JOURNAL OF VIROLOGY, July 1996, p. 4291–4298 Vol. 70, No. 7
0022-538X/96/$04.00+0

Equine Arteritis Virus Subgenomic mRNA Synthesis: Analysis of Leader-Body Junctions and Replicative-Form RNAs

JOHAN A. DEN BOON, MAURITS F. KLEIJNEN, WILLY J. M. SPAAN, AND ERIC J. SNIJDER*

Department of Virology, Institute of Medical Microbiology, Leiden University, Leiden, The Netherlands

Received 16 January 1996/Accepted 26 March 1996

In addition to the genomic RNA, a 3′ coterminal nested set of six subgenomic mRNAs is produced in equine arteritis virus (EAV)-infected cells. The seven viral RNAs are also 5′ coterminal, since they all contain a 206-nucleotide common leader sequence which is identical to the 5′ end of the genome. A conserved pentanucleotide sequence motif, 5′ UCAAC 3′, was shown to be present at the junctions between the leader and body sequences in each of the mRNAs. In addition, two alternative junction sites were detected for mRNA 3. Seven replicative-form (RF) RNAs (RFs I to VII), corresponding to the genomic RNA and each of the subgenomic EAV mRNAs, could be prepared from lysates of infected cells. The minus-strand RNA contents of these RF RNAs were analyzed by using an RNase protection assay with an RNA probe containing the mRNA 2 leader-body junction. It was established that RF II contained a negative-stranded copy of mRNA 2, including a complementary leader sequence. The presence of subgenomic minus-strand RNA in RFs is indicative of a function as a transcription template during the production of EAV subgenomic mRNAs.

Equine arteritis virus (EAV) is the prototype of the *Arteriviridae*, a newly established family of positive-stranded RNA viruses (for a review, see reference 19) which contains at least three other members: lactate dehydrogenase-elevating virus of mice (LDV) (10), porcine reproductive and respiratory syndrome virus (PRRSV) (4, 17), and simian hemorrhagic fever virus (11). The arteriviruses are evolutionarily related to coronaviruses and toroviruses, which is most clearly reflected by the presence of a set of homologous replicase domains and by common features of their genome organization and expression (6, 10, 28).

The EAV genome is a 12.7-kb RNA molecule that contains eight open reading frames (ORFs) (Fig. 1A). Upon infection, only ORFs 1a and 1b are translated directly from this RNA. Translation of ORF 1b requires ribosomal frameshifting to occur at the ORF 1a-1b overlap region (6). As a result, two large replicase polyproteins (1a and 1ab) are produced, which are extensively processed into smaller subunits (5, 29, 30). ORFs 2 to 7 are expressed from a 3′ coterminal nested set of six subgenomic (sg) mRNAs (Fig. 1B) (8, 35). They encode four structural proteins (9) and two proteins of unknown function. A common 5′ leader sequence of 206 nucleotides (nt), identical to the 5′ end of the genome, is fused to mRNA body sequences which are colinear with the genomic 3′ end (8). For EAV sg mRNAs 6 and 7, a conserved pentanucleotide sequence motif (5′ UCAAC 3′) has been shown to be present at the so-called leader-body junction site (8). In the genomic RNA, this conserved sequence is present at the 3′ end of the common leader sequence and at positions just upstream of each of ORFs 2 through 7 (6). Similar sequence elements, which are assumed to play a crucial role in sg mRNA transcription, have been described for the sg mRNAs of the other arteriviruses (3, 16, 38). A detailed analysis of the leader-body junctions in the LDV and PRRSV sg mRNAs revealed that the fusion of leader and sg RNA body sequences is not restricted to a single position within this sequence motif and that the underlying transcription mechanism may be imprecise (3, 16).

In a previous report, we have demonstrated that EAV sg mRNAs do not result from processing of a larger-genome-length precursor RNA, as has been suggested in the past (35), and that at least late in infection their synthesis is largely independent from that of genomic RNA (7). A comparison with the coronavirus mouse hepatitis virus (MHV) suggested that in EAV-infected cells, sg mRNAs are generated by a coronavirus-like discontinuous transcription mechanism (for reviews, see references 13, 31, and 37). In coronavirus-infected cells, negative-stranded counterparts for all viral mRNAs have been detected (24, 25). Furthermore, these minus strands have been shown to be present in so-called replicative-form (RF) RNAs, which suggests that they are transcriptionally active (22). For the arterivirus LDV, antisense genome-length RNA has been detected in lysates of infected cells, but indications for the presence of sg minus-strand RNA could be obtained only by reverse transcription (RT) in combination with PCR (2). The experiments described in this paper reveal information about the structure and synthesis of both positive- and negative-stranded sg RNAs of arteriviruses. We have analyzed the leader-body junction sites of all EAV mRNAs. Furthermore, direct evidence for the presence of antileader-containing sg minus-strand RNA and six sg RF RNAs in EAV-infected cells is presented.

MATERIALS AND METHODS

Cells and viruses. EAV (Bucyrus strain) was grown in baby hamster kidney cells (BHK-21 cells) at 39.5°C. MHV (A59 strain) and Sindbis virus (SIN) (HR strain) were propagated at 37°C in Sac(−) cells and BHK-21 cells, respectively. All experiments were carried out with a multiplicity of infection of between 10 and 50.

cDNA synthesis and PCR. The generation of cDNA clones of EAV sg mRNAs 6 and 7 has been described previously (8). To obtain additional clones, including those of the other viral RNA species, first-strand cDNA synthesis and PCR were carried out with Moloney murine leukemia virus reverse transcriptase (Gibco-BRL) and *Taq* DNA polymerase, respectively, in the buffers supplied by the manufacturers. The PCR consisted of 30 cycles, each consisting of 1 min of denaturation at 95°C, 1 min of annealing at 5°C below the lowest of the calculated melting temperatures of the primers, and 2 to 3 min of extension at 72°C. The oligonucleotides which were used are listed in Table 1. Because of the fact that the EAV mRNAs form a 3′ coterminal nested set, single RT reactions could

* Corresponding author. Mailing address: Department of Virology, Institute of Medical Microbiology, Leiden University, AZL gebouw 1 P4-26, Rijnsburgerweg 10, 2333 AA Leiden, The Netherlands. Phone: 31-715261652. Fax: 31-715266761. Electronic mail address: E.J.Snijder@Microbiology.MedFac.LeidenUniv.NL.

