing PAGE (shown in panel B). The numbers at the left identify the CCA repeat (initiation box numbers) from which product strands originate, #9 representing the 3'-most CCA. Each CCA initiation box produces three products, marked a, b, and c at the left, whose origins are explained in the text.

at an internal site, thus ensuring the complete end-to-end copying that is needed for sequence maintenance during replication. The results presented here provide a prescription for the minimal 3'-sequence requirement needed by Q β replicase for efficient 3'-end initiation on a template. This knowledge provides an explanation for the range of 3' sequences observed for natural and model RNAs amplifiable by Q β replicase and insight into the mechanism that ensures the propagation of full-length viral RNA with minimal interference in this process by other RNAs present in the cell. Although there is evidence that internal "promoter" elements specifically recognized by the replicase function in positive-strand RNA viral systems (see, for example, reference 12), including Q β (the M-site on Q β positive-sense RNA [27]), such elements are not absolutely required to ensure transcription by Q β replicase. In this system, we show that a small number of nucleotides at the 3' terminus of an RNA exert a crucial influence over the fate of that RNA as a template.

MATERIALS AND METHODS

Materials. Q β replicase, a generous gift from Michael Farrell (Vysis, Inc.), was prepared by overexpression in *E. coli* (20). Synthetic DNA oligomers used for the preparation of template RNAs were synthesized by Life Technologies, Inc., or at the Central Services Laboratory of the Center for Gene Research and Biotechnology, Oregon State University.

Preparation of RNA templates for Q β replicase assays. Template RNAs were prepared enzymatically with T7 RNA polymerase from DNA templates comprising annealed DNA oligomers as described previously (38) or from templates generated by PCR as described previously (30) in the case of *E. coli* tRNA^{His} and its derivatives. All of the DNA templates, other than those used to make CCCA9 and CCCCA7 RNAs (see Fig. 3) and GGA(CCA)₇X RNAs (see Fig. 5B), contained modified 2'-O-methyl uridine and/or 2'-O-methyl guanosine at the two 5'-most nucleotide positions to suppress the formation of *n* + 1 transcripts (17).

Transcripts were purified by 7 M urea–10% polyacrylamide gel electrophoresis (PAGE) with single-base resolution. Concentrations of RNA solutions were determined by spectrophotometry by using extinction coefficients calculated by Oligo 6.0 software.

Q β replicase assay and analysis of products. Typical Q β replicase reactions (25 μ l) contained 2.5 pmol (100 nM) of template RNA and 1.25 pmol (50 nM) of Q β replicase in 80 mM Tris-HCl (pH 7.5); 10 mM MgCl₂; 1 mM dithiothreitol; 200 μ M concentrations each of ATP, GTP, and UTP; and 50 μ M CTP, including 10 μ Ci of [α -³²P]CTP. The experiments of Fig. 1 were conducted in the presence of 21 mM MgCl₂ as in previous studies (38); we have adopted 10 mM MgCl₂ as our standard reaction condition to make our experiments more representative of conditions used by other researchers. Incubation was for 10 min at 37°C, and products were recovered by phenol extraction and ethanol precipitation. At the time of ethanol precipitation in the case of CCA9 RNA and its GGA(CCA)₇NNACCA derivatives, 250 pmol of DNA oligomer complementary to the T7 transcript of CCA12 RNA was added; this DNA promotes displacement of the template RNAs from the Q β replicase products during sample preparation for electrophoresis. In other cases, 250 pmol of DNA exactly complementary to the RNA template was added as described above, except for the RNAs analyzed in Fig. 6, for which no antisense nucleic acids were used.

Dried pellets recovered after ethanol precipitation were dissolved in 50 μ l of 90% formamide–10 mM EDTA–0.02% dyes, boiled for 5 min, and ice-chilled, and then 10 μ l was subjected to 7 M urea–PAGE. After electrophoresis, gels were fixed and dried, and radioactivity was detected and analyzed with a PhosphorImager with ImageQuant software (Molecular Dynamics).

Single-round transcription assays. Assays were performed as described above, except that CTP was initially withheld. After incubation for 1 min at 37°C, polyethylene sulfonate (PES) was added to 5 μ g/ml, followed by 50 μ M CTP, including 10 μ Ci of [α -³²P]CTP. The assay was completed by incubation at 37°C for 9 min, followed by product analysis as described above.

RESULTS

A CCCA 3' terminus is superior to CCA at directing initiation predominantly from the 3' end. Our previous experiments have shown that (CCA)_{*n*} RNAs of ca. 30 nucleotides

(nt) or longer are good templates for Q β replicase (38), supporting a cumulative level of RNA strand initiation similar to that observed with an amplifiable positive control RNA, DN3 (39). However, most initiation on such RNAs occurs upstream of the 3' end (38) (see also Fig. 1, lanes marked CCA9), resulting in sequence loss that is incompatible with the requirements of viral RNA replication. More precisely, predominant initiation occurs opposite the downstream C of the 3'-penultimate CCA repeat (CCACCA, in box #8; 59% of the initiation products were longer than 12 nt), with weaker initiation opposite the 3'-most C (CCACCA, in box #9; 17% of initiations; Fig. 1, lanes 2, 8, 14, and 21). Note that the band heterogeneity present as bands 8a, b, and c in the above lanes is principally due to differences at the 3' end of the product, a finding we have reported previously (38) and reconfirmed in this study (not shown). Whereas the longer CCA12 RNA (38) produces mainly doublets, CCA9 RNA yields the triplets labeled a, b, and c. Our studies indicate that (i) the "a" band represents transcription to the 5' end of the template, followed by addition of a nontemplated A (25, 36) (CCA 3' end), (ii) the "b" band represents a mixture of full-length transcripts that fail to acquire an additional A (CC 3' end) and those that terminate 1 nt early and acquire a nontemplated A (CA 3' end), while (iii) the "c" band represents termination 1 nt early with no adenylation (C 3' end). We have shown that this pattern of products holds true with initiation from 3'-CCCA by labeling RNAs AAA71, AAA51, and AAA31 (see Fig. 4) with [α - 32 P]CTP and inspecting RNase T $_1$ -released fragments (results not shown).

