RNA signals in the 3′ terminus of the genome of *Equine arteritis virus* are required for viral RNA synthesis

Nancy Beerens and Eric J. Snijder

Molecular Virology Laboratory, Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, The Netherlands

RNA virus genomes contain cis-acting sequences and structural elements involved in virus replication. Both full-length and subgenomic negative-strand RNA synthesis are initiated at the 3′ terminus of the positive-strand genomic RNA of *Equine arteritis virus* (EAV). To investigate the molecular mechanism of EAV RNA synthesis, the RNA secondary structure of the 3′-proximal region of the genome was analysed by chemical and enzymic probing. Based on the RNA secondary structure model derived from this analysis, several deletions were engineered in a full-length cDNA copy of the viral genome. Two RNA domains were identified that are essential for virus replication and most likely play a key role in viral RNA synthesis. The first domain, located directly upstream of the 3′ untranslated region (UTR) (nt 12610–12654 of the genome), is mainly single-stranded but contains one small stem–loop structure. The second domain is located within the 3′ UTR (nt 12661–12690) and folds into a prominent stem–loop structure with a large loop region. The location of this stem–loop structure near the 3′ terminus of the genome suggests that it may act as a recognition signal during the initiation of minus-strand RNA synthesis.

INTRODUCTION

Embedded in the genomes of RNA viruses are sequence motifs and structural elements that play a regulatory role in virus replication. Some of these signals are involved in selective interactions of viral RNAs with the RNA-synthesizing machinery, whilst others may play a role during viral protein synthesis or assembly. *Equine arteritis virus* (EAV) is a positive-stranded RNA virus and the prototype of the family *Arteriviridae*, to which Porcine reproductive and respiratory syndrome virus (PRRSV), Simian hemorrhagic fever virus (SHFV) and Lactate dehydrogenase-elevating virus also belong. The family *Arteriviridae* is grouped together with the corona-, toro- and roniviruses in the order *Nidovirales* (Snijder et al., 2005). Positive-strand RNA viruses replicate in the cytoplasm of infected cells, a process that is mediated by an enzyme complex containing the viral RNA-dependent RNA polymerase (RdRp). Genomic RNA serves as template for the production of negative-strand RNA, which is subsequently used as template for the synthesis of new plus strands. The amplification of viral RNA requires the recruitment of the RdRp complex to specific cis-acting sequence motifs or structures (‘replication signals’) within the termini of the viral RNA, where the RdRp complex initiates the synthesis of plus or minus strands (Buck, 1996; van Dijk et al., 2004).

In addition to the production of progeny genome copies, replication of arteriviruses and coronaviruses entails the synthesis of a nested set of 3′ and 5′ co-terminal subgenomic (sg) mRNAs that are generated by a unique mechanism involving a discontinuous step. Accumulating evidence supports a model in which the discontinuous step in sgRNA production occurs during negative-strand RNA synthesis (Sawicki & Sawicki, 1995, 2005; van Marle et al., 1999; Baric & Yount, 2000; Sawicki et al., 2001; Pasternak et al., 2001, 2006; Zúñiga et al., 2004; van den Born et al., 2004). Thus, for nidoviruses both genome replication and sgRNA synthesis would initiate at the 3′ end of the viral genome. Therefore, it is likely that at least part of the regulatory signals for these processes is implicit in the RNA sequence and structure of the genomic 3′ untranslated region (UTR).

Among positive-stranded RNA viruses, there is considerable variety of 3′ terminal structures: tRNA-like elements, poly(A) tails and terminal structures that fit neither of these two categories are found (Dreher, 1999). Nidoviruses have 3′-polyadenylated genomes and the structure of the 3′ UTR upstream of the poly(A) tail has been the focus of several studies. For the coronavirus *Murine hepatitis virus* (MHV), the structure of the 3′ UTR has been probed and several stem–loop structures involved in virus replication were identified (Liu et al., 2001). The minimal signal needed for
initiation of MHV negative-strand RNA synthesis has been mapped to the last 55 bases of the 3′ UTR (Lin et al., 1994). The 3′-proximal domain of the genome folds into a stem–loop structure that has been identified as the binding site for host proteins (Yu & Leibowitz, 1995). For Bovine coronavirus, a pseudoknot interaction between two stem–loop structures upstream in the 3′ UTR has been demonstrated to be required for virus replication (Williams et al., 1999). A similar pseudoknot interaction was found for MHV (Hsu & Masters, 1997; Hsu et al., 2000; Goebel et al., 2004) and predicted for Severe acute respiratory syndrome coronavirus (Goebel et al., 2004).