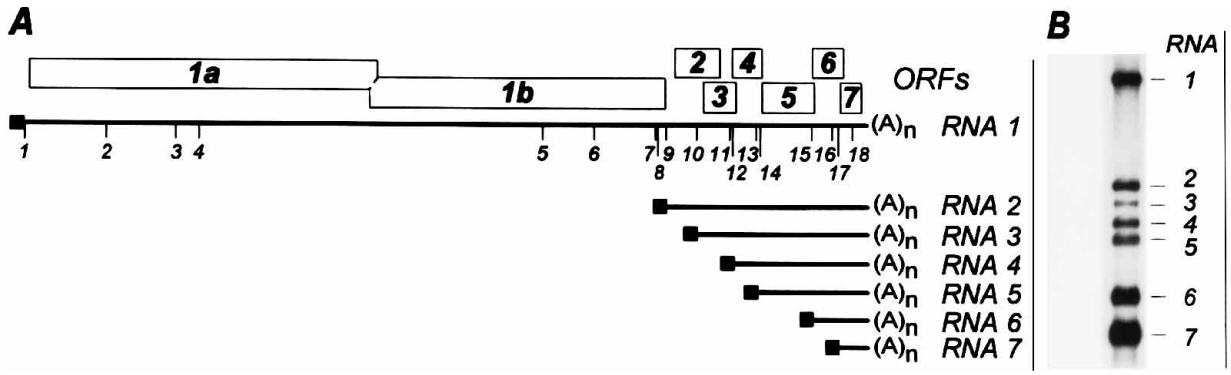

FIG. 1. EAV genome organization and RNAs. (A) The positions of the ORFs and all 5′ UCAAC 3′ sequence motifs, numbered 1 to 18, are indicated along the 12.7-kb positive-stranded genomic RNA. Also shown is the nested set of six sg mRNAs, from which ORFs 2 to 7 are expressed. The 5′ common leader sequence is represented by a black box. (B) Hybridization analysis of intracellular EAV RNA resolved by denaturing agarose gel electrophoresis. As a probe, a ^{32}P-labeled oligonucleotide which was complementary to the 3′ ends of all seven viral RNAs (oligonucleotide I [Table 1]) was used.

be used for the determination of multiple leader body-junction sites. An example of this strategy is shown in Fig. 2.

cDNA cloning and sequence analysis. Standard cloning procedures were carried out as described by Sambrook et al. (20). PCR fragments were cloned blunt ended into either the *Sma*I site of pBS$^-$ (Stratagene) or the *Eco*RV site of pUC21BM (Boehringer Mannheim). In the case of sg RNA 2, some of the clones were obtained by digestion of the PCR fragments with *Taq*I and subsequent ligation into the *Cla*I site of pBSSK$^-$ (Stratagene). Transcription plasmids pRPP21 and pRPT20 were constructed by inserting RNA 2-derived cDNA fragments downstream of a T7 promoter. With these plasmids, in vitro transcription was carried out to generate probes and control RNA to be used in RNase protection assays (see below). Sequence analysis was performed with the Pharmacia T7 sequencing kit and [α-^{33}P]dATP according to the manufacturer's instructions.

Analysis of intracellular RNA. The extraction of total intracellular RNA and denaturing RNA electrophoresis were carried out as described previously (7, 32). In order to prepare RF RNA, the lysis procedure described by Sawicki and Sawicki (22) was applied. Lysates were treated with DNase I (Gibco-BRL; 30 min, 37°C, 5,000 U/ml) and RNases A and T_1 (Boehringer Mannheim; 30 min, 37°C, 5 μg/ml and 100 U/ml, respectively). For direct analysis of double-stranded RF RNA, nondenaturing 1 to 1.25% agarose–Tris-borate-EDTA gels were used, essentially as described by Sawicki and Sawicki (22).

Preparation of oligonucleotide probes and hybridizations. For detection of EAV plus-strand RNA, oligonucleotide I (Table 1) was labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase and used for hybridization in dried agarose gels as described by Meinkoth and Wahl (15).

Metabolic RNA labeling. RNA synthesis during specific time intervals in infection was monitored by metabolic labeling with $H_3{}^{32}PO_4$ (100 μCi/ml) or [^{3}H]uridine (250 μCi/ml). One hour in advance and during labeling, cells were incubated in medium to which 10 μg of dactinomycin per ml was added in order to block host RNA synthesis. In the case of $H_3{}^{32}PO_4$ labeling, cells were first washed with a phosphate-free balanced salt solution (137 mM NaCl, 4 mM KCl, 1.6 mM $CaCl_2$, 1 mM $MgCl_2$, 23.8 mM $NaHCO_3$, 10 mM D-glucose) and were incubated under phosphate-free conditions until lysis.

Sucrose gradient ultracentrifugation of RF RNA. To separate individual RF RNA species prior to electrophoresis or RNase protection analysis, total RF preparations (equivalent to 10^7 EAV-infected BHK-21 cells) were loaded onto 3 to 30% nonlinear sucrose gradients buffered with 50 mM Tris HCl (pH 7.5)–100 mM LiCl–1 mM EDTA–0.1% sodium dodecyl sulfate (SDS). Centrifugation was carried out at 50,000 rpm for 2.25 h at 15°C, using a Beckman SW55Ti rotor. After centrifugation, 600-μl fractions were collected from bottom to top.

RNase protection assays. RNase protection assays were performed essentially as described by Novak and Kirkegaard (18). Control target RNA was prepared with 1 μg of linearized pRPT20 by in vitro transcription for 1 h at 37°C in a 20-μl reaction mixture containing 40 mM Tris HCl (pH 8.0), 8 mM $MgCl_2$, 2 mM spermidine, 25 mM NaCl, 400 ng of bovine serum albumin, 5 mM dithiothreitol, a 400 μM concentration of each of the four ribonucleotides, 20 U of RNAguard, and 20 U of T7 RNA polymerase. Transcripts derived from pRPT20 were 738 nt in length and contained 630 nt of EAV RNA 2 minus-strand-specific sequence (nt 10214 to 9717 and 213 to 86 in the published genomic sequence [6]). To prepare radioactively labeled probe transcripts, pRPP21 was used. The concentration of each of the four nucleotides in the reaction mixture was lowered to 80 μM, and [α-^{32}P]CTP was included at a concentration of 330 nM. pRPP21 transcripts were 508 nt long, of which 469 nt was EAV RNA 2 plus-strand specific (leader sequence nt 86 to 213 and body sequence nt 9717 to 10053 [6]). After transcription, reaction mixtures were diluted 10-fold in DNase buffer (40 mM Tris HCl [pH 7.5], 6 mM $MgCl_2$, 10 mM NaCl) containing 10 U of DNase I and were incubated for 30 min at 37°C. Prior to use, labeled transcripts were purified from 4% polyacrylamide gels by elution at 56°C for 1 h in a buffer containing 10 mM Tris HCl (pH 8.0), 1 mM EDTA, and 300 mM NaCl. Unlabeled transcripts were purified from low-melting-point agarose gels.