To test whether the predominant initiation at CCA box #8 in CCA9 RNA results from the suppression of initiation from the 3'-terminal CCA box #9, we studied the spectrum of transcription products in a family of 15 templates in which the C residues of box #8 were mutated with single or double substitutions to NNA. As shown in Fig. 1, all single or double mutations in CCA9 RNA resulted in the almost complete loss of initiation from box #8, with little or no other changes in the spectrum of initiations from other sites. Most significantly, mutation of box #8 did not result in increased and preferential initiation from the 3' end (box #9). The observations were uniform, regardless of whether the substitution of C was with the pyrimidine U or the purines A or G, at position -5, position -6, or both. These results indicate that the absence of predominant initiation from the 3' end in CCA9 RNA is mostly due to the absence of an appropriately strong initiation signal and not due to competition by an upstream CCA.

Thus, although the sequence CCA clearly does constitute a functional initiation box, the results of Fig. 1 indicate it is relatively weak in directing initiation from the 3' terminus of an RNA. Indeed, RNAs reported in the literature as being amplifiable by Q β replicase typically terminate in NCCCA (26, 34, 40), suggesting that a 3'-CCCA rather than 3'-CCA is necessary to ensure preferential initiation from the 3' terminus. To test this idea, we studied transcription from a second family of templates based on CCA9 RNA, in which the CCACCA terminus was replaced with NNNCCCA. Of the 14 3'-heptanucleotide sequences tested, 11 were derived from the published sequences of amplifiable RNAs (Fig. 2A). As is evident from Fig. 2B, a CCCA terminus ensures predominant initiation from the 3' end in the presence of a variety of tri-

nucleotide sequences immediately upstream (mutations in initiation box #8). A total of 59 to 87% of all products longer than 12 nt initiated from within the CCCA initiation box for all templates tested, except for UGGCCCA RNA (Fig. 2B, lane 12), which was a poor template overall (accessibility probing with RNases suggested that this RNA exists in solution with some undetermined secondary structure, which presumably interferes with template activity). In contrast, 59% of the products longer than 12 nt transcribed from CCA9 RNA originated from the 3'-penultimate CCA initiation box (#8), with only 20% of the initiations occurring from the 3'-CCA (box #9). No change in product spectrum was observed when representative RNAs from Fig. 1 and 2 were analyzed under single-round transcription conditions in the presence of 5 μ g of PES/ml (data not shown). There is therefore no indication of preferential reinitiation at certain sites.

Similar levels of initiation were observed when the CCCA 3'-terminal initiation box was adjacent to various sequences (Fig. 2, mutated box #8), indicating that these adjacent sequences have little influence on transcription by Q β replicase. The level of initiation from CCCA was similar to that from CCA box #8 of CCA9 RNA.

Relative initiation strength from 3'-most and 3'-penultimate initiation boxes is influenced by the number of cytosine residues. The results presented above show that a CCCA initiation sequence is superior to CCA at the 3' end, although CCA provides strong initiation from a position slightly internal to the 3' end. To explore the relationship between initiation from these two positions, we analyzed transcription from RNAs with different numbers of cytosines in either the 3' or 3'-penultimate initiation boxes. The addition of increasing numbers of C residues to box #9 of CCA9 RNA dramatically switches the initiation preference from box #8 to the 3' end (Fig. 3B, lanes 1 to 5). Thus, the ratio of products initiating from box #9 relative to box #8 increases from 0.29:1 for CCA9 to 1.1, 4.7, and 14:1 when the 3' box (#9) has the sequences C3A, C4A, and C5A, respectively (Fig. 3B).

In the absence of a functional 3'-penultimate initiation box (Fig. 3B, lanes 8 to 11), predominant end-to-end transcription is driven by 3' initiation boxes with three, four, or five C residues. In these cases, 80 to 90% of products longer than 12 nt originate from the 3' end (Fig. 3B). The level of transcription was some 30 to 60% higher from 3'-C4A or 3'-C5A than from 3'-C3A, indicating that C strings longer than three can provide modest additional initiation strength. Initiation from 3'-CCA is at about the same low level in the presence or absence of CCA at box #8 (Fig. 3B, lanes 7 versus 8), indicating that initiation boxes #8 and #9 act largely independently rather than in competition (compare also transcription from box #9 of Fig. 3, lanes 3 to 5 versus lanes 9 to 11). The decreasing level of transcription from box #8 in lanes 2 to 5 of Fig. 3B is attributable to the increasing distance from the 3' end as C residues are added to box #9. Thus, box #8 of CCAC5A RNA (lane 5) is actually at the same position relative to the 3' end as box #7 of CCA9 RNA (lane 2), and both show similar amounts of initiation.

