The arterivirus replicase is the only viral protein required for genome replication and subgenomic mRNA transcription

Richard Molenkamp, Hans van Tol, Babette C. D. Rozier, Yvonne van der Meer, Willy J. M. Spaan and Eric J. Snijder

Department of Virology, Center of Infectious Diseases, Leiden University Medical Center, LUMC P4-26, PO Box 9600, 2300 RC Leiden, The Netherlands

Equine arteritis virus (EAV) (Arteriviridae) encodes several structural proteins. Whether any of these also function in viral RNA synthesis is unknown. For the related mouse hepatitis coronavirus (MHV), it has been suggested that the nucleocapsid protein (N) is involved in viral RNA synthesis. As described for MHV, we established that the EAV N protein colocalizes with the viral replication complex, suggesting a role in RNA synthesis. Using an infectious cDNA clone, point mutations and deletions were engineered in the EAV genome to disrupt the expression of each of the structural genes. All structural proteins, including N, were found to be dispensable for genome replication and subgenomic mRNA transcription. We also constructed a mutant in which translation of the intraleader ORF was disrupted. This mutant had a wild-type phenotype, indicating that, at least in cell culture, the product of this ORF does not play a role in the EAV replication cycle.

The replication of eukaryotic positive-stranded RNA viruses relies on a distinctive process of cytoplasmic RNA-dependent RNA synthesis. The interplay between viral proteins and recognition signals in the RNA template secures the specificity of viral RNA synthesis. Some positive-stranded RNA viruses regulate their genome expression solely at the (post)-translational level, primarily by the synthesis and recognition of viral polyprotein precursors, which are processed into at least 12 to 15 smaller functional subunits by internal viral proteinases (Ziebuhr et al., 2000). The nidovirus replicase proteins are derived from an array of separate genes. These are located in the 3′-terminal part of the genome and are translated from a set of six to eight sg mRNAs. A key feature of these sg transcripts is that the 5′- and 3′-terminal sequences are identical to those of the viral genome. This 5′ and 3′ nested set structure is achieved by fusion of the genomic 5′ leader sequence to specific body transcription-regulating sequences (TRSs) in the 3′-terminal one-third region of the genome. Almost all of the viral genes in the genomic 3′-terminal region are preceded by a TRS, which is termed ‘intergenic sequence’ (IG) in the case of coronaviruses, and thereby they are positioned at the 5′ end of the resulting sg mRNA. The 3′ end of the genomic leader sequence also contains a TRS (leader TRS). Leader and body sequences are fused via an 1-3 partially understood mechanism of discontinuous transcription, which involves base-pairing between the genomic leader TRS and the body TRS complements in the viral minus-strand (van Marle et al., 1999).

Equine arteritis virus (EAV) is the prototype member of the Arteriviridae family (Snijder & Meulenberg, 1998). The 12.7 kb EAV genomic RNA (den Boon et al., 1991) is encapsidated by an isometric nucleocapsid containing a single nucleocapsid protein (N). The envelope which surrounds this core structure contains two major (M and GP1) and three or four minor (E, GP2E, GP3 and GP4) structural proteins (de Vries et al., 1992; Snijder et al., 1999; Snijder & Meulenberg, 1998). As in all nidoviruses, the EAV replicate gene consists of two ORFs, 1a and 1b, which are both expressed from the genomic RNA, the latter by ribosomal frameshifting (den Boon et al., 1991). The EAV ORF1a and ORF1b polyproteins are processed by three viral proteinases, generating 12 nonstructural proteins...
(nsp1–12) and a large number of processing intermediates [for reviews, see Snijder & Meulenberg (1998) and Ziebuhr et al. (2000)]. A number of hydrophobic domains in the ORF1a polyprotein presumably anchor the EAV replication complex to intracellular membranes (van der Meer et al., 1998; van Dinten et al., 1996), resulting in their modification into characteristic double-membrane vesicles (Pedersen et al., 1999).

In addition to the two replicase ORFs and the seven structural genes in its 3′-terminal quarter, the EAV genome contains a 10th potential ORF, which is entirely located within the 211 nt genomic leader sequence (Kheyar et al., 1996). This leader ORF (L-ORF; nt 14–124) encodes a hypothetical 37 amino acid protein. Due to the nested structure of the EAV mRNAs, they all contain the L-ORF in their 5′-terminal region. However, it is unclear whether the EAV L-ORF is indeed expressed, in particular because translation initiation would have to occur very close to the 5′ end of the RNA.