Equal amounts of purified probe transcript were added to each of the samples containing either RF RNA or a control transcript. Subsequently, 1 μg of tRNA was added, and the samples were phenol extracted and precipitated with 96% ethanol. The RNA pellets were resolved in 30 μl of hybridization buffer (80% deionized formamide, 40 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)] [pH 6.4], 400 mM NaCl, 1 mM EDTA), denatured for 5 min at 85°C, and incubated at 50°C for at least 8 h. The RNase treatment was carried out by addition of 300 μl of RNase mixture (300 mM NaCl, 10 mM Tris HCl (pH 7.5), 5 mM EDTA, 15 μg of RNase A per ml, 300 U of RNase T_1 per ml) and subsequent incubation at the ambient temperature for 1 h. RNase activity was stopped by incubating the samples at 37°C for 15 min upon addition of 10 μl of 10% SDS and 10 μl of proteinase K (1 mg/ml). The samples were supplemented with 0.25 μg of tRNA, phenol extracted, and ethanol precipitated, and the RNA pellets were dissolved in 6 μl of loading buffer (90% formamide, 100 mM Tris HCl [pH 8.7], 100 mM H_3BO_3, 2 mM EDTA, 0.01% xylene cyanol, 0.01% bromophenol blue). Prior to being loaded on a standard 4% polyacrylamide sequencing gel, the RNA was denatured at 95°C for 3 to 5 min.

TABLE 1. Oligonucleotides

Oligonucleotide	Sequence (5′ →3′)	Polarity	Position[a]	Usage
A	GTAAATCCTAGAGGGCTTTC	+	81–100	Leader-specific PCR
B	GTAGGCACGACCCAT	−	10321–10307	RNA 2- and 3-specific PCR
C	GGCCCAATCCATGAC	−	10827–10813	RNA 4-specific PCR
D	CAACGTACCAGTCGTA	−	11137–11122	RNA 5-specific PCR
E	CCGCGATCCGTCAGC	−	11265–11251	RNA 2- to 5-specific RT
F	GCTCCCATACCTCAG	−	11909–11895	RNA 6-specific PCR
G	ACGTGATCGTCTTGAC	−	12334–12319	RNA 7-specific PCR
H	TGGTTCCTGGGTGGC	−	12706–12692	RNA 6- and 7-specific RT
I	TTTGGTTCCTGGGTGGC	−	12708–12692	3′ end-targeted hybridization

[a] According to the genomic sequence numbering published by den Boon et al. (6).

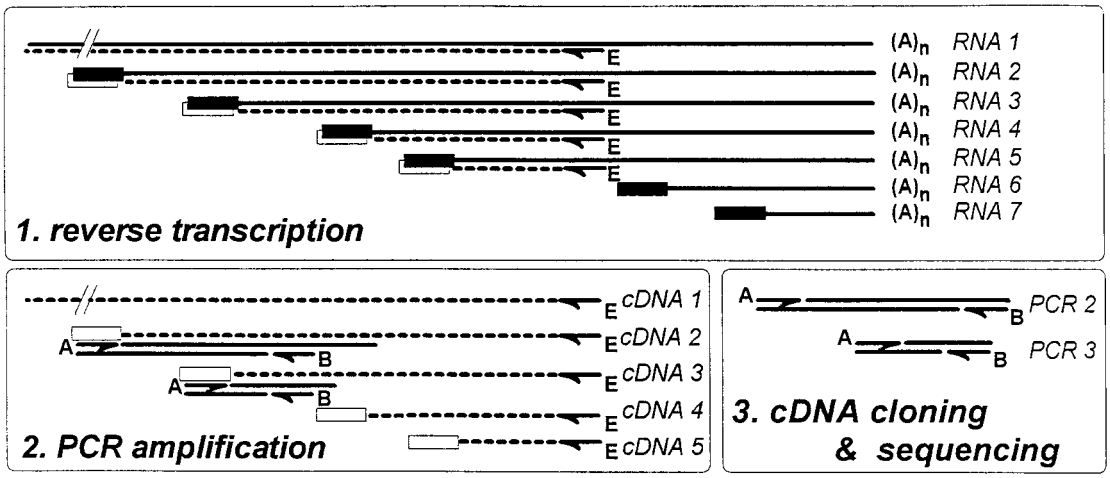

FIG. 2. Cloning strategy for EAV leader-body junction sequences. As an example, the nested RT-PCR strategy used to analyze the leader-body junctions of EAV sg RNAs 2 and 3 is shown. First-strand cDNA was synthesized by using an oligonucleotide primer (E [Table 1]) which was complementary to a part of the ORF 5 sequence, which is present in RNAs 1 to 5. PCR amplification with a primer (A [Table 1]) derived from the leader sequence and a second primer (B [Table 1]) specific for the body sequences of RNAs 1 to 3 yielded DNA fragments which were subsequently cloned and sequenced.

RESULTS

The conserved leader-body junction sequence 5′ UCAAC 3′ is present in all sg mRNAs. Previously, EAV sg mRNA leader-body junction sites have been determined by using three cDNA clones which were derived from mRNA 6 (one clone) and mRNA 7 (two clones) (8). The sequence 5′ UCAAC 3′, which is present in both the 3′ end of the leader sequence and the region upstream of ORF 6 and ORF 7, was found to be the site where leader and body sequences were fused. The analysis of the entire EAV genomic RNA sequence revealed as many as 18 copies of this pentanucleotide sequence motif (Fig. 1A) (6). Since only six sg mRNAs are produced (Fig. 1B), this sequence cannot be the sole determinant for leader-to-body joining during sg mRNA transcription. Upstream of ORFs 4, 5, and 7, two 5′ UCAAC 3′ sequences are present, and upstream of ORF 2, even three 5′ UCAAC 3′ sequences are present. Determining which of those are functional in EAV sg mRNA transcription could shed light on additional sequence requirements. Furthermore, this information would allow a comparison with the recently reported leader-body junction sequences in the sg mRNAs of LDV, PRRSV, and simian hemorrhagic fever virus (3, 16, 38).

The leader-body junction sites of all EAV sg mRNAs were determined by sequence analysis of available cDNA clones and RT-PCR clones obtained by the strategy exemplified in Fig. 2. First, RT was carried out with an oligonucleotide primer which was specific for a subset of the sg mRNAs. cDNA was subsequently amplified by PCR with a primer containing sequences from the EAV leader sequence and a second primer, which was selective for an even smaller subset of the sg mRNAs. To determine whether heterogeneity in the junction sequences, as observed in the PRRSV and LDV sg mRNAs (3, 16), also occurs in those of EAV, at least eight cDNA clones representing each sg mRNA were analyzed. The sequences of the various leader-body junction sites identified in these clones and their positions in the genome (as indicated in Fig. 1A) are listed in Fig. 3. The 5′ UCAAC 3′ sequence was confirmed to be present at the junctions in mRNAs 6 and 7 and was also identified in sg mRNAs 2, 4, and 5. Except for in sg mRNA 6, a conserved U was consistently present downstream of the 5′ UCAAC 3′ motif. Therefore, the consensus of the junction site sequence can be extended to 5′ UCAACu 3′.