Increased numbers of C residues at the 3'-penultimate initiation site leads to a higher ratio of initiation from the 3'-penultimate relative to 3' boxes (Fig. 3B, lanes 16 versus 13). With more C residues at both sites, 3' initiation becomes

FIG. 2. A CCCA 3'-initiation box directs strong 3'-terminal initiation from a wide range of adjacent sequences. (A) RNAs in which the 3'-CCACCA of CCA9 RNA was changed to 3'-NNNCCCA were incubated in the presence of Qβ replicase and [α - 32 P]CTP for 10 min at 37°C as described in Materials and Methods, with MgCl₂ present at 10 mM. As indicated, many of the 3'-heptanucleotide sequences were derived from the 3' termini of RNAs exponentially amplifiable by Qβ replicase: Qβ genome(+), WSI (26); Qβ genome(-) (34); MDV-1 (24); DN3 (40); RQ120 (21); 50#1, 50#2, and 77#1 (7); MNV-1 (2); SV7 (GenBank accession no. L07339); and SV11 (GenBank accession no. L07337). (B) Analysis of transcription products labeled with [α - 32 P]CTP and separated by 12.5% denaturing PAGE, with templates identified by the sequence of their modified initiation box #8. The relative levels of transcription (with reference to box #8 of CCA9) originating from box #9 of each RNA is given at the foot of each lane (average of three experiments; typical standard deviation = 10 to 20%). Below the panel is shown the proportion (% of total) of transcripts >12 nt in length originating from box #9 (box #8 for CCA9). *, Quantitation of transcription from box #8 in the case of CCA9 RNA.

relatively more favored, although internal initiation is still important (Fig. 3B, lanes 14 and 15). These results indicate that the proportion of end-to-end transcription is strongly influenced by the C_nA sequences present at and adjacent to the 3' end. At both positions, but particularly at the 3' end, initiation is favored by a larger number of C residues. Our results refine previous findings (15) that somewhere between 5 and 20 3'-C

residues are able to activate transcription of an RNA by Qβ replicase.

To what extent do the upstream CCA boxes of AAACCCA and related RNAs contribute to transcription? The template activities of AAACCCA and related RNAs analyzed in Fig. 2 indicate the capacity of 3'-CCCA to independently direct initiation, unsupported by a neighboring CCA box. However, we

FIG. 3. C₃₋₅A sequences provide optimal 3' initiation sites. The RNAs shown in panel A were analyzed as transcriptional templates for Q β replicase (panel B) as described in Fig. 2. The relative levels of transcription (% of total, with reference to box #8 of CCA9) originating from boxes #8 and #9 of each RNA are given below each lane (average of three experiments; typical standard deviation = 10 to 20%); note that for lanes 14 and 15, the numbers refer to initiation from the 3' and 3'-penultimate boxes, respectively. Below these figures are shown the proportions of transcripts >12 nt in length originating from box #9 (as well as box #8 for CCA9) as a percentage.

were interested to know whether the seven contiguous upstream CCA boxes were important contributors to the strong initiation from the 3' end. Such a contribution seemed plausible in view of the well-known ability of Q β replicase to transcribe poly(C) but not other homopolymers (6) and the prevalence of pyrimidine-rich sequences in replicons (7). To experimentally address this question, we analyzed transcription from variants of AAACCCA RNA in which (CCA)₃ segments were changed to A₄UA₄ (RNAs AAA31 to AAA35), (CCA)₅ segments were changed to A₄UA₅UA₄ (RNAs AAA51 to AAA53), and (CCA)₇ to (A₂UA₂UA)₃ (AAA71 RNA) (Fig. 4A). The interspersed U residues were designed to prevent the replicase slippage anticipated on longer A tracts. As well as testing a role for C clusters as nonspecific transcriptional enhancers, these RNAs permitted testing a role for a short C cluster at an appropriate spacing upstream of the 3'-initiation box in enhancing transcription (proposed in reference 18), perhaps by binding to so-called Site II of the EF-Tu subunit (8).

All of the modified RNAs showed robust predominant initiation from the 3' end (Fig. 4). Similar results were obtained with variants in which CCA boxes #1 to 7 were individually

changed to AAA (not shown). Most convincingly, AAA71 RNA (Fig. 4B, lane 12), which lacks any C residues upstream of the 3'-CCCCA initiation box, supported similar levels of end-to-end transcription as the AAACCCA parental RNA (lane 3); essentially all transcription was end to end. Similarly strong end-to-end transcription was observed from AAA72 RNA (Fig. 5, lane 3; 170% transcription relative to box #8 of CCA9), an RNA related to AAA71 RNA, but with only three U residues in the A-rich tract upstream of the 3'-CCCCA. Mutation of each CCA box individually from the various RNAs tested in Fig. 4 resulted in the loss of initiation from that site (indicated with a dot next to each lane) but only modest (if any) changes in the initiation levels from adjacent CCA boxes. These results indicate that all CCA boxes serve as independent initiation sites and that the upstream CCA tract provides little or no enhancement of transcription initiation from the 3'-CCCCA. The notion of a role for an upstream C cluster (18) has been questioned by the inapparent conservation of such a feature among RNAs amplifiable by Q β replicase (26).

