The EMBO Journal Vol. 20 No. 24 pp. 7220–7228, 2001

Sequence requirements for RNA strand transfer during nidovirus discontinuous subgenomic RNA synthesis

Alexander O.Pasternak, Erwin van den Born, Willy J.M.Spaan and Eric J.Snijder[1]

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

[1]Corresponding author
e-mail: e.j.snijder@lumc.nl

A.O.Pasternak and E.van den Born contributed equally to this work

Nidovirus subgenomic mRNAs contain a leader sequence derived from the 5′ end of the genome fused to different sequences ('bodies') derived from the 3′ end. Their generation involves a unique mechanism of discontinuous subgenomic RNA synthesis that resembles copy-choice RNA recombination. During this process, the nascent RNA strand is transferred from one site in the template to another, during either plus or minus strand synthesis, to yield subgenomic RNA molecules. Central to this process are transcription-regulating sequences (TRSs), which are present at both template sites and ensure the fidelity of strand transfer. Here we present results of a comprehensive co-variation mutagenesis study of equine arteritis virus TRSs, demonstrating that discontinuous RNA synthesis depends not only on base pairing between sense leader TRS and antisense body TRS, but also on the primary sequence of the body TRS. While the leader TRS merely plays a targeting role for strand transfer, the body TRS fulfils multiple functions. The sequences of mRNA leader–body junctions of TRS mutants strongly suggested that the discontinuous step occurs during minus strand synthesis.

Keywords: arterivirus/nidovirus/RNA recombination/subgenomic mRNA synthesis/transcription-regulating sequences

Introduction

The genetic information of RNA viruses is organized very efficiently. Practically every nucleotide of their genome is utilized, either as protein-coding sequence or as *cis*-acting signals for translation, RNA synthesis or RNA encapsidation. As part of their genome expression strategy, several groups of positive-strand RNA (+RNA) viruses produce subgenomic (sg) mRNAs (reviewed by Miller and Koev, 2000). The replication of their genomic RNA, which is also the mRNA for the viral replicase, is supplemented with the generation of sg transcripts to express structural and auxiliary proteins, which are encoded downstream of the replicase gene in the genome. Sg mRNAs of +RNA viruses are always 3′-co-terminal with the genomic RNA, but different mechanisms are used for their synthesis. Some viruses, such as brome mosaic virus, initiate sg mRNA synthesis internally on the full-length minus strand RNA template (Miller *et al*., 1985). Others, exemplified by red clover necrotic mosaic virus (RCNMV), may rely on premature termination of minus strand synthesis from the genomic RNA template, followed by the synthesis of sg plus strands from the truncated minus strand template (Sit *et al*., 1998). Members of the order Nidovirales, which includes coronaviruses and arteriviruses, have evolved a third and unique mechanism, which employs discontinuous RNA synthesis for the generation of an extensive set of sg RNAs (reviewed by Brian and Spaan, 1997; Lai and Cavanagh, 1997; Snijder and Meulenberg, 1998). Nidovirus sg mRNAs differ fundamentally from other viral sg RNAs in that they are not only 3′-co-terminal, but also 5′-co-terminal with the genome (Figure 1A). A 5′ common leader sequence of 65–221 nucleotides, derived from the 5′ end of the genomic RNA, is attached to the 3′ part of each sg RNA (the 'mRNA body').

Various models have been put forward to explain the co-transcriptional fusion of non-contiguous parts of the nidovirus genome during sg RNA synthesis (Figure 1B and C). Central to each of these models are short transcription-regulating sequences (TRSs), which are present both at the 3′ end of the leader and at the 5′ end of the sg RNA body regions in the genomic RNA. The TRS is copied into the mRNA and connects its leader and body part (Spaan *et al*., 1983; Lai *et al*., 1984). Synthesis of sg mRNAs initially was proposed to be primed by free leader transcripts, which would base-pair to the complementary TRS regions in the full-length minus strand, and would be extended subsequently to make sg plus strands (Figure 1B; Baric *et al*., 1983, 1985). This model, however, was based on the report that sg minus strands were not present in coronavirus-infected cells (Lai *et al*., 1982). The subsequent discovery of such molecules (Sethna *et al*., 1989) resulted in reconsideration of the initial 'leader-primed transcription' model. Sawicki and Sawicki (1995) have proposed an alternative model (Figure 1C), in which the discontinuous step occurs during minus instead of plus strand RNA synthesis. In this model, minus strand synthesis would be attenuated after copying a body TRS from the plus strand template. Next, the nascent minus strand, with the TRS complement at its 3′ end, would be transferred to the leader TRS and attach by means of TRS–TRS base pairing. RNA synthesis would be reinitiated to complete the sg minus strand by adding the complement of the genomic leader sequence. Subsequently, the sg minus strand would be used as template for sg mRNA synthesis, and the presence of the leader complement at its 3′ end might allow the use of the same RNA signals that direct genome synthesis from the full-length minus strand.