The results obtained by analyzing sg mRNA 3-derived cDNA clones were suggestive of a mixed population of sg mRNA 3 molecules in EAV-infected cells. The sequence 5′ UCAACU 3′ was present in some of these cDNA clones, but junction sequence variation was detected in others (Fig. 3). The sequence found in RNA 3 type 3.2.2 suggested that the actual border between leader and body sequences in this particular sg RNA was located 2 or 3 nt upstream of the 5′ UCAACU 3′ motif. Interestingly, additional heterogeneity was observed. Two types of RNA 3 junctions were detected, which apparently were generated by use of two alternative sequences upstream of ORF 3 that resemble the 5′ UCAAC 3′ sequence motif (Fig. 3). Because of the low level of sg mRNA 3 in infected cells (Fig. 1B), we do not have direct evidence for the heterogeneous nature of the EAV sg RNA 3 population. It cannot be formally excluded that the observed heterogeneity resulted from PCR artifacts. However, we consider this possibility unlikely, since similar heterogeneity was not found in any of the cDNA clones derived from mRNAs other than mRNA 3, including the sg mRNA 2 clones, which were derived from the same PCR as the mRNA 3 clones. Furthermore, other regions of the PCR-amplified RNAs were devoid of any mutations. Preliminary data obtained by differential oligonucleotide hybridization techniques suggest that RNA 3 molecule junction types 3.1 and 3.3 are indeed present in infected cells (data not shown).

Detection of a nested set of RF RNAs. The amount of negative-stranded RNA in lysates of EAV-infected cells was found to be too small to be detected by routinely used hybridization techniques (4a). As an alternative approach, we decided to prepare and analyze so-called RF RNA. Replicative RNA intermediates (RIs), which are the transcriptionally active double-stranded RNA molecules, were isolated from infected cells. Single-strand-specific RNases were then used to convert RIs into RF RNA, by removing nascent stretches of RNA that have become detached from the template RNA (27). Since the bulk of viral RNA produced in the infected cell is of positive polarity, it is assumed that most of the minus strands are

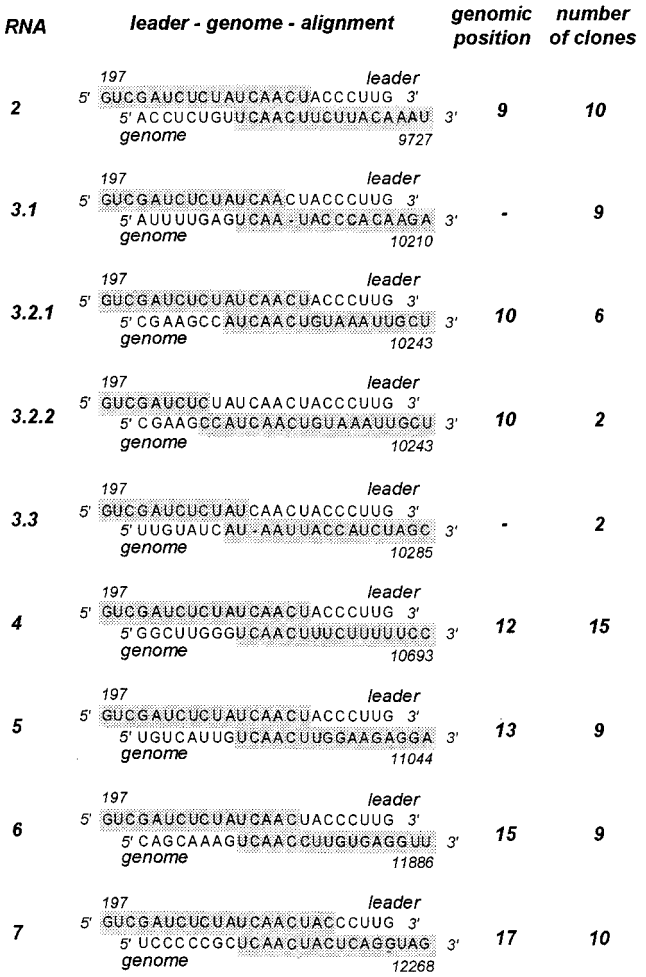

FIG. 3. EAV leader-body junction sequences. For each sg mRNA, the 3′ end of the leader sequence as it is present in the genomic RNA is aligned with the genomic sequence region which contains the 5′ end of the sg body sequence. Indicated are the positions to which the leader-body junction sites correspond on the genomic RNA, as shown in Fig. 1A. For sg RNA 3, a subdivision in different types of molecules, based on the ocurrence of multiple functional leader-body junction sites, is shown. Those sequences which are present in the sg RNAs are indicated by shading.

present in the form of double-stranded RIs, in which they serve as the template for plus-strand synthesis. Therefore, RF RNAs are thought to contain a full-length minus strand which is protected from single-strand RNase activity by short stretches of nascent plus-strand RNA.

EAV-infected BHK cells were metabolically labeled with [^{3}H]uridine, and RF RNAs were prepared upon lysis. After nondenaturing gel electrophoresis, seven prominent RF species could be observed, which were named RFs I to VII in order of decreasing size (Fig. 4A). However, a number of additional, minor bands were seen, the nature of which remained unclear. To check whether the procedure which was followed to obtain the EAV RF RNAs could have been inappropriate, two additional RF RNA preparations were made. RF RNA analyses for both the coronavirus MHV and the alphavirus SIN have been previously described. Seven RF RNAs, which represent the genomic transcription unit and the six sg transcription units, have been described for MHV (22). The genome-length negative-stranded RNA of SIN is known to be the only transcription template for production of both the genomic 49S RNA and the sg 26S RNA (21, 34). From lysates of SIN-infected cells, three RFs can be prepared. RF I corresponds to the full-length 49S transcription unit. RFs II and III are derived from RIs involved in sg RNA transcription and result from cleavage of an RNase-sensitive site at the position of the 26S promoter (27). Our results obtained with lysates of MHV- and SIN-infected cells were consistent with the previously published data (Fig. 4B). Therefore, it was concluded that the RF RNA preparation was carried out properly and that the appearance of additional double-stranded RNA molecules was specific for the EAV RF RNA preparation. Furthermore, the analysis of the three different samples in the same gel allowed a comparison between the sizes of the RF RNAs of EAV, MHV, and SIN. On the basis of the known sizes of the single-stranded RNA species of MHV and SIN (Fig. 4C), it was concluded that the migration of the major EAV RF RNAs corresponded to the nested set of mRNAs which is present in EAV-infected cells. For both EAV and MHV, the genomic RF RNA (RF I) was found to be most abundant.