Contribution of the 3'-A to initiation box function. The role of the 3'-A in the CCCA 3'-initiation box was tested by ob-

FIG. 4. Upstream CCA boxes contribute little to initiation from a 3'-CCCA sequence. (A) Derivatives of AAACCCA RNA in which CCA boxes #1 to #7 have been replaced progressively with A-rich sequences (underlined). (B) Analysis of the same RNAs as transcriptional templates for Q β replicase, performed as described in Fig. 2. The dots placed to the left of lanes indicate absent signals from the mutated initiation boxes. The relative level of transcription (percentage of total, with reference to box #8 of CCA9) originating from box #9 of each RNA is given below each lane (average of three experiments; typical standard deviation = 10 to 20%), as is the proportion (percentage of total) of transcripts >12 nt in length originating from box #9.

serving transcription from AAA72 RNA with the 3'-A present or absent (3'-CCCA and 3'-CCC termini, respectively). AAA72 RNA lacking the 3'-A yielded less than one-tenth the amount of transcript produced from full-length AAA72 RNA (Fig. 5, lanes 3 and 4), indicating the importance of the addition of nontemplated A by Q β replicase to the 3' end of product strands. Transcription from the 3'-CCC terminus was even poorer than from AAA72 RNA terminating in 3'-CCA (Fig. 5, lane 2). Comparison of the same three termini on two other RNAs likewise emphasized the importance of the 3'-terminal A residue: with both GGA(CCA)₇AAAX and GGA(CCA)₇CCUX RNAs, where X is CCA, CCCA, or CCC, about one-fifth the level of transcription observed from CCCA ends was observed for both CCA and CCC ends (not shown). Our results are consistent with studies reporting 33% template activity for Q β genomic RNA lacking the 3'-A of Q β genomic RNA (25), suggesting fundamentally similar recognition of the 3' termini of the viral RNA and the synthetic templates we have used here.

A specific role for A at the 3' end was tested with a GGA(CCA)₇C₅N RNA family, in which N was A, G, U, or C. Transcription was strongest with a 3'-A (Fig. 5, lane 6), and 56 to 66% that level with the other nucleotides at the 3' end (lanes 7 to 9) when quantitated on the basis of the entire cluster of bands derived from the C-rich initiation box (bracketed "C" in Fig. 5B). The yield of full-length transcript (band N) was only 34 to 35% from RNAs with a 3'-C or -U (relative to RNA with

3'-A), reflecting the presence of considerable band heterogeneity in the "C" clusters of lanes 8 and 9. This heterogeneity includes an additional band N + 1, which could reflect initiation opposite the additional 3'-pyrimidine, stuttering during initiation, or a gel artifact: despite the use of antisense DNA to hybridize the template molecules (see Materials and Methods), some replicase products are harder than others to completely denature, and a shadow or faint band at the N + 1 position is occasionally seen. For the RNAs with 3'-U or -C (lanes 8 and 9), the band heterogeneity is also expressed as the N band being less prominent than the bands immediately below: we do not know whether this reflects staggered initiation at alternative C residues or a differential heterogeneity introduced during termination.

In summary, a 3'-A is an important feature of a template with three C residues in the 3'-initiation box, although it is less important when more C residues (five or six) are present. In the latter case, a 3'-A additionally has a potential role in minimizing initiation heterogeneity.

A 3'-CCCA initiation box alone supports transcription at levels similar to that seen with replicon RNAs. Having concluded from the above study that no specific sequence other than a 3'-CCCA (or closely related sequences) is required for transcription by Q β replicase, we wanted to test whether the transcription levels seen with our short, linear RNAs are comparable to those supported by RNAs capable of amplification. Transcription of CCA9, AAACCCA, and AAA71 RNAs (Fig.

FIG. 5. Role of the 3'-terminal A residue. Two families of RNAs are analyzed as transcriptional templates for Q β replicase as described in Fig. 2. In lanes 2 to 4, derivatives of AAA72 RNA with the indicated 3' ends are analyzed: the AAA72 RNA family has the sequence GGA₅UA₆UA₅UA₆ upstream of the indicated C-rich initiation box. In lanes 6 to 9, derivatives of GGA(CCA)₇X RNA, with the indicated 3' ends representing X, are analyzed. The major end-to-end transcription product is marked N at the right side of the figure, with the upper band evident in lanes 8 and 9 marked N+1. The entire cluster of bands originating from the initiation boxes of the RNAs in lanes 6 to 9 is bracketed and labeled "C." The relative percent transcription yields are given below each lane, separately for C, N, and N+1 in panel B. The yields of products in lanes 3 and 6 were 170 and 125%, respectively, relative to transcription from box #8 of CCA9 RNA. The RNAs in lanes 6 to 9 were previously analyzed (38) at 21 mM MgCl₂, which favors internal over 3'-end initiation.

4) was compared with product synthesis from two replicon RNAs capable of amplification by Q β replicase: MDV and DN3 RNAs. Single-round transcription was studied through the addition of the polymerase scavenger PES (see Materials and Methods). In the presence of PES, AAACCCA and AAA71 RNAs (each 31 nt long) yielded similar amounts of transcript as DN3 RNA, a replicon of similar length (34 nt [40]) (Table 1). Some 3.5 times more transcript was obtained from MDV RNA, a 225-nt-long replicon (Table 1). We have previously observed a marked length effect with short transcripts, CCA12 (39 nt) supporting about three times the transcription derived from CCA9 RNA (38). The increased transcription from MDV relative to AAA71 RNA may be due to its greater length or other properties, such as the presence of a replicase recognition site (22). Nevertheless, the model RNA AAA71 with transcription driven solely by a 3'-CCCA initiation box supports transcription levels comparable to those of replicon RNAs. Similar results were observed at 10-fold-lower

template and replicase concentrations of 10 and 5 nM, respectively (data not shown).

Can a 3'-CCCA transcription signal provide sufficient template specificity in vivo?: the case of tRNA^{His}. While the results presented above indicate that an accessible CCCA positioned at the 3' end of an RNA is sufficient to make that RNA a strong template for Q β replicase, we were concerned that this may not be sufficiently specific to prevent transcription of some of the abundant stable RNAs present in an *E. coli* cell, especially some of the tRNAs, all of which end in the conserved 3'-CCA. Perusal of the sequences of all *E. coli* tRNAs (32; <http://www.uni-bayreuth.de/departments/biochemie/trna/>) and other RNAs (Blattner et al., <http://www.genome.wisc.edu/k12.htm>) revealed that only tRNA^{His} has a 3'-CCCA. Note, however, that tRNA^{His} is also unique in possessing an additional 5'-G relative to other tRNAs, resulting in the underlined "discriminator base" being base paired (Fig. 6A).