© European Molecular Biology Organization

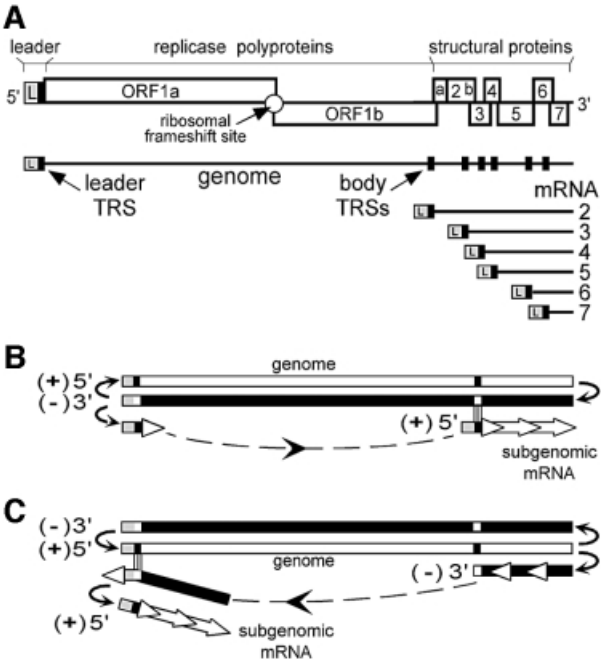

Fig. 1. (**A**) Schematic diagram of the genome organization and expression of EAV, the arterivirus prototype. The regions of the genome specifying the leader (L) sequence, the replicase gene (ORFs 1a and 1b) and the structural genes are indicated. The nested set of EAV mRNAs (genome and sg mRNAs 2–7) is depicted below. The black boxes in the genomic RNA indicate the position of leader and major body TRSs. (**B** and **C**) Alternative models for nidovirus discontinuous sg RNA synthesis. The discontinuous step may occur during either plus strand (B) or minus strand (C) RNA synthesis. In the latter case, sg mRNAs would be synthesized from an sg minus strand template. For details see text.

Using site-directed mutagenesis of TRSs of the arterivirus equine arteritis virus (EAV), we have shown previously that base pairing between the sense leader TRS and antisense body TRSs is crucial for sg mRNA synthesis (van Marle *et al*., 1999a). However, base pairing is only one step of the nascent strand transfer process and is essential in both models outlined in Figure 1. The EAV genomic RNA contains several sequences that match the leader TRS precisely, but nevertheless are not used for sg RNA synthesis (den Boon *et al*., 1996; Pasternak *et al*., 2000). This suggests that leader–body TRS similarity alone is, though necessary, not sufficient for the strand transfer to occur.

To gain further insight into the *cis*-acting signals regulating sg RNA synthesis, we performed a comprehensive site-directed mutagenesis study of the EAV leader and body TRSs. Every nucleotide of the TRS (5′-UCAACU-3′) was substituted with each of the three alternative nucleotides. Our analysis revealed a number of striking similarities with the process of copy-choice RNA recombination, as it occurs in RNA viruses. Whereas the leader TRS plays only a targeting role in translocation of the nascent strand, body TRS nucleotides appear to fulfil diverse position-specific and base-specific functions. In addition, the sequence of the leader–body junctions of the sg mRNAs produced by these mutants provided strong evidence for the discontinuous minus strand extension model.

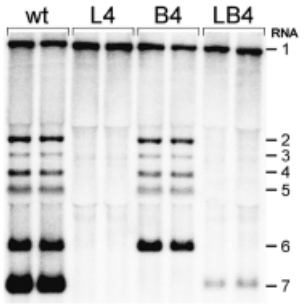

Fig. 2. Northern analysis of EAV-specific RNA isolated from cells transfected with RNA transcribed either from the wild-type EAV infectious cDNA clone or from TRS pentanucleotide mutants (UCAAC to AGUUG). The results of two independent experiments are shown.

Results and discussion

EAV genome replication is not significantly affected by leader TRS and body TRS mutations

To dissect EAV RNA synthesis, we routinely use a full-length cDNA clone (van Dinten *et al*., 1997), from which infectious EAV RNA is *in vitro* transcribed. Following transfection of the RNA into baby hamster kidney (BHK-21) cells, intracellular RNA is isolated and analysed by northern blot hybridization and RT–PCR (van Marle *et al*., 1999a). Due to differences in transfection efficiency, the total amount of virus-specific RNA (genomic RNA and sg mRNA) isolated from transfected cell cultures is somewhat variable. Thus, the accurate quantitation of sg mRNA synthesis by TRS mutants requires an internal standard for transfection efficiency. The amount of viral genomic RNA can be this standard, but only if its amplification is not dramatically affected by the TRS mutations. To prove that this is the case, we used the previously described mutants L4, B4 and LB4 (van Marle *et al*., 1999a), in which five nucleotides of the TRS (5′-UCAAC-3′) were replaced by the sequence 5′-AGUUG-3′, either in the leader TRS (L4), RNA7 body TRS (B4) or both TRSs (LB4).

The three mutants were tested in three independent experiments. Intracellular RNA was isolated at 14 h post-transfection, early enough to prevent spread of the wild-type control virus to non-transfected cells (first cycle analysis). Transfection efficiencies were determined by immunofluorescence assays (see Materials and methods) and varied between 10 and 23% (data not shown). Prior to RNA analysis, the amount of isolated intracellular RNA was corrected for the transfection efficiency of the sample, so that each lane in Figure 2 represents EAV-specific RNA from an approximately equal number of EAV-positive cells. Phosphoimager quantitation revealed that genomic RNA replication of mutants L4, B4 and LB4 varied by not more than 30% (Table I). These differences could reflect, for example, a slight influence of RNA secondary structure changes in the TRS regions on genomic RNA synthesis. Remarkably, however, the genomic RNA level of the leader–body TRS double mutant LB4 was not affected by more than 10%. In view of the results obtained with these pentanucleotide TRS mutants, we assumed that the amount of genomic RNA could indeed be used as an internal standard during the analysis of mutants containing only single nucleotide replacements in leader TRS and/or RNA7 body TRS.