RF II contains negative-stranded RNA 2 with an antileader sequence. To obtain more information about the composition of the EAV RFs, we used an RNase protection assay to investigate the presence of minus-strand RNA. The approach was similar to that described by Novak and Kirkegaard for poliovirus minus-strand analysis (18) and is summarized in Fig. 5A. After denaturation of RF RNA, minus strands were hybridized to radioactively labeled, in vitro-synthesized transcripts of positive polarity and analyzed by a second round of RNase treatment.

The experiment was designed to discriminate between negative-stranded genomic and sg RNAs and at the same time to provide information about the presence of antileader sequences in sg minus strands. RF RNA was prepared from two

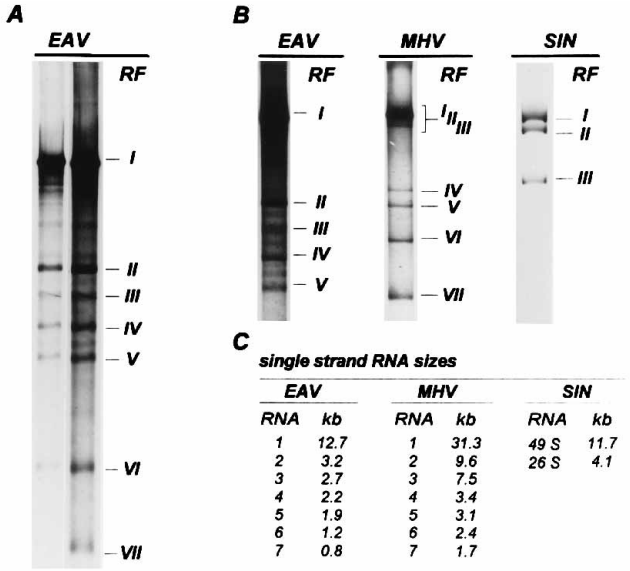

FIG. 4. RF RNA. (A) The RNA in EAV-infected BHK-21 cells was metabolically labeled between 1.5 and 10 h postinfection. After lysis and DNase I and RNase A and T_1 treatment, RF RNA prepared from 3×10^6 cells was resolved in a nondenaturing agarose gel. Two different exposures of the same lane are shown. (B) EAV RF RNAs I to V resolved in the same gel as RF RNAs prepared from 5×10^5 MHV- or SIN-infected cells, which were labeled between 1.5 and 7 h postinfection. (C) Lengths of the individual single-stranded RNA species which are present in EAV-, MHV-, or SIN-infected cells. RNA sizes were obtained from reference 6 and from references cited in references 31 and 33.

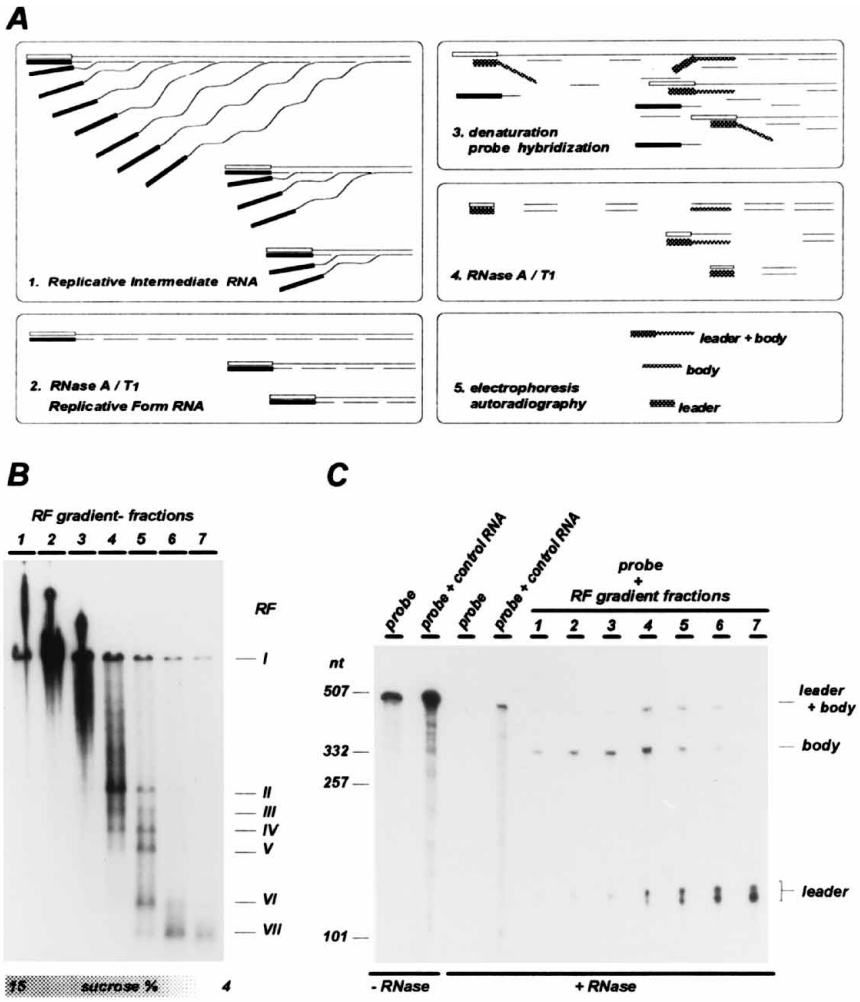

FIG. 5. Analysis of negative-stranded RNA 2 by RNase protection. (A) Schematic diagram of the procedure which was followed. Solid symbols indicate positive-stranded RNA; open symbols represent minus strands. The positive-stranded in vitro-synthesized probe transcript is indicated by hatching. (B) Metabolically labeled RF RNA, prepared from lysates of EAV-infected cells, was size fractionated in a sucrose gradient and analyzed by nondenaturing agarose gel electrophoresis and autoradiography. (C) RNase protection assays on the unlabeled equivalents of the RF RNA gradient fractions shown in panel B. The ^{32}P-labeled probe transcript containing positive-polarity sequences derived from the leader-body junction region of sg RNA 2 is shown in the leftmost lane. The next three control lanes show that an unlabeled transcript, containing the same and extra flanking viral sequences in negative polarity, can, upon incubation with RNase A and RNase T_1, fully protect the viral sequences in the probe. In the seven RF fractions (lanes 1 to 7), protected probe fragments are detected which contain either leader (133-nt) or RNA 2 body (342-nt) sequences, or both (470 nt), indicative of the presence of full-length minus-strand sg RNA 2 in RF II. RNA size markers are indicated on the left.

cultures of infected cells, one of which had been metabolically labeled with $H_3{}^{32}PO_4$ from 5.5 to 10 h postinfection. The RF RNAs were size fractionated by sucrose gradient ultracentrifugation. Figure 5B shows a direct analysis of the labeled gradient fractions in a nondenaturing agarose gel. This confirmed once more that seven RF RNAs could be prepared from EAV-infected cells. Corresponding fractions from the gradient containing the unlabeled RFs were used in the RNase protection assay. During this analysis, we focused on the RF RNA which corresponded to the largest sg RNA, RNA 2.