To see whether tRNA^{His} could serve as a template for Q β replicase, we synthesized by in vitro transcription a form of the RNA lacking the usual base modifications (Fig. 6A). Only weak end-to-end transcription was supported by the tRNA^{His} transcript (4.5% yield relative to transcription from box #8 of CCA9; Fig. 6B, lanes 1 and 2). To test whether this weak transcription was due to the poor accessibility of the 3'-CCCA, transcription was also tested from three variants which either lack the 5'-G or 5'-GGU or have an additional AACCCA at the 3' end. Removal of bases from the 5' end progressively resulted in increased transcription (Fig. 6B, lanes 3 and 4), while addition of the 3'-AACCCA resulted in a still-higher level of transcription (78% relative to box #8 of CCA9). This latter level of transcription is comparable to that observed from AAA71 and other RNAs in Fig. 4 that terminate in 3'-CCCA, indicating that the body of the tRNA neither enhances nor represses transcription directed by the CCCA initiation box. These results demonstrate that host tRNA^{His} is only a poor template for Q β replicase and that this is due to the inaccessibility of the initiation site due to base pairing.

Transcription initiation from a 3'-UCCA terminus. The results shown in Fig. 1 to 6 clearly demonstrate the ability of 3'-CCCA and related sequences with additional C residues to direct transcription by Q β replicase, explaining the presence of such sequences at the 3' ends of almost all replicons. However, we know of two instances in the literature that deviate from this rule: a host factor-independent variant of Q β RNA that is able to amplify in vivo and has a CCUCCA 3' end (28) and

TABLE 1. Single-round transcription of RNAs by Q β replicase ^a	
RNA template	% Relative molar transcription (+ PES) (mean $\pm$ SD)
DN3	86 $\pm$ 5
MDV	347 $\pm$ 63
CCA9	54 $\pm$ 7
AAACCCA	88 $\pm$ 1
AAA71	100

^a Transcription reactions were performed as described in Materials and Methods in the presence of the polymerase scavenger PES, added after a 1-min preincubation in the absence of CTP. DNA3 RNA is a 34-nt amplicon with the 3' sequence GAUCCA (40). The MDV RNA was a variant of MDV-1 (19) referred to as Syn5.0 RNA (10). The figure entered for CCA9 RNA refers to transcription initiating from CCA box #8.

FIG. 7. Transcription from a 3'-UCCA terminus. Transcription from GGA(CCA)₇CCUCCA and CCA9 RNAs by Q β replicase was compared as described in Fig. 2. CCA9 RNA (30 nt) is the shorter of the two RNAs by 1 nt.

These experiments provide a clear demonstration of the importance of a 3'-CCCA terminus in directing productive end-to-end copying by Q β replicase; its importance has previously been deduced from comparison of the sequences of amplifiable RNAs (26).

The 3'-most A residue of a CCCA initiation box plays a critical role, its removal leading to a 5- to 10-fold decrease in end-to-end transcription (Fig. 5A and experiments not shown). With an initiation box containing five C residues, the role of a terminal A is less distinct, its substitution with another base resulting in less than a twofold drop in transcription level (Fig. 5B). However, since Q β RNA and its related short replicons typically possess 3'-initiation boxes with three or four C residues (e.g., see Fig. 2A), the 3'-A must be a critical feature of RNAs replicated in the cell by Q β replicase. It is interesting that this A is not templated by the RNA being transcribed but rather is added by the replicase in a step associated with termination (25, 36). Nontemplated 3'-terminal residues are common among eukaryotic positive-strand RNA viruses (9, 13, 37) and have also been observed to be necessary for efficient transcription from the 3' end (see, for example, references 14 and 31). In cases where the viral RdRp itself is responsible for the nontemplated addition, this activity could be a viable target for antiviral drugs.

We have previously reported the inability of Q β replicase to initiate transcription from a base-paired initiation box (39), and the importance of a 3' end free of secondary structure was earlier deduced from the properties common to replicating RNAs (4, 26). The results shown in Fig. 6 further demonstrate the importance of an accessible initiation site. Although the 3'-CCCA of *E. coli* tRNA^{His} lacking its normal 5'-G (Fig. 6, lane 3) is not base paired, it is poorly used by Q β replicase. Transcription is increased by positioning the 3'-A at the end of

a 6-nt unpaired tail (lane 4) and even more by positioning it at the end of a 9-nt single-stranded tail (lane 5). These results seem to indicate that an unstructured 3' end more than 6 nt long is preferred by the enzyme active site for initiation at the 3' end. The existence of base pairing at the 3' end of Q β RNA (3) is thought to be an important contributing factor in the requirement for host factor (Hfq) in positive-strand RNA transcription. 6S RNAs, which have unstructured 3' ends, are transcribed independent of the host factor (6). Interestingly, the addition of CC to the 3' end of Q β RNA largely removes the Hfq dependence (28); applying the conclusions from our study, we can attribute this effect to a more extended single-stranded initiation site and the increased transcriptional strength derived from the additional C residues.