The RNA–RNA interaction between the leader and body TRSs is not the only factor that regulates EAV sg RNA synthesis

There are numerous examples of regulatory RNA–RNA interactions in both eukaryotic and prokaryotic cells, as well as in RNA viruses. Essential processes such as translation, replication and encapsidation of RNA virus genomes frequently depend on RNA–RNA interactions and higher order RNA structures. Regulation of sg RNA synthesis of +RNA viruses by RNA–RNA interactions is also not without precedent. In tomato bushy stunt virus, an RNA element located 1000 nucleotide upstream of the sg RNA2 promoter base-pairs with the promoter and is necessary for sg RNA production (Zhang *et al*., 1999). Similarly, base pairing interactions between complementary sequences in the 5′ end of the potato virus X genomic RNA and sequences upstream of two major sg RNA promoters are required for efficient sg RNA synthesis (Kim and Hemenway, 1999). In RCNMV, an intermolecular RNA–RNA interaction is required for sg RNA synthesis (Sit *et al*., 1998).

Table I. Effect of pentanucleotide TRS mutations on genomic RNA synthesis

Mutant	Genomic RNA synthesis (% of wild type)
L4	111 ± 12
B4	72 ± 14
LB4	89 ± 2

Recently, we have established the pivotal role of an interaction between sense and antisense RNA sequences in the life cycle of EAV (van Marle *et al*., 1999a). In that study, the role of TRS nucleotides C_2 and C_5 was tested by substituting them with G. It was concluded that base pairing between the sense leader TRS and the antisense body TRS plays a crucial role in nidovirus sg RNA synthesis. We now took a more systematic approach and performed an extensive site-directed co-variation mutagenesis study of the entire leader TRS and RNA7 body TRS, which directs the synthesis of the most abundant EAV sg RNA. Every nucleotide of the TRS (5′-UCA-ACU-3′) was replaced with each of the other possible nucleotides. As in the study of van Marle *et al*. (1999a), every mutation was introduced into leader TRS, RNA7 body TRS and both TRSs, resulting in 54 mutant constructs. Each mutant was given a unique name: e.g. BU_1A refers to a mutant in which a U has been changed to A at position 1 of the body TRS; LU_1A refers to the same substitution in the leader TRS; and DU_1A means that these two substitutions were combined in one double mutant construct. The amount of sg RNA7 was quantitated by phosphoimager scanning of hybridized gels and was corrected for the amount of genomic RNA in the same lane (as outlined above). Figure 3 shows the relative sg RNA7 level of the 54 mutants, compared with the RNA7 level of the wild-type control. For a selection of 11 interesting mutants (see below), the analysis was repeated three times (Figure 4), without observing significant variations in sg RNA synthesis.

The comprehensive analysis of the effects of TRS mutations considerably expanded our understanding of

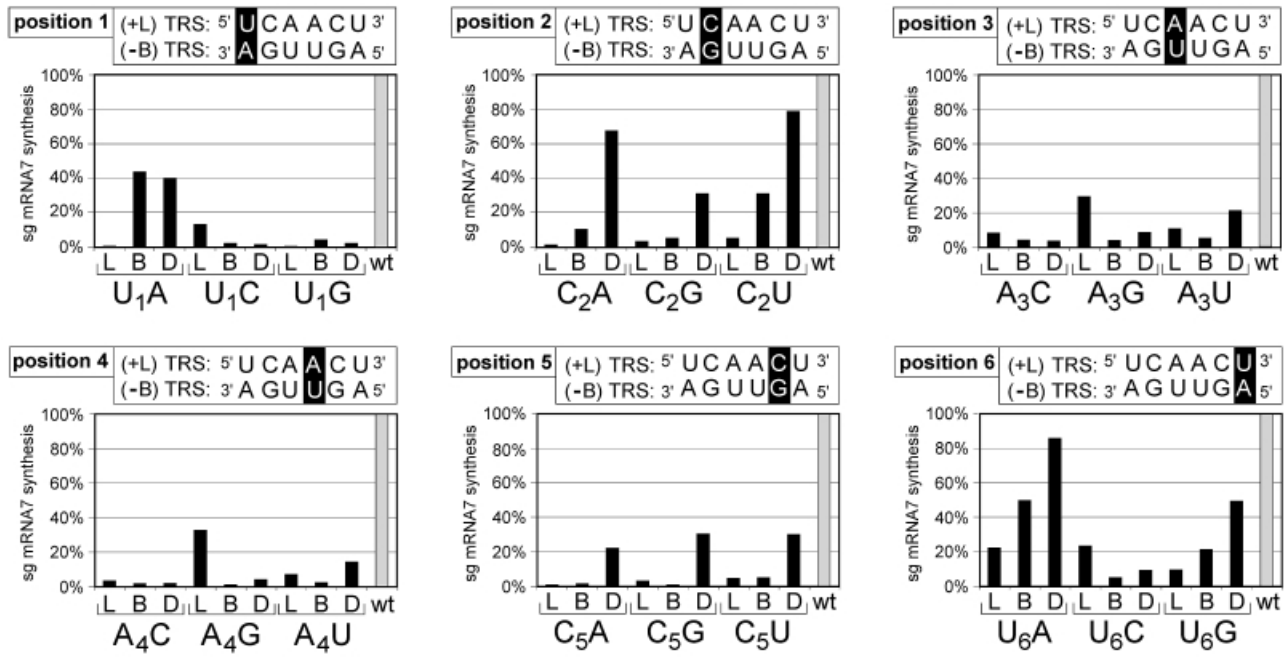

Fig. 3. Relative amounts of sg RNA7 produced by EAV TRS mutants. For each set of transfections (representing one experiment), the EAV wild-type construct pEAV030 (van Dinten *et al*., 1997) was taken along as a positive control. For every mutant, the level of sg RNA7 synthesis was calculated as $[(sg/g)/(sg/g)_{wt}] \times 100\%$: it was corrected for the level of genomic RNA (used as an internal standard; see text) and subsequently was related to the level of sg RNA7 produced by the wild-type control in the same experiment, which was also corrected for the corresponding genomic RNA level. The relative sg RNA7 level of the wild-type control was set at 100%.