For the arterivirus PRRSV, a kissing-loop interaction was proposed between a stem–loop structure in the 3′ UTR and an upstream hairpin located in open reading frame (ORF) 7 (Verheije et al., 2002). Recently, the structure of the 3′ UTR of SHFV was investigated by ribonuclease probing experiments (Maines et al., 2005) and a stem–loop structure was identified that contains binding sites for two cellular proteins. To gain insight into the molecular details of EAV RNA synthesis, we developed an RNA secondary structure model for the 3′-proximal region of the genome using chemical and enzymic probing in combination with computer-assisted structure prediction. Based on this structure model, and using an EAV reverse genetics system, several deletions were engineered in the 3′-proximal domain of the genome. These mutants were tested in vivo for their effect on virus replication in general and RNA synthesis in particular. Two RNA domains were identified that are essential for virus replication and most probably play a key role in viral RNA synthesis.

**METHODS**

**RNA secondary structure probing and prediction.** The 3′-proximal region of the EAV genome (nt 12331 to 12704) was PCR amplified with the sense primer E799 (nt 12231 to 12250, with 5′-flanking T7 RNA polymerase promoter sequence) and an antisense primer E804 (nt 12703 to 12704, with 3′-flanking T20) and used for in vitro transcription. The RNA (500 ng) was treated with dimethyl sulfate (DMS; Fluka) at a concentration of 0.1, 0.2 or 0.4% in buffer D [50 mM sodium cacodylate (pH 7.0), 10 mM MgCl₂, 50 mM KCl] or with 1-cyclohexyl-3-[2-morpholinoethyl] carbodiimide metho-p-toluenesulfonate (CMCT; Aldrich) at a concentration of 4, 8 or 18 mg ml⁻¹ in buffer C [50 mM sodium borate (pH 8.0), 10 mM MgCl₂, 50 mM KCl]. For enzymic probing, the RNA was treated with 0.25, 0.5 or 1 U RNase T1 (Invitrogen), 2.5 × 10⁻⁵, 5 × 10⁻⁵ or 10 × 10⁻⁵ U RNase T2 (Invitrogen) or 0.5 × 10⁻⁵, 1 × 10⁻⁵ or 2 × 10⁻⁵ U RNase V1 (Kemotex Bio) in buffer R [50 mM Tris/HCl (pH 7.5), 10 mM MgCl₂, 50 mM KCl]. All reactions were incubated for 15 min at 37°C. The nucleic acids were recovered by phenol extraction and ethanol precipitation. The sites of modification or cleavage were determined by primer extension analysis using the radioactively labelled antisense primers E804 or E806 (nt 12551 to 12570). Primers were annealed by incubation at 70°C for 10 min, followed by cooling to room temperature over a 30 min period. Subsequently, 25 U Superscript II reverse transcriptase (Invitrogen), dNTPs (100 μM) and 10 × RT buffer (Invitrogen) were added and the reaction was incubated for 1 h at 42°C. The samples were analysed on a denaturing 6% polyacrylamide/urea sequencing gel. Using the same end-labelled primers and the Ladderman Dideoxy Sequencing kit (Takara), a sequence ladder was generated, which was run on the same gel to locate the exact positions of modification or cleavage. RNA secondary structure was predicted using the Zuker algorithm (Zuker, 1989) on the MFOLD web server (Zuker, 2003).