In addition to their function in genome encapsidation, RNA virus coat or N proteins can be involved in genome replication or mRNA transcription. Alfalfa mosaic virus coat protein associates with the viral replicase (Quadt et al., 1991) and plays a role in the initiation of infection and in asymmetric plus-strand accumulation (van der Kuyl et al., 1991). The N proteins of the negative-stranded rhabdoviruses and orthomyxoviruses are part of the helical ribonucleoprotein structure which is the template for viral RNA transcription and replication (Lamb & Kolakofski, 1996; Lamb & Krug, 1996; Wagner & Rose, 1996). A role in viral RNA synthesis has also been postulated for the coronavirus N protein. Antibodies specific for mouse hepatitis coronavirus (MHV) N protein were able to almost completely inhibit viral RNA synthesis in an in vitro system (Compton et al., 1987) and could specifically immuno-precipitate all leader-containing MHV mRNAs, as well as replicative intermediates (Baric et al., 1988; Stohlman et al., 1988). This finding led Baric et al. to propose that the coronavirus N protein might be involved in discontinuous subgenomic RNA synthesis. More recently, immunofluorescence and electron microscopy studies (van der Meer et al., 1999) demonstrated that MHV N protein colocalizes with the viral replicase in membrane-associated complexes that are involved in RNA synthesis. MHV N protein was also shown to interact with the leader and IG sequences in in vitro binding assays (Nelson et al., 2000). Furthermore, it was reported to interact with the cellular heterogeneous nuclear ribonucleoprotein A1 (Wang & Zhang, 1999), which has been suggested to be involved in coronavirus sg RNA synthesis (Li et al., 1997).

Also in arterviruses, the N protein is encoded by the most 3′ structural gene and is abundantly expressed from the smallest sg mRNA (mRNA7). To assess a possible role of EAV N in viral RNA synthesis, we first investigated its subcellular localization by confocal immunofluorescence microscopy. Three different cell lines [baby hamster kidney (BHK-21), rabbit kidney (RK-13) and African green monkey kidney (Vero)] were infected with EAV as described previously (van der Meer et al., 1998). N expression was monitored during time-course experiments and visualized by using monoclonal antibody 51A (Glaser, 1995). N expression was first detected around 6 h post-infection in BHK-21 and RK-13 cells (data not shown) and about 2 h later in Vero cells (Fig. 1), in which EAV
replication is somewhat slower. In all cell lines, the replication cycle was rather asynchronous, since replicase-positive, but N-negative cells could be observed up to 3 h later. The early N signal was characterized by staining of the nucleoli of the infected cell, an observation previously also made for the porcine arterivirus (Rowland et al., 1999), and a perinuclear, cytoplasmic staining that overlapped almost completely with that for the EAV replicase (Pedersen et al., 1999; van der Meer et al., 1998). The partial colocalization with the replicase remained visible throughout infection, also when N accumulated and staining of the entire cytoplasm of the infected cell was observed (Fig. 1).

The above observations, which suggested that EAV N might play a role in viral RNA synthesis, prompted us to systematically investigate the involvement of all EAV structural proteins and the (potential) L-ORF protein in genome replication and sg mRNA transcription. Using site-directed mutagenesis of an EAV infectious cDNA clone (van Dinten et al., 1998), we have recently shown that the proteins encoded by ORFs 2a (E) and 2b (GP) are fully dispensable for viral RNA synthesis (Snijder et al., 1999). This analysis was now extended to ORFs 3–7, the latter encoding the N protein. Mutant full-length clones (Table 1) in which one of these genes was disrupted were generated either by mutagenesis of its translation initiation codon (ORFs 3, 4 and 7), the introduction of a frameshift mutation (ORF6) or by a large internal deletion (ORF5). Since the genes in the 3'-terminal part of the EAV genome all overlap with adjacent ORFs, care was taken to engineer the mutations in such a manner that they did not affect the coding sequences of any of the other genes. In vitro-transcribed, full-length RNA was transfected into BHK-21 cells as described previously (van Dinten et al., 1997). At 12–24 h post-transfection, cells were fixed and analysed using double immunofluorescence assays (IFA). As documented before (van Dinten et al., 1997; van Marle et al., 1999), genome replication and sg mRNA transcription can be demonstrated by staining with an nsp3-specific rabbit antiserum (Pedersen et al., 1999) and GP2- [6D10; Balasuriya et al. (1997)] or N-specific [3E2; MacLachlan et al. (1998)] mouse monoclonal antibodies (data not shown). The results of this analysis are summarized in Table 1. For the ORF2a to 7 knockout mutants, as judged by double IFA, replication and sg mRNA transcription were comparable to those of the wild-type (wt) full-length EAV RNA (EAV030). However, none of these mutants produced infectious progeny virus: spread of the infection to initially untransfected cells was not observed and no plaques were formed in infectious centre assays using cells transfected with these mutants. Incidentally, the production of infectious progeny was observed very late in the experiment (48–72 h post-transfection) for the knockout mutants of ORFs 3, 4 and 7, but this was attributed to reversion of the mutation at the AUG codon.