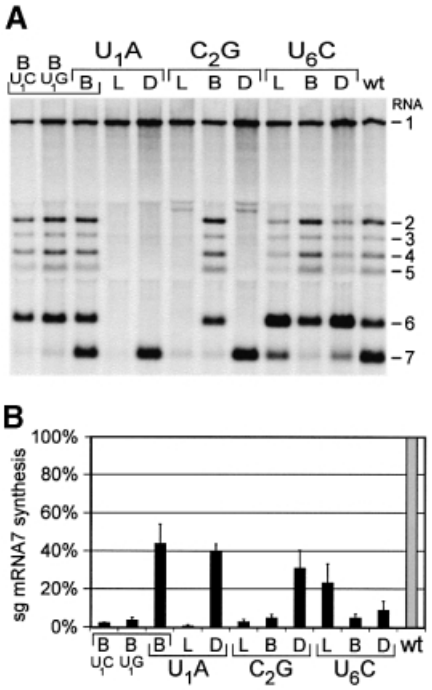

Fig. 4. (**A**) Northern analysis of EAV-specific RNA isolated from cells transfected with wild-type EAV RNA or 11 selected TRS mutants. (**B**) Relative amount of sg RNA7 produced by the selection of 11 TRS mutants. See the legend to Figure 3 for the calculation of sg RNA7 levels. For each construct in (B), experiments were repeated at least three times and average sg RNA levels are shown.

discontinuous sg RNA synthesis. Remarkably, the effects of single (leader or body) TRS mutations were mostly base specific, i.e. different nucleotide substitutions at the same position affected sg RNA7 synthesis to different extents. For example, at position 1, the BU_1A mutant retained 44% of the wild-type RNA7 synthesis level, whereas both the BU_1C and BU_1G mutants lost RNA7 synthesis almost completely. Conversely, when U_1 of the leader TRS was changed to A or G, RNA7 synthesis was completely abolished, whereas 13% of the wild-type level was still maintained by LU_1C. For position 2, only the BC_2U mutant retained 30% of the wild-type RNA7 synthesis level, while all the other position 2 single mutants have lost 90% or more of wild-type RNA7 synthesis. Another example is position 6: BU_6C left only 5% of wild-type RNA7 synthesis, whereas BU_6A produced much higher RNA7 levels. This implied that for some positions (1, 2 and 6), certain mismatches in the duplex between plus leader TRS and minus body TRS, such as U–U (BU_1A and BU_6A) or C–A (LU_1C and BC_2U), are allowed to a limited extent. In contrast, no mismatches were allowed for position 5, where all single nucleotide substitutions abolished RNA7 synthesis almost completely.

Surprisingly, both body TRS U to C substitutions at positions 1 and 6 (BU_1C and BU_6C) resulted in low levels of RNA7, despite the fact that these mutations allow the formation of a G–U base pair between the plus leader TRS, providing the U nucleotide, and the minus body TRS, providing the G. On the other hand, for positions 3 and 4, G–U base pairing was shown to be functional, because mutants LA_3G and LA_4G, which can form G–U base pairs between the G in the plus leader TRS and U in the minus body TRS, were the only position 3 and 4 single mutants that produced reasonable levels of RNA7. Taken together, these findings suggest that other factors, besides leader–body base pairing, also play a role in sg RNA synthesis and that the primary sequence (or secondary structure) of TRSs may dictate strong base preferences at certain positions. Our analysis of the degree of complementation by the double mutants provided strong support for this assumption.

Differentiating between effects at the level of primary TRS sequence and the level of leader–body duplex formation

For some TRS nucleotides (2, 5 and 6, except in the case of DU_6C), the RNA7 level of double mutants was clearly higher than that of the corresponding single mutants. This means that base pairing between these leader and body TRS nucleotides is involved in sg RNA synthesis. However, none of these double mutants reached the wild-type sg RNA7 level. In the other double mutants (all position 1, 3 and 4 mutants, and DU_6C), in clear contradiction to the predictions of the 'base pairing model', RNA7 synthesis was not significantly restored. Moreover, a comparison of the values for the B and D mutants in Figure 3 showed that, for almost all of these mutants (e.g. the position 1 mutants), the amount of sg RNA7 produced by the double mutant appeared to be limited by the level allowed by the body TRS mutation. Sometimes the RNA7 level of the double mutant was even less than that of the leader mutant (DU_1C, DA_3G, DA_4G or DU_6C). Clearly, for these substitutions, restoration of the possibilities for leader–body duplex formation did not restore sg RNA synthesis. Apparently this is because the effect of body TRS mutations at the level of primary sequence or secondary structure can be 'dominant' over the duplex-restoring effects of the double mutations.

Body TRS mutants thus fell into two distinct types, determined by the position and chemistry of the substitution. In mutants of the first type, sg RNA synthesis was impaired mainly because of the disruption of the leader–body TRS duplex. This effect could be compensated for by introduction of the corresponding mutation in the leader TRS and, in the double mutant, sg RNA synthesis was restored compared with the corresponding single mutants. In mutants of the second type, sg RNA synthesis was down-regulated as a consequence of both TRS duplex disruption and disruption of the primary sequence (or secondary structure) of the body TRS. Obviously, the latter effect could not be compensated for by mutating the leader TRS, and the corresponding double mutants did not show restoration of sg RNA synthesis.

In contrast to our findings with the body TRS mutants, we did not obtain leader TRS mutations that appeared to determine the level of sg RNA7 synthesis of the corresponding double mutant (Figure 3). Thus, effects of mutations in the leader TRS were not 'dominant' over the duplex-restoring effects of the double mutations, suggesting that they only affected duplex formation. This indicated that the leader TRS probably does not have an additional, sequence-specific function in sg RNA synthesis in addition to its participation in TRS–TRS base pairing. The fact that single leader TRS mutations at all six

positions severely repressed RNA7 synthesis indicated that base pairing of every TRS nucleotide contributes to sg RNA production. In this respect, it was significant that the two leader TRS mutants with the highest RNA7 levels, LA_3G and LA_4G, can form G–U base pairs to maintain the duplex.