**Site-directed mutagenesis, RNA transfections and analysis of virus replication.** Deletions in the 3′-proximal region of the EAV genome were generated by standard site-directed PCR mutagenesis and introduced into full-length cDNA clone pEAV211 (van den Born et al., 2004). The mutations were verified by sequence analysis. Following in vitro transcription from the full-length cDNA clones, the RNA concentration was measured by UV spectroscopy and the integrity of the RNA was verified by agarose gel electrophoresis. Equal amounts (30 μg) of full-length EAV RNA were transfected into BHK-21 cells by electroporation as described previously (van Dinten et al., 1997). Immunofluorescence dual labelling assays with EAV-specific antisera for non-structural protein 3 (nsp3) (rabbit) and GP5 (mouse monoclonal antibody) were performed as described previously (van der Meer et al., 1998) at different time points after transfection. For RNA isolation, cells were lysed at 14 h post-transfection and intracellular RNA was isolated using the acidic phenol method, as described previously (Pasternak et al., 2000). Viral RNA was analysed on denaturing formaldehyde/agarose gels and hybridized with the radioactively labelled oligonucleotide probe E868 (antisense, nt 12270–12289), which recognizes both genomic and subgenomic positive-strand RNA.

**RESULTS AND DISCUSSION**

**RNA secondary structure model for the 3′ terminus of the EAV genome**

To identify putative regulatory RNA secondary structures in the 3′ end of the EAV genome, we first performed structure-probing experiments. *In vitro*-synthesized RNA representing the 3′-proximal region of the EAV genome [including a 20 residue poly(A) tail] was treated with structure-specific probes (Fig. 1). Nucleotides sensitive to the chemicals DMS and CMCT or the RNases T1 and T2 are assumed not to be involved in base-pairing or base-stacking interactions, whereas RNase V1 is specific for structured or double-stranded regions (Ehresmann et al., 1987). The sites of modification or cleavage were determined by primer-extension analysis. Control experiments with untreated RNA were performed in parallel to detect natural pauses in reverse transcription on the EAV template.

The sites of modification or cleavage that were consistently observed in several experiments are summarized in Fig. 1. Two representative chemical probing experiments are shown in Fig. 2. Computer-based thermodynamic analysis of the EAV 3′ UTR (nt 12646–12704) predicted a stem–loop structure with a large loop region (SL5 in Fig. 1). The structure-probing experiments were in agreement with the predicted structure, as the loop region was found to be highly accessible to single-strand-specific probes. Probing of the region upstream of SL5 (nt 12600–12658) demonstrated that this region was highly exposed to the chemicals DMS and CMCT. Although there were several base-pairing possibilities predicted by MFOLD analysis in this region, the
probing experiments were consistent with the presence of a large single-stranded region and the small stem–loop structure SL4. Thermodynamic analysis of nt 12401–12600 predicted a large extended stem–loop structure in this region with several protruding bulges, internal loops and small stem–loops. The probing results were in good agreement with the MFOLD prediction. The predicted top of the extended stem–loop structure was formed by stem–loop structure SL3. The loop region was highly sensitive to single-strand-specific probes and, in addition, several predicted bulges in the stem were exposed. The upper part of the stem was cleaved by RNase V1, confirming the double-stranded nature of this region. Upstream of SL3, two protruding stem–loops could be recognized, SL2 and SL1. The loop

Fig. 1. RNA secondary structure model of the 3′-proximal region of the EAV genome. Model for the RNA secondary structure of the 3′-terminal 300 nt of the EAV Bucyrus strain genome (GenBank accession no. NC_002532). Nucleotide numbers refer to positions in the genomic RNA, with nt 1 being the capped G residue. The reactive sites for the individual structure-specific probes are indicated. The RNA secondary structure model was obtained using MFOLD-assisted thermodynamic energy minimization and adapted according to the probing results. The deletions that were introduced into the EAV genome are marked in colour.

Fig. 2. Examples of chemical probing experiments. The 3′-proximal region of the EAV genome was treated with increasing concentrations of DMS or CMCT. Untreated control incubations were performed in parallel (−). Sites of modification were detected by primer-extension analysis. The left-hand panel shows the analysis of nt 12620–12690 with primer E804, whilst primer E806 was used to analyse nt 12450–12530 in the right-hand panel. A sequencing ladder was run in parallel to determine the exact positions of modification or cleavage. Ten-nucleotide intervals are indicated on the right-hand side of the autoradiographs.
region of both structures was found to be sensitive to single-strand-specific probes, whereas the stem region was cleaved by RNase V1.