Optimal templates lack CCA or a closely related sequence adjacent to the 3' end. The results in Fig. 3 show that predominant initiation at the 3' end (and consequently end-to-end transcription that is productive for replication) is not guaranteed by a 3'-CCCA. When C₂₋₄A is present immediately upstream, considerable internal initiation can occur: in this position, CCA is about as potent an initiation site as CCCA is at the 3' end (Fig. 3). Neither CCA, CCCA, nor CCCCCA serve as strong initiation sites when placed further from the 3' end (Fig. 3, lanes 13 to 15), although we have observed substantial initiation from a C₈A tract 15 nt from the 3' end of an RNA (39). A priority in the evolution of a Q β replicon would thus clearly be the avoidance of C₂₋₄A or related C-rich sequences immediately upstream of the 3' initiation box or at least the placement of such sequences in base-paired structures. In fact, RNAs amplified by Q β replicase typically lack C-rich sequences in this position (see box #8 in the RNAs listed in Fig. 2A). An exception is Q β positive-strand RNA, in which CCU lies upstream of the 3'-CCCA. We have previously shown that A, and to a lesser extent G but not U, punctuating a run of C residues is critical in defining an internal initiation site (38): that is, initiation will occur from the 3' end of a run of C-rich pyrimidine residues, either adjacent to a purine (preferably A) or from the 3' end of the RNA. Thus, the CCUCCA 3' end of Q β positive-strand RNA should not suffer from excessive internal initiation.

It is interesting that CCA suffices for strong internal initiation from near the 3' end, whereas an additional C (i.e., CCCA) is required for a comparable level of initiation from the 3' terminus. The polymerase active site has a clear preference for engaging a C-rich initiation box with a short 3' overhang, which presumably stabilizes the template-enzyme interaction. The recent crystal structure of the RdRp of the double-stranded RNA bacteriophage ϕ 6 (11), which has revealed strong similarities to the RdRp of the positive-strand RNA hepatitis C virus, suggests how such an overhang may provide initial binding stability. Cocrystallization of the RdRp with a DNA template mimic and subsequently with initiating nucleotide GTP showed that the template initially binds with its 3'-C positioned in a "specificity pocket" that is one nucleotide-equivalent beyond the active site. With GTP present, the template ratchets back to position the 3'-C in the active site to template initiation. Q β replicase may have an analogous "specificity pocket," but since initiation occurs opposite the 3'-penultimate nucleotide, there would be no need for the template to ratchet back. For internal initiation, low-specificity contacts

to 3'-proximal nucleotides could provide sufficient binding stabilization so that only two C residues suffice. It is proposed that the "specificity pocket" of $\phi 6$ RdRp can swivel aside to provide a path for the template that has already been read to feed through (11), and a similar scheme for Q β replicase could pertain to internal initiation.

Can an accessible 3'-CCCA be a sufficiently restrictive prescription for specific viral amplification in vivo?: the role of replicase binding sites. Our studies described above defined a good template for end-to-end transcription by Q β replicase as one possessing a 3'-CCCA and lacking CC(A/G) immediately upstream. Such a sequence will appear at the 3' end of an RNA with a probability of ca. 1 in 264. This rather modest specificity is increased to a degree that cannot be quantitated by the requirement that the 3'-CCCA be accessible. Until in vivo experiments with decoy RNAs are conducted, we cannot judge whether the transcription dependent on a CCCA initiation box that we have described in this study is sufficiently specific for viable viral replication. However, because endogenous decoy RNAs that could serve as nonproductive templates will do so in proportion to their concentration in the cell, we have perused the range of stable, abundant *E. coli* RNAs for the presence of a 3'-CCCA. This sequence is found only in tRNA^{His}, which was demonstrated in Fig. 6 to be a poor template because of the proximity of the 3'-initiation box to a base-paired helix. tRNA^{His} should thus only be used to a limited degree as a template by Q β replicase in vivo, and perhaps in practice not at all, considering that tRNAs are typically engaged in interactions with aminoacyl-tRNA synthetases, EF-Tu, or the ribosome.

A viable variant of Q β RNA with a 3'-UCCA has been reported (28), and we have confirmed that Q β replicase can effectively initiate transcription from this sequence, at least when juxtaposed next to a CCU upstream sequence (Fig. 7). Among the stable RNAs of *E. coli*, a 3'-UCCA terminus is only present in tRNA^{Cys} (CCUCCA) and tRNA^{Gly} (GCUCCA; underlined nucleotides are base paired). Based on the template activity of tRNA^{His}- $\Delta 5'$ G (Fig. 6, lane 3), which also has a 4-nt non-base-paired 3' end, these tRNAs should not be efficient templates.

While the above considerations suggest that transcriptional control solely by a 3'-initiation box may be feasible, this idea will certainly need to be tested experimentally. Specificity could be readily augmented with *cis*-acting elements that bind replicase, and such elements have been reported in the Q β system: the M and S sites of Q β positive-strand RNA (6) and short RNA sequences that were selected from random sequences by sequential in vitro binding (site I and II ligands that are bound by the S1 and EF-Tu subunits, respectively, of Q β replicase [8]). Such binding sites could be of benefit in ensuring that replicase will preferentially interact with viral RNAs in its search for a template, limiting its interaction with other cellular RNAs. They could also be beneficial as promoters or enhancers, augmenting the transcriptional strength provided by the initiation boxes described here. No experiments have yet tested the contribution of replicase binding sites to the specific transcription or replication of Q β RNA in vivo. It should also be remembered that other features in addition to an appropriate initiation site need to be provided in a successful replicon. One example is a high degree of secondary structure as has

been shown necessary for Q β replicons in vivo (1) and which has been interpreted to permit positive and negative strands to remain separate and to provide protection against RNases.