In addition to the structural gene knockouts, a mutant was constructed in which we removed the translation initiation codon of the potential L-ORF, the only AUG codon upstream of the translation initiation codon of the EAV replicase. On transfection, the L-ORF mutant was able to replicate and synthesize sg mRNAs (Table 1). In addition, similar virus titres were recovered from the medium of cells transfected with this mutant and wild-type EAV030. Progeny virus of the L-ORF mutant was plaque-purified and the presence of the original mutation was demonstrated in three independent plaques by RT–PCR and sequence analysis. These results prove that the L-ORF protein is neither required for EAV replication nor for transcription, and indicate that it is unlikely to play any role at all in the EAV life-cycle, at least in cell culture.

### Table 1. Characteristics of EAV ORF mutants

<table>
<thead>
<tr>
<th>ORF</th>
<th>Type of mutation</th>
<th>Replication*</th>
<th>Transcription†</th>
<th>Infectivity‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Point mutation AUG → GUG</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2a§</td>
<td>Point mutation AUG → AUA</td>
<td>+</td>
<td>+</td>
<td>—</td>
</tr>
<tr>
<td>2b§</td>
<td>Point mutation AUG → ACG</td>
<td>+</td>
<td>+</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>Point mutation AUG → GUG</td>
<td>+</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>Point mutation AUG → ACG</td>
<td>+</td>
<td>+</td>
<td>—</td>
</tr>
</tbody>
</table>
| 5   | Deletion nt 11222–11738 ||||*
| 6   | Frameshift mutation † | +            | +              | —           |
| 7   | Point mutation AUG → ACG | +            | +              | —           |

* As judged by replicase IFA.
† As judged by structural protein IFA.
‡ As judged by spread of progeny virus and/or infectious centre assay.
§ Snijder et al. (1999).
|| Deletion resulted in a +1 frameshift.
§§ Filling in the Xba restriction site at nt 11948 resulted in a +1 frameshift.
Fig. 2. (A) Schematic representation of the structural gene region of the EAV genome and the deletion mutants used in this study. Restriction sites used for the generation of deletions are indicated and deleted sequences are indicated with dashed lines. Note that a BspEI site was engineered in pEAV030 for cloning purposes. The 5' and 3' boundaries of the deletions are also indicated. In constructs 030-2282 and 030-2319, which retain all and a substantial part of the N gene, respectively, the fusion between ORF2b and ORF7 is not in-frame. A '+' sign indicates replication or transcription comparable to wt EAV030. The ' ± ' sign for 030-2511 replication indicates that a faint signal could only be observed at 48 h post-transfection by IFA. The ' ‡ ' sign for 2a-2594 transcription is derived from the results in panel (B), which shows analysis of replication and transcription of wt EAV030 and deletion mutant 2a-2594. Intracellular RNA isolated at 12 h post-transfection was analysed by hybridization with an oligonucleotide recognizing all viral mRNAs. The positions of the genomic RNA (RNA1) and the sg mRNAs (RNA2 to RNA7) are indicated. The position of the truncated mRNA2 of 2a-2594, which was present at low levels, is indicated by '2*'.

Our results with the gene knockouts also indicated that none of the structural proteins (the products of ORFs 2a–7) was required for RNA synthesis. However, the ORF5 and ORF6 mutants that we engineered could still express N-terminal fragments of the GP3 and M proteins (25 and 16 aa, respectively). In addition, we could not rule out that knockout mutants produced small amounts of protein or truncated protein, e.g. due to aberrant or internal initiation of translation or reversion of mutated AUG codons. To circumvent these potential problems, we constructed a set of deletion mutants (Fig. 2A) which lacked various parts of the structural protein coding region of the EAV genome. Each of these deletions completely ruled out the possibility of expression of one or more of the structural genes. Furthermore, they enabled us to assess which RNA sequences in the 3’ end of the EAV genome are required for replication. The nomenclature of the deletion mutants reflected the size (in nt) of the deletion, for example 030-1615 lacked a 1615 nt sequence. Since the RNA2 body TRS of EAV is located within the replicase gene, all mutants were (in principle) able to produce (a truncated) mRNA2. Some of the mutants had also retained the RNA6 and/or RNA7 body TRSs. The largest deletion (2a-2594) extended from nt 9756 to 12350. This mutant lacked the complete sequence of ORFs 2a–6 and lacked the 5’-terminal 40 nt of ORF7.