The observation that leader TRS mutations could be 'rescued' by introducing complementary mutations in the body TRS, but that many body TRS mutations could not be 'rescued' by corresponding changes in the leader TRS, is clearly illustrated by the U_1A mutants. Due to the restoration of TRS base pairing possibilities, the RNA7 synthesis of double mutant DU_1A was significantly increased compared with that of LU_1A, but not above the level of BU_1A. Thus, restoration of the leader–body duplex in DU_1A exerted a clear effect on sg RNA7 production compared with LU_1A, but had no effect on sg RNA synthesis compared with BU_1A. This exemplified the dominant nature of a mutation in the primary sequence of a body TRS. In contrast, for instance, the BC_2U mutation probably affected duplex formation only, because RNA7 synthesis was restored almost to wild-type levels in the DC_2U double mutant.

These results indicate that there are strong base preference constraints for some body TRS positions. To interpret these base preferences accurately, it is necessary to limit the analysis to the double mutants only, because in these mutants the down-regulation of sg RNA synthesis was only due to the sequence changes in the body TRS, and not to the disruption of the leader–body TRS duplex. There were strict preferences for positions 1, 3 and 4 of the body TRS: at position 1, only the U to A substitution allowed for a significant RNA7 level (~40% of wild-type); and at positions 3 and 4, only the A to U mutants retained 15–20% of the wild-type level. For positions 2 and 5, the sequence constraints were less stringent (all substitutions allowed for >20% of wild-type level), but still only DC_2A and DC_2U reached >50%. At position 6 of the body TRS, only U to C was not allowed, whereas the other two double mutants still produced 50% or more of RNA7. In other words, the functional EAV RNA7 body TRS (based on the analysis of our single nucleotide substitutions) can be described as $U_1(C/u/a)_2A_3A_4C_5(U/a/g)_6$, with wild-type nucleotides shown in upper case and nucleotides that allowed for at least 50% of the wild-type RNA7 level shown in lower case. Remarkably, TRS nucleotides A_3, A_4 and C_5 are conserved in the TRSs of all other arteriviruses (Snijder and Meulenberg, 1998). Also the fact that DC_2U retained 80% of RNA7 synthesis corresponded nicely to the presence of a U at this position in other arteriviruses.

Until recently (Almazan *et al*., 2000; Thiel *et al*., 2001), infectious cDNA clones were lacking for coronaviruses. Consequently, most studies on coronavirus sg RNA synthesis were carried out using defective interfering (DI) RNAs. These replicons carried body TRSs from which moderate levels of sg mRNAs could be produced in the presence of helper virus. Using this system, Joo and Makino (1992) and van der Most *et al*. (1994) performed body TRS mutagenesis studies for the murine coronavirus (MHV). Joo and Makino systematically mutagenized the core of the MHV body TRS. In contrast to our results, they found that in only two of 21 body TRS mutants was sg RNA synthesis from the DI RNA genome abolished, whereas all others supported normal levels of sg RNA production. Thus, it is possible that the MHV TRS which was used in that study is more tolerant to single-nucleotide mismatches than the EAV sg RNA7 TRS.

In a similar study, van der Most *et al*. (1994) observed that U to C substitutions at positions 1 and 3 of the MHV body TRS, which maintained the duplex by changing a U–A base pair into a U–G base pair, reduced sg RNA levels more strongly than substitutions that disrupted the duplex (van der Most *et al*., 1994). This implies that, as in the case of EAV, leader–body TRS duplex formation is not the only factor that determines coronavirus sg RNA synthesis. However, because of the limitations of the DI RNA system, the leader TRS could not be mutagenized in these studies, and body TRS-specific effects could not be distinguished from effects at the level of leader–body duplex formation.

The discontinuous step in nidovirus sg RNA synthesis occurs during minus strand RNA synthesis

Due to recent studies of arterivirus and coronavirus sg RNA synthesis (van Marle *et al*., 1999a; Baric and Yount, 2000; Sawicki *et al*., 2001), the discontinuous minus strand extension model (Figure 1C) has been gaining more and more ground. This model predicts that the TRS-derived sequence that forms the leader–body junction in the sg mRNA is a copy of the body TRS, and not of the leader TRS. The leader-primed transcription model predicts the opposite (Figure 1B). Therefore, determining the origin of the leader–body junction of sg mRNAs would help to distinguish between the two models. However, in the wild-type situation, EAV leader and body TRSs are identical and consequently one cannot determine the origin of the sg mRNA leader–body junction. This problem could be overcome by tracing the mutations introduced in leader or RNA7 body TRS mutants, most of which retained part of their ability to produce mRNA7. In a previous study (van Marle *et al*., 1999a), we found that nucleotides 2 and 5 of the mRNA7 leader–body junction sequence were derived exclusively from the body TRS, and not from the leader TRS. This was shown by direct sequencing of RT–PCR products obtained from the residual mRNA7 produced by mutants BC_2G, LC_2G, BC_5G and LC_5G (van Marle *et al*., 1999a).

Using the same approach, we analysed mRNA7 from mutants BC_2A and BC_2U, and these transcripts also contained the mutated nucleotide derived from the body TRS (data not shown). Assuming that only one crossover event occurs during leader–body joining, we could thus map this crossover between positions –1 and +2 of the sg RNA junction sequence. This left the intriguing question of whether the crossover site could be mapped even more precisely. In other words, was nucleotide +1 of the junction sequence derived from the body TRS or the leader TRS?