Three Q β replicase binding sites have been implicated in providing transcriptional strength through in vitro transcription studies: the M site in Q β positive-strand RNA (29) and the site I and II ligand RNAs (8). However, some observations suggest that *cis*-acting sites that tightly bind replicase do not enhance transcription or are not needed in RNAs with an accessible initiation box. First, as reported in Table 1, we observe comparably strong transcription from templates comprised of a 3'-CCCA initiation box coupled to A-rich sequences that are not tightly bound by Q β replicase (6) as from the amplifiable RNA DN3. Second, we observed no improvement in the transcription of an RNA similar to CCA9 when the site II ligand sequence was appended to its 5' end (39). Third, no *cis*-acting sites common to the various RNAs amplifiable by Q β replicase have yet been discerned, although there is evidence that the so-called site II on the EF-Tu subunit is able to accommodate a rather wide range of single- and double-stranded pyrimidine-rich sequences (23). While such site II binding may allow a range of RNAs to bind the replicase, the loose specificity could render this site ineffective in discriminating against nontemplate RNAs in the cell. The clear requirement for M-site sequences (29) is a peculiarity of Q β positive-strand RNA that likely relates to the regulated access of the replicase to an 3'-initiation site that is deliberately inaccessible in its default state (due to base pairing). It has already been convincingly shown that the S site functions in the regulation of translation and not transcription (34).

cis-Acting signals that function as replicase binding sites offer the appealing mechanistic view that productive transcription occurs after templates attract replicase molecules to the vicinity of the active site. On the other hand, transcriptional control by kinetic as distinct from binding discrimination has been well documented in at least two examples (16, 35) and could be the way in which the "accessibility" of the 3'-initiation box contributes to initiation discrimination by Q β replicase. Indeed, such specificity control was suggested by Blumenthal (5), based on the observation that template specificity can be overcome by altering Mn²⁺, glycerol, and GTP concentrations, and that weak templates required higher GTP concentrations for initiation. One can imagine that the half-life of an initiation site poised in the active site awaiting initial phosphodiester bond formation will vary depending on the surrounding sequence and the tendency for the initiation site vicinity to participate in base pairing and folding. A tendency to fold will antagonize active site binding, and bond formation will only occur rapidly when the GTP concentration is high enough to ensure that the first two GTPs are present in the active site whenever the unfavored, unfolded conformation of the initiation site enters the enzyme active site.

Further experiments are needed to examine the contribution of the transcription control mechanisms discussed above to RNA synthesis by Q β replicase. A biological system likely benefits from the use of diverse mechanisms, and transcription may be controlled by both replicase binding and kinetic control strategies. Nevertheless, the experiments presented in this study make a strong case that transcription by Q β replicase is

heavily dependent on control by 3'-CCCA (or closely related) initiation boxes.

ACKNOWLEDGMENTS

We are grateful to Michael Farrell of Vysis, Inc., for the generous gift of Q β replicase.

This work was supported by NIH grant GM54610.