In addition to the structure-probing experiments, sequence variation in the 3′-terminal 300 nt of different EAV isolates was analysed to obtain phylogenetic support for the proposed RNA secondary structure. An alignment of the six most diverse EAV cDNA sequences retrieved from the GenBank database is shown in Fig. 3. However, sequence variation appeared limited in this region of the EAV genome and co-variations supporting the predicted structure were not identified. Most nucleotide variations were found in single-stranded or loop regions. Three nucleotide changes (nt 12440, 12476 and 12534; shaded grey in Fig. 3) that did not disrupt base pairing were identified within base-paired regions and only two nucleotide changes in SL3 (nt 12543 and 12546; indicated in black in Fig. 3) were identified that were predicted to disrupt base pairing. Although these changes affected base pairing in the lower part of the stem, they did not interfere with formation of the SL3 structure. Therefore, similar RNA secondary structures could be predicted for all EAV strains.

Deletions in the 3′ terminus of the EAV genome affect virus replication

To study the role of the identified RNA secondary structures in viral RNA synthesis, several deletions (Figs 1 and 4) were engineered in an EAV full-length cDNA clone (van Dinten et al., 1997), which has been used as the basis for reverse genetics experiments. The deletions were designed carefully so as not to affect the folding of other structural elements in this region. The predicted RNA secondary structure of the deletion mutants is shown in Supplementary Fig. S1 (available in JGV Online). The small protruding SL1 and SL2 structures were deleted in mutants ΔSL1 and ΔSL2, respectively, whereas SL3, representing the top of the large extended stem–loop structure, was deleted in mutant ΔSL3. Interestingly, this region is followed by a lengthy single-stranded region containing only the small SL4 structure. In mutant ΔSL4, SL4 and the single-stranded flanking sequences on either side of this structure were deleted. The 3′ UTR folds into SL5, a stem–loop structure with a large loop region. To disrupt this RNA secondary structure, the left side of the stem region was deleted in mutant ΔS5. In addition, mutant ΔL5 was generated, in which the large loop region was deleted and replaced by four A residues to maintain the SL5 stem.

BHK-21 cells were transfected with infectious RNA transcribed in vitro from wild-type and mutant EAV cDNA clones. After approximately one cycle of virus replication (at 14 h post-transfection), the production of viral proteins was monitored by an immunofluorescence assay (IFA) using antisera recognizing nsp3 and the structural protein GP5 (Fig. 4), which served as indicators for genome replication and sg mRNA synthesis, respectively. As SL1–SL4 are located within ORF7 (Fig. 4a), encoding the nucleocapsid protein, we anticipated that the production of infectious progeny would be affected for these deletion mutants, if they proved to be replication competent. Indeed, positive IFA results were obtained for mutants ΔSL1, ΔSL2 and ΔSL3, but, in contrast to the wild-type control transfection, no
spread of virus to (initially untransfected) neighbouring cells was observed (Fig. 4b). In addition, supernatants harvested from the transfected-cell cultures were tested for the presence of progeny virus using plaque assays. Even at low dilutions, plaques could not be detected, confirming that these mutants were unable to produce infectious progeny (results not shown). No IFA signal was observed for mutants D₅L₁, D₅L₂, and D₅L₅ up to 56 h post-transfection (Fig. 4b) and no infectious progeny virus could be detected using plaque assays (results not shown), suggesting that important replication signals had been deleted or affected in these mutants.

Deletions in the 3’ terminus of the EAV genome affect RNA synthesis

To investigate the synthesis of viral RNA (genomic and subgenomic), intracellular RNA was isolated from transfected cells at 14 h post-transfection. At this time point, the wild-type control virus had not yet spread to uninfected cells (first-cycle analysis) and could therefore be compared with the non-spreading mutants carrying deletions in the nucleocapsid protein gene. We did not anticipate an effect of these deletions on viral RNA synthesis, as our previous studies had demonstrated that all EAV structural proteins, including the nucleocapsid protein, are dispensable for both genome replication and sg mRNA synthesis (Molenkamp et al., 2000).

The isolated intracellular RNA was separated in denaturing formaldehyde/agarose gels and hybridized to an oligonucleotide probe detecting all virus-specific plus-strand RNA molecules. As shown in Fig. 5, for the wild-type virus the full-length genome (RNA1) as well as sg mRNAs 2–7 were detected. All of these RNAs were also produced by mutants D₅L₁, D₅L₂, and D₅L₃, albeit at reduced levels compared with wild-type. Both genome replication and sg mRNA synthesis were affected to the same extent and a reduction to approximately 40 % of the level of the wild-type control was measured for all RNA species by phosphorimager analysis.