Using the position 1 mutants described above, we could answer this question (Figure 5). The most striking result was that mRNA7 of mutants BU_1A, BU_1G and LU_1C contained exclusively the body TRS-derived nucleotide at position +1. Thus, for these mutants, the crossover site could be mapped precisely between TRS nucleotide positions –1 and +1, meaning that the complete leader–

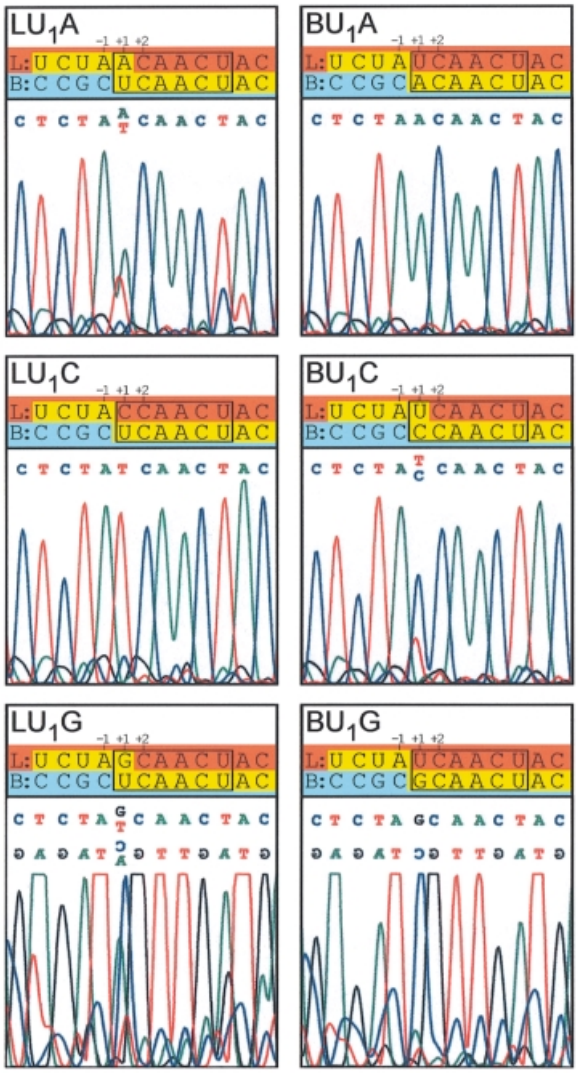

Fig. 5. Sequence analysis of mRNA7 leader–body junctions from position 1 TRS mutants. Sequences were determined directly from sg mRNA7-specific RT–PCR products. For the U_1A and U_1C mutants, the sequence shown corresponds to the plus strand of sg RNA7. For sequencing-related technical reasons, the minus strand sequence was determined for the U_1G mutants; a mirror image of the electropherogram is shown with the corresponding plus strand sequence listed at the top of the panel. For every mutant, a sequence alignment of the leader (red) and body (blue) TRSs and surrounding sequences is shown (TRSs are boxed). The mRNA7 leader–body junctions detected by our sequence analysis are shown in yellow.

body junction sequence in an EAV sg mRNA can be body TRS derived. On the other hand, sg RNAs from mutants LU_1A, BU_1C and LU_1G contained mixed populations of leader TRS- and body TRS-derived nucleotides at position +1 (Figure 5): A and U for LU_1A, C and U for BU_1C, and G and U for LU_1G. Remarkably, this pattern correlated with the relative amounts of sg mRNA7 produced by these mutants (Figure 3). Mutants that produced populations of sg RNAs that were mixed with respect to the origin of the nucleotide at position +1 of the leader–body junction had lost RNA7 synthesis almost completely. On the other hand, mutants that contained exclusively the body nucleotide at position +1 retained higher levels of RNA7 synthesis. This observation may be explained as follows: in the wild-type situation, the large majority of the crossovers probably occur between positions –1 and +1, leading to a body TRS-derived nucleotide at position +1 in the sg RNA; however, a low number of crossovers take place between nucleotides +1 and +2, resulting in a leader TRS-derived nucleotide at position +1. Mutants in which almost all sg RNA synthesis is blocked by a substitution at position +1 may somehow be deficient in the crossover between –1 and +1, but may have retained the ability for crossovers between +1 and +2, which were detected by sequence analysis. Conversely, in position +1 mutants that retain reasonable sg RNA synthesis, most crossovers occur between positions –1 and +1, and they obscure the minority of crossovers between +1 and +2 in the sequencing electropherogram. Alternatively, position +1 TRS mutations that strongly interfere with sg RNA synthesis may force a shift of the crossover site in the remaining molecules.

We believe that our present findings strongly support the discontinuous minus strand extension model. Indeed, the fact that a complete body TRS can be copied into the sg RNA is very difficult to reconcile with the alternative model, in which sg RNA synthesis from the genomic minus strand template is primed by free plus strand leader transcripts that contain the leader TRS at their 3′ end (Figure 1B). To explain the presence of a complete copy of the body TRS in the sg mRNA in this model, one would have to assume that a 3′–5′ exonuclease activity trims back the free leader transcript prior to its extension into an sg mRNA (Baker and Lai, 1990). Note that there would not be a single base pair left to hold these ‘trimmed’ leader molecules on the template. Such an enzymatic activity, which is unprecedented in +RNA viruses, exists in yeast retrotransposon Ty5 (Ke *et al*., 1999), in which reverse transcription is primed by an internal region in a tRNA. However, in this system, it is not a part of the duplex that is removed, but the single-stranded 3′ tail of the tRNA, which cannot base-pair with the Ty5 RNA.