The double-stranded RNA was denatured and hybridized to an RNA probe which contained 128 nt of the leader, the 5′ UCAAC 3′ sequence, 337 nt of the sg mRNA 2 body sequence, and some flanking vector-derived sequences. The result of the assay is shown in Fig. 5C. The lengths of the various protected RNA fragments were compared with those of in vitro-transcribed RNA size markers. It was concluded that the complete mRNA 2-specific sequence in the probe (469 nt) could be protected from RNase degradation especially by RNA from those gradient fractions which were enriched for RF II (Fig. 5B). This demonstrated that RF II most likely contained a full-length complementary strand of sg RNA 2, including an antileader sequence. Protection of parts of the probe could be achieved with the RNAs from all gradient fractions, which was in agreement with the facts that the anti-RNA 2 body sequence is also present in the genome-length minus strands and that there is a complementary leader sequence in all seven minus-strand RNAs. A 342-nt RNA 2 body sequence (including the 5′ UCAAC 3′ sequence) was protected by the RNA from fractions containing RFs I and II. This fragment was expected to arise from hybridization of the probe to the complementary RNA 2 body sequence which is part of the genome-length minus strand of RF I. Alternatively, it could originate from binding of the probe to RF II-derived minus strands that do not contain an antileader sequence, a possibility that could not be investigated with this experimental setup. Two bands were detected in the 125- to 135-nt size range, where the 133-nt probe fragment (including the 5′ UCAAC 3′ sequence) that

could be protected by the antileader sequence was expected. On the basis of the size markers, we assume that the upper band is the protected leader fragment. So far, we can not explain the presence of the lower band, which was especially prominent when fractions containing the smallest RF RNAs were used in the assay.

DISCUSSION

By performing UV transcription mapping analyses, we have recently demonstrated that EAV sg RNA synthesis involves a transcription process which most likely is similar to that of the coronaviruses (7). The generation of the sg mRNAs of coronaviruses which contain a 5′ common leader sequence does not involve conventional *cis* splicing but involves discontinuous transcription. This process may take place during either plus-strand or minus-strand transcription (for reviews, see references 13, 31, and 37). The ability of the polymerase to terminate and reinitiate transcription during RNA synthesis is the key feature in most sg RNA transcription models that have been proposed. By imposing a regulatory mechanism, such that transcription reinitiation occurs at preferential positions, these viruses are thought to utilize this polymerase ability to direct the synthesis of sg mRNAs. The complementary sequences of the arterivirus leader-body junction sites and those of the so-called intergenic sequences of coronaviruses probably specify such positions and can therefore be regarded as promoter sequences for subgenomic plus-strand transcription. A model in which coronavirus intergenic sequences play an additional role as attenuators of sg minus-strand transcription has also been proposed (37).

Base pairing between the 3′ end of the leader sequence and the position where reinitiation of transcription takes place may facilitate discontinuous transcription by aligning the leader-replicase complex along the sg promoter sequence. Alternatively, if it can occur in the absence of the replicase complex, base pairing may even be the initial step in sg RNA transcription. In that case, its function is to facilitate priming of transcription by leader transcripts, which are subsequently elongated by the incoming polymerase. In the case of EAV, base pairing possibilities are present because of the conservation of a junction site sequence at positions upstream of all ORFs downstream of ORF 1. The consensus sequence found at the junctions in the sg mRNAs was shown to be 5′ UCAACu 3′ (Fig. 3). For other arteriviruses, similar sequence elements have been described: the sequences 5′ UnUAACC 3′ (LDV) (3) and 5′ UnAACC 3′ (PRRSV [16] and simian hemorrhagic fever virus [38]) were identified as junction site sequence motifs. For EAV, the only deviation from the 5′ UCAACu 3′ consensus was seen in sg RNA 6, which contained a C instead of a U at the 3′ position. Examination of the 11 remaining 5′ UCAAC 3′ pentanucleotide sequences found in the genomic sequence (Fig. 1A) (6) revealed that three of these are also followed by a U. Thus, although the consensus EAV junction motif displays a bias towards 5′ UCAACU 3′, the demand for additional determinants of sg promoter activity remains. So far, no common primary or secondary sequence elements have been detected in the sequences flanking the leader-body junction sites.

Upstream of four ORFs, multiple 5′ UCAAC 3′ (or even 5′ UCAACU 3′) motifs are present, and it was anticipated that additional RNA species could be synthesized as a result. However, no indications were found for the use of more than one junction site for the transcription of RNAs 2, 4, 5, and 7. Surprisingly, the use of alternative leader-body junction sites was observed in the case of sg mRNA 3. Only one of the sg mRNA 3 leader-body junction sites is of the consensus type 5′ UCAACu 3′ (Fig. 3, RNA 3 junction types 3.2). The most upstream one (type 3.1) lacks the C at position 5, but five additional nucleotides (5′ UACCC 3′) immediately downstream from this nucleotide match with those found in the 3′ end of the genomic leader sequence. This results in extended base pairing possibilities, as shown in Fig. 6. The downstream junction site found for EAV sg RNA 3 (Fig. 3, type 3.3), which on the basis of hybridization results (data not shown) and the number of cDNA clones found is likely to represent only a minor RNA 3 subpopulation, is even less conserved. Again, however, this seems to be compensated for by a significant flanking sequence complementarity (Fig. 6). The junction site used for the most abundant sg mRNA, RNA 7, also shows extra matching sequences at this position. These observations suggest that the extent of sequence complementarity between the 3′ end of the leader RNA and the complementary junction sites in the antisense template plays a role in the process of leader-to-body joining. In terms of this base pairing, the deviation of the sg mRNA 6 junction site from the 5′ UCAACu 3′ consensus might be tolerated because of the fact that instead of an A · U base pair, a G · U base pair can still be formed (Fig. 6). On the other hand, the 5′ UCAAC 3′ motif located at position 11 of the genomic map in Fig. 1A does contain three 3′ extending base pairing possibilities (6) but still is not used as a leader-body junction site.