No RNA signal was detected for mutants D₅L₄, D₅L₅, and D₅L₆. These results indicated that key signals for viral RNA synthesis are located in the 3’ terminal 100 nt of the genome and that a moderate role in viral RNA synthesis can be attributed to the more upstream sequences deleted in mutants D₅L₁, D₅L₂, and D₅L₃. Mutant D₅L₄ contained a...
Fig. 5. RNA synthesis of wild-type EAV and deletion mutants. Infectious RNA was transfected into BHK-21 cells and intracellular RNA was isolated at 14 h post-transfection. The RNA was separated in a denaturing agarose gel and analysed by hybridization to an oligonucleotide detecting all positive-strand viral RNAs. The positions of the genome (RNA1) and sg mRNAs (RNA2–RNA7) are indicated. The RNAs produced by the mutants migrated somewhat faster than those of the wild-type control virus (wt) due to the deletion (of variable size) introduced in the 3’ end of their genome.

large deletion of 44 nt that included both SL4 and its flanking sequences. Two smaller and more specific deletions were introduced in the SL5 structure. Replacing the loop sequence by four A residues in mutant ΔL5 completely blocked RNA synthesis, suggesting that the sequence and/or conformation of the loop are essential for RNA synthesis. Deletion of the left side of the stem in mutant ΔS5 also inhibited RNA synthesis. This deletion was predicted to disrupt the SL5 structure, suggesting that the stem region is important for the formation and correct presentation of the loop. Alternatively, the structure of the stem region itself or a sequence motif in this region may be important for viral RNA synthesis. The contribution of distinct sequences or structural elements within SL5 to the process of EAV RNA synthesis is the subject of ongoing studies.

Role of the identified RNA signals in virus replication and RNA synthesis

Thermodynamic analysis also predicts stem–loop structures at the 3’ terminus of other arterivirus genomes, although these structures are not identical to the SL5 structure in the EAV genome. For SHFV, structure-probing experiments provided evidence for a stem–loop structure at the genomic 3’ terminus, which also contains a top region highly sensitive to single-strand-specific RNases (Maines et al., 2005). Two cellular proteins were found to bind to this domain and were identified as polypyrimidine-tract-binding protein and aldolase A (Maines & Brinton, 2001; Maines et al., 2005). The same proteins have been reported to interact with the 3’ UTR of EAV and PRRSV (Maines et al., 2005). The biological role and importance of the binding of these proteins to the 3’ UTR remains to be investigated. As there is no obvious sequence conservation between the 3’ UTRs of different arteriviruses, the binding sites of these proteins may be determined by RNA secondary structure. For other nidovirus 3’ UTRs, various interactions involving structure elements have been reported that are important for RNA synthesis and virus replication. For PRRSV, a kissing-loop interaction has been proposed between a predicted stem–loop structure in the 3’ UTR and an upstream hairpin (Verheijen et al., 2002). For several coronaviruses, a pseudo-knot interaction between two stem–loop structures in the 3’ UTR has been reported (Hsue & Masters, 1997; Williams et al., 1999; Hsue et al., 2000; Goebel et al., 2004). Whether the SL5 structure in the EAV genome is involved in such tertiary interactions is not currently known. Since SL5 is located in the 3’ UTR of the EAV genome, it is tempting to speculate that it acts as a direct recognition signal for the initiation of minus-strand RNA synthesis. Alternatively, RNA synthesis may initiate on the poly(A) tail, in which case the RNA structure near the 3’ terminus of the viral genome could provide the specificity to the process. This would allow the RdRp complex to discriminate between the polyadenylated viral RNA genome and polyadenylated cellular mRNAs, which are abundantly present in the host cell’s cytoplasm.

ACKNOWLEDGEMENTS

We thank Erwin van den Born, Jessika Zevenhoven-Dobbe and Willy Spann for technical assistance and helpful discussions. N.B. was supported by Veni grant 700.53.405 from the Council for Chemical Sciences of the Netherlands Organization for Scientific Research (NWO-CW).

REFERENCES