Removal of the TRS at the 3′ end of the nidovirus leader, which has already base paired with the template, would be very energetically unfavourable for the RdRp. Instead of starting elongation using the intact and properly positioned leader as a primer, it would have to disrupt the newly formed duplex, degrade part of the leader RNA and then reinitiate polymerization, without any base pairing between primer and template. It has been shown that influenza virus transcription does not require a sequence match between the (cellular) RNA primer and the (viral) template (Plotch *et al*., 1981). However, if in the nidovirus system the ‘trimmed’ leader RNA could also be fixed on the template solely by RNA–protein interactions, the targeting of the nascent strand by TRS base pairing would be extremely puzzling.

Sequence data of sg RNA leader–body junctions from other arteriviruses are also difficult to reconcile with the leader-primed transcription model. For the porcine and simian arteriviruses (Meulenberg *et al*., 1993; Godeny *et al*., 1998), the leader–body junctions of some sg RNAs mapped two nucleotides upstream of the body TRS, which again would not leave a single nucleotide to hold the putative free leader on the template after the hypothetical ‘back trimming’. On the other hand, these findings and our data can be explained readily by the discontinuous minus strand extension model (Figure 1C). The six-nucleotide

duplex formed between the body TRS complement at the 3′ end of the leaderless sg minus strand and the leader TRS in the genomic RNA template should suffice to position the nascent minus strand properly for subsequent elongation to add the complement of the leader sequence. In most cases, the nascent minus strand contains the entire body TRS complement at its 3′ end at the moment of strand transfer, leading to a body TRS-derived leader–body junction sequence in the sg mRNA molecule. In a small number of transcripts, however, minus strand synthesis appears to be interrupted before nucleotide +1 of the body TRS is copied and, after strand transfer, resumes by incorporating the complement of the +1 nucleotide of the leader TRS. As stated above, we postulate that the detection of this phenomenon is determined by the level of crossovers between the –1 and +1 position that is allowed by the mutations introduced at the +1 position of body TRS or leader TRS. We cannot, however, formally exclude that a 'back trimming' activity degrades the 3′-terminal nucleotide of the minus strand before or after strand transfer. However, note that in the discontinuous minus strand extension model (Figure 1C), such an activity would not disturb the proper positioning of the nascent minus strand on the leader template, because the TRS–TRS duplex would be shortened by one nucleotide only.

Nidovirus discontinuous minus strand extension resembles similarity-assisted, copy-choice RNA recombination

Due to their discontinuous sg RNA synthesis, nidoviruses occupy a special 'niche' in the +RNA virus world. Their mode of sg RNA production is clearly different from that of other +RNA viruses and resembles another well-documented +RNA virus feature: RNA recombination (for recent reviews see Nagy and Simon, 1997; Aaziz and Tepfer, 1999; Worobey and Holmes, 1999). Most of the experimental evidence supports an RdRp template switch (Kirkegaard and Baltimore, 1986) as the main mechanism of RNA recombination. Mechanistically, such a template switch involves the transfer of a nascent strand from one RNA template (donor) to the other (acceptor). Also, nidovirus discontinuous sg RNA synthesis involves transfer of a nascent RNA strand, the sg RNA, but now from one site to another in the same template.

Based on the data currently available, we refer to the discontinuous minus strand extension model as our working model for nidovirus sg RNA synthesis. If one applies the 'recombination terms' to this model (Chang *et al*., 1996; Brian and Spaan, 1997; van Marle *et al*., 1999a), the donor strand would be the body part of the genomic RNA template, the acceptor strand would be the leader part of the genomic RNA template and the nascent strand would be the discontinuously synthesized minus strand. Nagy and Simon (1997) have defined three main classes of RNA recombination: similarity-essential, similarity-non-essential and similarity-assisted recombination. The latter is defined as a mechanism in which strand transfer is determined by both sequence similarity between the parental RNAs and additional RNA determinants, present in only one of the parental RNAs.

The results of our present study strongly suggest that nidovirus discontinuous sg RNA synthesis can be considered a special case of high-frequency similarity-assisted RNA recombination. While the only obvious function of the leader TRS is to ensure the fidelity of the strand transfer by base pairing with the 3′ end of the nascent strand, the body TRS in the donor template indeed has additional, sequence-specific functions. One of these functions apparently is to pause (or terminate) nascent strand synthesis and thereby provide the opportunity for strand transfer. In addition, body TRS-derived nucleotides may play a role in the reinitiation of nascent strand synthesis on the acceptor template. Given the compact nature of the EAV TRS, it is quite possible that some nucleotides fulfil multiple tasks.

RNA secondary structure of the body TRS may regulate sg RNA synthesis

The sequence-specific function of the body TRS, revealed in this study, may be exerted at the level of either primary sequence or secondary structure. For a number of +RNA viruses, RNA secondary structure motifs located in the (proximity of) sg RNA promoters are vital for sg RNA synthesis. In alfalfa mosaic virus (Haasnoot *et al*., 2000), turnip crinkle virus (TCV) (Wang *et al*., 1999) and barley yellow dwarf virus (Koev *et al*., 1999), stem–loop structures in sg RNA promoter regions of the template strand are required for sg RNA synthesis. The sg RNA1 promoter of the latter virus is especially interesting, since it contains two stem–loop domains. For one of them, secondary structure, but not the primary sequence, is important for sg RNA synthesis, whereas the other domain acts through primary sequence, and not secondary structure (Koev *et al*., 1999). Similarly, RNA secondary structure may play only a minor role in the sequence-specific recognition of the BMV sg RNA promoter by the RdRp (Adkins *et al*., 1997; Siegel *et al*., 1997).