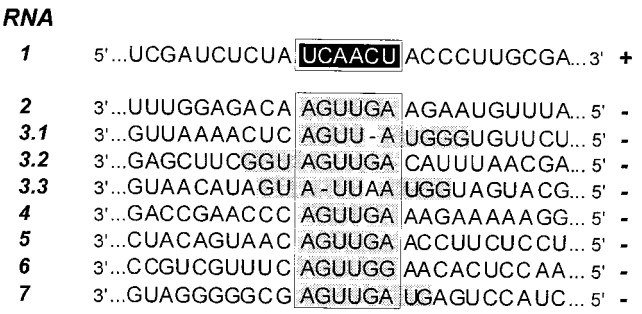

FIG. 6. Possibilities of base pairing between the 3′ end of the positive-strand EAV leader, as it is present in the genomic RNA, and the complementary leader-body junction sites for each of the sg RNAs. The conserved sequence elements which are thought to be involved in discontinuous transcription are boxed. Shading indicates base pairing possibilities.

The occurence of leader-body junctions like those found in sg mRNAs 3.1, 3.2, and 3.3 raises an additional point of interest. The leader-derived sequences which eventually are present in these mRNAs do not include the entire base pairing region. If base pairing between leader transcripts and the sg promoters on the minus-strand template was indeed involved in their generation, a regressive polymerase or RNase activity, which removes 3′ sequences from the leader transcript, must have preceded the reinitiation of transcription. This suggests that base pairing functions only to initially align the sequences and becomes superfluous as soon as the reinitiating replicase is positioned correctly. It also raises the question of whether the length of priming leader transcripts extends beyond the conserved junction site sequence. The heterogeneity in the type 3.2 leader-body junctions of the EAV mRNA 3 molecules resembles the sequence variation found in the junctions of the LDV and PRRSV sg mRNAs (3, 16) and can be explained by a mechanism using the aforementioned nuclease activity. Several observations which support a similar mechanism have been made for coronavirus sg mRNA transcription (1, 14, 26, 36).

Since the presence of negative-stranded counterparts for genomic and sg RNAs in coronavirus-infected cells has been established (24, 25), it has become feasible that discontinuous transcription can occur during plus-strand and/or minus-strand synthesis. For the arteriviruses, genome-length minus-strand RNA has been detected in LDV-infected cells (2). By performing PCR to analyze intracellular negative-stranded RNA, Chen et al. obtained indications that sg minus strands containing antileader sequences were present as well (2). Mainly because of their apparent extremely low abundance, the functionality of these sg minus strands during LDV sg mRNA transcription was doubted by those authors.

In this report, we have shown that seven RF RNAs could be prepared from EAV-infected cells (Fig. 4A). In addition to the seven RF RNAs, a number of minor bands were present. The most likely explanation for their appearance was that the EAV RI RNAs contained RNase-sensitive sites (other than the single-stranded parts of the nascent plus strands). A second possibility was that, because of the nondenaturing conditions during electrophoresis, the extra double-stranded RNA species represented the same seven RF RNA molecules which displayed differential migrational behavior. Alternatively, they could have arisen as a result of partial denaturation of the RF RNAs, although in that case comigration with the single-stranded plus strands would have been expected, which was never observed (data not shown).

The RNase protection experiments showed that sequences which were complementary to the leader-body junction region of sg mRNA 2 were present in the EAV RF RNA preparation (Fig. 5). The fact that these were present in RF II as a continuous stretch of RNA rules out the possibility that this sg RF RNA originated from digestion of an RI that contained a genome-length minus strand. Theoretically, the protecting minus strands could have been nascent transcripts in RF RNAs which contained a positive-stranded transcription template. However, the appearance of a single, uniformly sized RNase protection fragment contradicts this hypothesis. Therefore, RF II must have contained a full-length sg RNA 2 minus strand, which implies that sg minus strands can be used as templates for positive-strand transcription, as has been postulated for coronaviruses (22, 25). The analysis of the transcriptional activities of a number of MHV *ts* mutants has recently supported this assumption, by showing that sg minus strands are likely to be used as transcription templates late in infection (23). As an alternative possibility, it has been postulated that the coronavirus sg minus-strand RNAs are dead-end products which result from negative-strand synthesis with positive-strand sg mRNAs as a template (12).

In view of our results, it can be expected that minus strands complementary to each of the other EAV sg mRNAs can be found in the corresponding RF RNAs. Therefore, a nested set of RIs containing minus strands with common 5′- and 3′-terminal sequences must be present in EAV-infected cells. The existence of minus strands without an antileader sequence cannot be excluded on the basis of our results and remains an intriguing possibility. Once more, our results underline the previously established evolutionary relationship between the arteri- and coronaviral replicases (6, 28), which apparently is linked to similarities between the transcriptional mechanisms used by these two virus groups to generate sg mRNAs.

ACKNOWLEDGMENTS

We very much appreciated the stimulating discussions with Dorothea and Stanley Sawicki. We are grateful to Derrick Louz for technical assistance. We also thank Guido van Marle, Leonie van Dinten, and Fred Wassenaar for helpful comments and Twan de Vries for providing cDNA clones of sg RNAs 6 and 7.

REFERENCES

1. **Baker, S. C., and M. M. C. Lai.** 1990. An in vitro system for the leader-primed transcription of coronavirus mRNAs. EMBO J. **9:**4173–4179.
2. **Chen, Z., K. S. Faaberg, and P. G. W. Plagemann.** 1994. Detection of negative-stranded subgenomic RNAs but not of free leader in LDV-infected macrophages. Virus Res. **34:**167–177.
3. **Chen, Z., L. Kuo, R. R. R. Rowland, C. Even, K. S. Faaberg, and P. G. W. Plagemann.** 1993. Sequence of 3′-end of genome and 5′-end of ORF 1a of lactate dehydrogenase-elevating virus (LDV) and common junction motifs between 5′ leader and bodies of seven subgenomic mRNAs. J. Gen. Virol. **74:**643–660.
4. **Conzelmann, K. K., N. Visser, P. V. Woensel, and H. J. Thiel.** 1993. Molecular characterization of porcine reproductive and respiratory syndrome virus, a member of the arterivirus group. Virology **193:**329–339.

4a. **den Boon, J. A.** Unpublished results.