Mutant and wt full-length RNA transcripts were transfected into BHK-21 cells. Transfection efficiencies, as determined by IFA at 12 h post-transfection, were similar for all mutants except 030-2511 (see below). At 12 h post-transfection, intracellular RNA was isolated and analysed by gel electrophoresis and hybridization with a 32P-labelled oligonucleotide recognizing all viral mRNAs. Fig. 2(A) summarizes the results of this experiment. Although not quantified in any detail, all deletion mutants, with the exception of 030-2511, replicated with an efficiency comparable to that of wt EAV030. In addition, they were all able to synthesize sg mRNA(s). Mutant 030-2511, which retained not more than 147 nt of the genomic 3’ end, was severely compromised in genome replication. Virus-specific RNA was not detected for this mutant by hybridization and only a few replicate-positive cells could be observed by IFA after 48 h (data not shown). This strongly
suggested that in mutant 030-2511 an important 3’ cis-acting RNA replication signal had been disrupted, which most likely resides between nt 12350 and 12557, since constructs 030-2319 and 2a-2594 replicated with wt efficiency (Fig. 2A).

The largest deletion that did not significantly affect genome replication was that in construct 2a-2594. However, this mutant did show a severe reduction of mRNA2 synthesis compared to the wt EAV030 (Fig. 2B). Mutant 030-2319, which contained the same 3’-terminal sequences but retained an intact ORF2a, produced wt amounts of mRNA2 (data not shown). Since we have previously shown that the ORF2a product (the E protein) is dispensable for genome replication and transcription (Snijder et al., 1999), the mRNA2 transcription defect of mutant 2a-2594 is most likely due to an effect of the deletion on the cis-acting RNA sequences required for the synthesis of sg mRNA2. An RNA secondary structure prediction of the RNA2 body TRS region in 2a-2594 showed considerable differences compared to that of the same region in wild-type EAV030 (data not shown). Alternatively, the reduction of 2a-2594 mRNA2 levels might reflect differences in stability of this transcript, which contains an AUG to AU A mutation of the ORF2a translation initiation codon and is probably untranslated.

By deletion analysis of an EAV defective interfering (DI) RNA, we have previously mapped the boundary of the EAV 3’ replication signals to the region between nt 354 and 1066 upstream of the genomic 3’ end (Molenkamp et al., 2000). A construct containing a deletion leaving only 354 nt of the 3’ end failed to replicate (EDIC2-3457; Molenkamp et al., 2000). Mutant 030-2319, however, contains a similar deletion and apparently is able to replicate normally. This clearly shows that the sequence requirements for EAV DI RNA replication might be similar, but are not necessarily identical, to those of full-length genome replication. Similar findings have been reported for the alphavirus Sindbis virus (Niesters & Strauss, 1990).

The experiments in this paper suggest that expression of the EAV replicase gene only is required for genome replication and sg mRNA transcription. Furthermore, the products of EAV ORFs 2a–7 were all found to be essential for the production of infectious virus particles, whereas the functionality of the L-ORF has become highly doubtful. Formally, we cannot rule out the possibility that the L-ORF protein or one of the EAV structural proteins exerts a modest effect on viral RNA synthesis, but it is clear that their presence is not essential for the two main processes of RNA synthesis in the EAV life-cycle. The transcriptional involvement of the N protein in particular is unlikely, since the inactivation of its expression in different ways produced consistent results. In a previous study (van Marle et al., 1999), inactivation of the RNA7 body TRS did not affect genome replication or transcription of the other sg mRNAs. For a number of these TRS mutants, mRNA7 transcription could not be detected even by sensitive RT–PCR methods, suggesting the complete absence of N protein expression. Here, both mutagenesis of the ORF7 AUG codon and deletion of the first 40 nt of ORF7 did not affect replication or transcription. Also mutant 030-2511, containing only the 3’-terminal 80 nt of the N gene, showed a low level of genome replication.

Despite the fact that it apparently is dispensable for RNA synthesis, part of the EAV N protein clearly colocalized with the replication complex (Fig. 1). In our opinion, the most obvious explanation for this colocalization is that it points towards the site of EAV genome encapsidation. Although genome replication and virus budding occur at distinct sites in arterivirus-infected cells, little is known about the intermediate stage of nucleocapsid formation. Newly made genomes may be encapsidated almost immediately, resulting in a preformed nucleocapsid structure that subsequently migrates to the site of virion budding. Possibly, the colocalization of the coronavirus replicase and N protein should be explained in a similar manner, although the pronounced differences between arterivirus and coronavirus N proteins also leave room to speculate on auxiliary functions of the latter, e.g. in viral RNA or protein synthesis. To address this issue, the recent generation of infectious cDNA clones for coronaviruses (Almazán et al., 2000; V. Thiel and others, personal communication) will prove to be instrumental.

We are grateful to Amy Glaser, Udeni Balasurya and James MacLachlan for providing monoclonal antibodies. We thank Leonie van Dinten for providing constructs 030-1615 and 030-2511 and Sasha Pasternak for helpful discussions. R.M. was supported by grant 700-31-020 from the Council for Chemical Sciences of the Netherlands Organization for Scientific Research.

References


Received 19 May 2000; Accepted 30 June 2000