REFERENCES

- Arora, R., C. C. Priano, A. B. Jacobson, and D. R. Mills. 1996. *cis*-Acting elements within an RNA coliphage genome: fold as you please, but fold you must! *J. Mol. Biol.* **258**:433–446.
- Avota, E., V. Berzins, E. Grens, Y. Vishnevsky, R. Luce, and C. K. Biebricher. 1998. The natural 6S RNA found in Q β -infected cells is derived from host and phage RNA. *J. Mol. Biol.* **276**:7–17.
- Beekwilder, M. J., R. Nieuwenhuizen, and J. van Duin. 1995. Secondary structure model for the last two domains of single-stranded RNA phage Q β . *J. Mol. Biol.* **247**:903–917.
- Biebricher, C. K., and R. Luce. 1993. Sequence analysis of RNA species synthesized by Q β replicase without template. *Biochemistry* **32**:4848–4854.
- Blumenthal, T. 1980. Q β replicase template specificity: different templates require different GTP concentrations for initiation. *Proc. Natl. Acad. Sci. USA* **77**:2601–2605.
- Blumenthal, T., and G. G. Carmichael. 1979. RNA replication: function and structure of Q β replicase. *Annu. Rev. Biochem.* **48**:525–548.
- Brown, D., and L. Gold. 1995. Selection and characterization of RNAs replicated by Q β replicase. *Biochemistry* **34**:14775–14782.
- Brown, D., and L. Gold. 1996. RNA replication by Q β replicase: a working model. *Proc. Natl. Acad. Sci. USA* **93**:11558–11562.
- Buck, K. W. 1996. Comparison of the replication of positive-stranded RNA viruses of plants and animals. *Adv. Virus Res.* **47**:159–251.
- Burg, J. L., P. B. Cahill, M. Kutter, J. E. Stefano, and D. E. Mahan. 1995. Real-time fluorescence detection of RNA amplified by Q β replicase. *Anal. Biochem.* **230**:263–272.
- Butcher, S. J., J. M. Grimes, E. V. Makeyev, D. H. Bamford, and D. I. Stuart. 2001. A mechanism for initiating RNA-dependent RNA polymerization. *Nature* **410**:235–240.
- Chapman, M. R., and C. C. Kao. 1999. A minimal RNA promoter for minus-strand RNA synthesis by the brome mosaic virus polymerase complex. *J. Mol. Biol.* **286**:709–720.
- Collmer, C. W., and J. M. Kaper. 1985. Double-stranded RNAs of cucumber mosaic virus and its satellite contain an unpaired terminal guanosine: implications for replication. *Virology* **145**:249–259.
- Deiman, B. A. L. M., A. K. Koenen, P. W. G. Verlaan, and C. W. A. Pleij. 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.
- Feix, G., and H. Sano. 1975. Initiation specificity of the poly(cytidylic acid)-dependent Q β replicase activity. *Eur. J. Biochem.* **58**:59–64.
- Gaal, T., M. Bartlett, W. Ross, C. L. Turnbough, Jr., and R. Gourse. 1997. NTP concentration as a regulator of transcription initiation: control of rRNA synthesis in bacteria. *Science* **278**:2092–2097.
- Kao, C., M. Zheng, and S. Rudisser. 1999. A simple and efficient method to reduce nontemplated nucleotide addition at the 3' terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5**:1268–1272.
- Küppers, B., and M. Sumper. 1975. Minimal requirements for template recognition by bacteriophage Q β replicase: approach to general RNA-dependent RNA synthesis. *Proc. Natl. Acad. Sci. USA* **72**:2640–2643.
- Mills, D. R., F. R. Kramer, and S. Spiegelman. 1973. Complete nucleotide sequence of a replicating RNA molecule. *Science* **180**:916–927.
- Moody, M. D., J. L. Burg, R. DiFrancesco, D. Lovern, W. Stanick, J. Lin-Goerke, K. Mahdavi, Y. Wu, and M. P. Farrell. 1994. Evolution of host cell RNA into efficient template RNA by Q β replicase: the origin of RNA in untemplated reactions. *Biochemistry* **33**:13836–13847.
- Munishkin, A. V., L. A. Voronin, and A. B. Chetverin. 1988. An *in vivo* recombinant RNA capable of autocatalytic synthesis by Q β replicase. *Nature* **333**:473–475.
- Nishihara, T., D. R. Mills, and F. R. and Kramer. 1983. Localization of the Q β replicase recognition site in MDV-1 RNA. *J. Biochem.* **93**:669–674.
- Preuss, R., J. Dapprich, and N. G. Walter. 1997. Probing RNA-protein interactions using pyrene-labeled oligodeoxynucleotides: Q β replicase efficiently binds small RNAs by recognizing pyrimidine residues. *J. Mol. Biol.* **273**:600–613.
- Priano, C., F. R. Kramer, and D. R. Mills. 1987. Evolution of the RNA coliphages: the role of secondary structures during RNA replication. *Cold Spring Harb. Symp. Quant. Biol.* **52**:321–330.
- Rensing, U., and J. T. August. 1969. The 3'-terminus and the replication of phage RNA. *Nature* **224**:853–856.
- Schaffner, W., K. J. Ruegg, and C. Weissmann. 1977. Nanovariant RNAs: nucleotide sequence and interaction with bacteriophage Q β replicase. *J. Mol. Biol.* **117**:877–907.
- Schuppli, D., I. Barrera, and H. Weber. 1994. Identification of recognition elements on bacteriophage Q β minus strand RNA that are essential for template activity with Q β replicase. *J. Mol. Biol.* **243**:811–815.
- Schuppli, D., J. Georgijevic, and H. Weber. 2000. Synergism of mutations in bacteriophage Q β RNA affecting host factor dependence of Q β replicase. *J. Mol. Biol.* **295**:149–154.
- Schuppli, D., G. Miranda, S. Qiu, and H. Weber. 1998. A branched stem-loop structure in the M-site of bacteriophage Q β RNA is important for template recognition by Q β replicase holoenzyme. *J. Mol. Biol.* **283**:585–593.
- Singh, R. N., and T. W. Dreher. 1997. Turnip yellow mosaic virus RNA-dependent RNA polymerase: initiation of minus strand synthesis *in vitro*. *Virology* **233**:430–439.
- Sivakumaran, K., and C. C. Kao. 1999. Initiation of genomic plus-strand RNA synthesis from DNA and RNA templates by a viral RNA-dependent RNA polymerase. *J. Virol.* **73**:6415–6423.
- Sprinzel, M., C. Horn, M. Brown, A. Ioudovitch, and S. Steinberg. 1998. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **26**:148–153.
- Trown, P. W., and P. L. Meyer. 1973. Recognition of template RNA by Q β polymerase: sequence at the 3'-terminus of Q β 6S RNA. *Arch. Biochem. Biophys.* **154**:250–262.
- van Duin, J. 1988. Single-stranded RNA bacteriophages, p. 117–167. *In* R. Calendar (ed.), *The bacteriophages*. Plenum Press, New York, N.Y.
- Villemain, J., R. Guajardo, and R. Sousa. 1997. Role of open complex instability in kinetic promoter selection by bacteriophage T7 RNA polymerase. *J. Mol. Biol.* **273**:958–977.
- Weber, H., and C. Weissmann. 1970. The 3'-termini of bacteriophage Q β plus and minus strands. *J. Mol. Biol.* **51**:215–224.
- Wengler, G., G. Wengler, and H. S. Gross. 1979. Replicative form of Semliki Forest virus RNA contains an unpaired guanosine. *Nature* **282**:754–756.
- Yoshinari, S., and T. W. Dreher. 2000. Internal and 3' RNA initiation by Q β replicase directed by CCA boxes. *Virology* **271**:363–370.
- Yoshinari, S., P. D. Nagy, A. E. Simon, and T. W. Dreher. 2000. CCA initiation boxes without unique promoter elements support *in vitro* transcription by three viral RNA-dependent RNA polymerases. *RNA* **6**:698–707.
- Zamora, H., R. Luce, and C. K. Biebricher. 1995. Design of artificial short-chained RNA species that are replicated by Q β replicase. *Biochemistry* **34**:1261–1266.