5. **den Boon, J. A., K. S. Faaberg, J. J. M. Meulenberg, A. L. M. Wassenaar, P. G. W. Plagemann, A. E. Gorbalenya, and E. J. Snijder.** 1995. Processing and evolution of the N-terminal region of the arterivirus replicase ORF1a protein: identification of two papainlike cysteine proteases. J. Virol. **69:** 4500–4505.
6. **den Boon, J. A., E. J. Snijder, E. D. Chirnside, A. A. F. de Vries, M. C. Horzinek, and W. J. M. Spaan.** 1991. Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. J. Virol. **65:**2910–2920.
7. **den Boon, J. A., W. J. M. Spaan, and E. J. Snijder.** 1995. Equine arteritis virus (EAV) subgenomic RNA transcription: UV inactivation and translation inhibition studies. Virology **213:**364–372.
8. **de Vries, A. A. F., E. D. Chirnside, P. J. Bredenbeek, L. A. Gravestein, M. C. Horzinek, and W. J. M. Spaan.** 1990. All subgenomic mRNAs of equine arteritis virus contain a common leader sequence. Nucleic Acids Res. **18:** 3241–3247.
9. **de Vries, A. A. F., E. D. Chirnside, M. C. Horzinek, and P. J. M. Rottier.** 1992. Structural proteins of equine arteritis virus. J. Virol. **66:**6294–6303.
10. **Godeny, E. K., L. Chen, S. N. Kumar, S. L. Methven, E. V. Koonin, and M. A. Brinton.** 1993. Complete genomic sequence and phylogenetic analysis of the lactate dehydrogenase-elevating virus. Virology **194:**585–596.
11. **Godeny, E. K., L. Zeng, S. L. Smith, and M. A. Brinton.** 1995. Molecular characterization of the 3′ terminus of the simian hemorrhagic fever virus genome. J. Virol. **69:**2679–2683.
12. **Jeong, Y. S., and S. Makino.** 1992. Mechanism of coronavirus transcription: duration of primary transcription initiation activity and effects of subgenomic RNA transcription on RNA replication. J. Virol. **66:**3339–3346.
13. **Lai, M. M. C.** 1990. Coronavirus—organization, replication and expression of genome. Annu. Rev. Microbiol. **44:**303–333.
14. **Makino, S., L. H. Soe, C. K. Shieh, and M. M. C. Lai.** 1988. Discontinuous transcription generates heterogeneity at the leader fusion sites of coronavirus mRNAs. J. Virol. **62:**3870–3873.
15. **Meinkoth, J., and G. Wahl.** 1984. Hybridization of nucleic acids immobilized on solid supports. Anal. Biochem. **138:**267–284.
16. **Meulenberg, J. J. M., E. J. de Meijer, and R. J. M. Moormann.** 1993. Subgenomic RNAs of Lelystad virus contain a conserved leader-body junction sequence. J. Gen. Virol. **74:**1697–1701.
17. **Meulenberg, J. J. M., M. M. Hulst, E. J. de Meijer, P. L. J. M. Moonen, A. den Besten, E. P. de Kluyver, G. Wensvoort, and R. J. M. Moormann.** 1993. Lelystad virus, the causative agent of porcine epidemic abortion and respiratory syndrome (PEARS), is related to LDV and EAV. Virology **192:**62–72.
18. **Novak, J. E., and K. Kirkegaard.** 1991. Improved method for detecting poliovirus negative strands used to demonstrate specificity of positive-strand encapsidation and the ratio of positive to negative strands in infected cells. J. Virol. **65:**3384–3387.
19. **Plagemann, P. G. W., and V. Moennig.** 1992. Lactate dehydrogenase-elevating virus, equine arteritis virus and simian haemorrhagic fever virus, a new group of positive strand RNA viruses. Adv. Virus Res. **41:**99–192.
20. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
21. **Sawicki, D. L., and S. G. Sawicki.** 1987. Alphavirus plus and minus strand RNA synthesis, p. 251–259. *In* M. Brinton and R. Ruckert (ed.), Positive-strand RNA viruses. Alan R. Liss, Inc., New York.
22. **Sawicki, S. G., and D. L. Sawicki.** 1990. Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. J. Virol. **64:**1050–1056.
23. **Schaad, M. C., and R. S. Baric.** 1994. Genetics of mouse hepatitis virus transcription: evidence that subgenomic negative strands are functional templates. J. Virol. **68:**8169–8179.
24. **Sethna, P. B., N. A. Hofmann, and D. A. Brian.** 1991. Minus-strand copies of replicating coronavirus mRNAs contain antileaders. J. Virol. **65:**320–325.
25. **Sethna, P. B., S. L. Hung, and D. A. Brian.** 1989. Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. Proc. Natl. Acad. Sci. USA **86:**5626–5630.

26. **Shieh, C. K., L. H. Soe, S. Makino, M. F. Chang, S. A. Stohlman, and M. M. C. Lai.** 1987. The 5′ end sequence of the murine coronavirus genome: implications for multiple fusion sites in leader-primed transcription. Virology **156:**321–330.
27. **Simmons, D. T., and J. H. Strauss.** 1972. Replication of Sindbis virus II. Multiple forms of double-stranded RNA isolated from infected cells. J. Mol. Biol. **71:**615–631.
28. **Snijder, E. J., and W. J. M. Spaan.** 1995. The coronaviruslike superfamily, p. 239–255. *In* S. G. Siddell (ed.), The Coronaviridae. Plenum Press, New York.
29. **Snijder, E. J., A. L. M. Wassenaar, and W. J. M. Spaan.** 1994. Proteolytic processing of the replicase ORF1a protein of equine arteritis virus. J. Virol. **68:**5755–5764.
30. **Snijder, E. J., A. L. M. Wassenaar, W. J. M. Spaan, and A. E. Gorbalenya.** 1995. The arterivirus nsp2 protease: an unusual cysteine protease with primary structure similarities to both papain-like and chymotrypsin-like proteases. J. Biol. Chem. **270:**16671–16676.
31. **Spaan, W. J. M., D. Cavanagh, and M. C. Horzinek.** 1988. Coronaviruses: structure and genome expression. J. Gen. Virol. **69:**2939–2952.
32. **Spaan, W. J. M., P. J. M. Rottier, M. C. Horzinek, and B. A. M. van der Zeijst.** 1981. Isolation and identification of virus-specific mRNAs infected with mouse hepatitis virus (MHV-A59). Virology **108:**424–434.
33. **Strauss, E. G., C. M. Rice, and J. H. Strauss.** 1984. Complete nucleotide sequence of the genomic RNA of Sindbis virus. Virology **133:**92–110.
34. **Strauss, J. H., and E. G. Strauss.** 1994. The alphaviruses: gene expression, replication, and evolution. Microbiol. Rev. **58:**491–562.
35. **Van Berlo, M. F., M. C. Horzinek, and B. A. M. van der Zeijst.** 1982. Equine arteritis virus-infected cells contain six polyadenylated virus-specific RNAs. Virology **118:**345–352.
36. **van der Most, R. G., R. J. de Groot, and W. J. M. Spaan.** 1994. Subgenomic RNA synthesis directed by a synthetic defective interfering RNA of mouse hepatitis virus: a study of coronavirus transcription initiation. J. Virol. **68:** 3656–3666.
37. **van der Most, R. G., and W. J. M. Spaan.** 1995. Coronavirus replication, transcription, and RNA recombination, p 11–31. *In* S. G. Siddell (ed.), The Coronaviridae. Plenum Press, New York.
38. **Zeng, L., E. K. Godeny, S. L. Methven, and M. A. Brinton.** 1995. Analysis of simian hemorrhagic fever virus (SHFV) subgenomic RNAs, junction sequences, and 5′ leader. Virology **207:**543–548.