We have suggested previously that RNA secondary structure of body TRS regions contributes to their attenuating potential and thereby determines the relative portion of the nascent minus strands that is transferred to the leader TRS in the template (Pasternak *et al*., 2000). At present, it is unknown whether EAV body TRSs are part of an RNA structural motif that is essential for body TRS function, or whether they are recognized by a protein factor in a sequence-specific manner. However, the latter seems less likely than the former, since even LB4 (Figure 2), in which five TRS nucleotides were substituted, still produced some sg RNA7, although ~30-fold less than the wild-type control. The fact that some sequences in the EAV genome match the leader TRS perfectly, but are not used for sg mRNA synthesis, also argues against the recognition of a specific sequence (Pasternak *et al*., 2000). More probably, mutagenesis of the RNA7 body TRS disturbed an RNA structure that is necessary for its function. This could, for example, explain the fact that the BU_6C substitution reduced the amount of RNA7 by 20-fold (and could not be rescued by the same mutation in the leader TRS), whereas the wild-type RNA6 body TRS contains a C at the same position. If a protein factor were involved in sequence-specific TRS recognition, then one would expect it to recognize all TRSs similarly. If RNA structure is important for recognition by such a protein, then the BU_6C substitution probably disturbs a structural motif of the RNA7 TRS, which is not present in the RNA6 TRS. On the other hand, conservation of part of the TRS in other arteriviruses

suggests a sequence-specific recognition. Further studies are required to distinguish between these possibilities.

In the TCV satellite RNA recombination system, the hairpin structure in the acceptor strand, as well as the donor–acceptor homology region, are necessary for the template switch (Nagy *et al*., 1998). The hairpin has been postulated to bind the RdRp, whereas the homology region targets the nascent strand to the crossover site. The TCV RdRp probably recognizes the secondary and/or tertiary structure of the hairpin, while individual nucleotides play a less important role (Nagy and Simon, 1998). In EAV, the leader TRS in the acceptor template is predicted to reside in the loop of an extensive hairpin, and its base pairing interaction with the body TRS complement at the 3′ end of the nascent minus strand would resemble certain antisense RNA-regulated control mechanisms that are based on interactions between single-stranded tails and hairpin loops (van Marle *et al*., 1999a, and references therein). It is possible that the EAV RdRp, or its accessory proteins, also binds to the stem of the long hairpin that presents the leader TRS. In any case, the leader TRS itself does not seem to be recognized by a protein in a sequence-specific manner.

The body TRS is a better candidate to serve as a protein recognition site. This protein would then mediate the pausing of the nascent strand synthesis and/or nascent strand transfer. This would resemble the DNA-dependent RNA polymerase I termination system, in which specific DNA-binding terminator proteins bind to termination sequences (Reeder and Lang, 1997), or a function of the HIV nucleocapsid protein, which promotes the minus strand strong-stop DNA transfer (Guo *et al*., 1997). The EAV replicase component nsp1, which recently was shown to possess an sg RNA synthesis-specific activity (Tijms *et al*., 2001), may be a good candidate for such a regulatory role. Residues predicted to form a zinc finger structure in nsp1 were shown to be necessary for sg RNA synthesis. Interestingly, zinc finger structures in the HIV nucleocapsid protein facilitate strand transfer (Guo *et al*., 2000). Finally, it should be noted that the RNA structure of the nascent strand may also influence pausing, strand transfer or reinitiation, as illustrated by the fact that stable hairpin structures in the nascent strand promote termination of transcription by *Escherichia coli* RNA polymerase (Wilson and von Hippel, 1995).

Materials and methods

Site-directed mutagenesis, RNA transfections and immunofluorescence analysis

Site-directed mutagenesis of EAV leader and body TRSs was carried out as described by van Marle *et al*. (1999a), and all mutant constructs were sequenced. Following *in vitro* transcription from infectious cDNA clones, full-length EAV RNA was introduced into BHK-21 cells by electroporation, as described by van Dinten *et al*. (1997). Immunofluorescence assays with EAV-specific antisera were performed at 14 h post-transfection as described by van der Meer *et al*. (1998). To visualize the nuclei for cell counting, nuclear DNA was stained with 5 μg/ml Hoechst B2883 (Sigma). Cells were counted using the Scion Image software (Scion Corporation) and the percentage of transfected cells was calculated on the basis of the number of cells positive for the EAV replicase component nsp3 (Pedersen *et al*., 1999).

RNA isolation and analysis

For RNA analyses, cells were lysed at 14 h post-transfection. Intracellular RNA isolation was performed using the acidic phenol method as described by Pasternak *et al*. (2000). Total intracellular RNA was resolved in denaturing agarose–formaldehyde gels. Hybridization of dried gels with the radioactively labelled oligonucleotide probe E154, which is complementary to the 3′ end of the EAV genome and recognizes all viral mRNA molecules (genomic and subgenomic), and phosphoimager quantitation of individual bands were performed as described by Pasternak *et al*. (2000). To determine the leader–body junction sequence of sg mRNA7, mRNA7-specific RT–PCRs were carried out as described by van Marle *et al*. (1999b) using an antisense (RT and PCR) primer from the RNA7 body region and a sense PCR primer matching a part of the leader sequence. RT–PCR products were sequenced directly as described by Pasternak *et al*. (2000) using the leader-derived primer, an ABI PRISM™ sequencing kit (Perkin Elmer) and an ABI PRISM™ 310 Genetic Analyser (Perkin Elmer).

Acknowledgements

We are grateful to Stanley Sawicki, Dorothea Sawicki, Paul Masters, Alexander Gultyaev, Kees Pleij, Marieke Tijms and Richard Molenkamp for helpful discussions and comments. We acknowledge Jessika Dobbe for technical assistance. A.O.P. was supported by grant 700-31-020 from the Council for Chemical Sciences of the Netherlands Organization for Scientific Research.

References

Aaziz,R. and Tepfer,M. (1999) Recombination in RNA viruses and in virus-resistant transgenic plants. *J. Gen. Virol.*, **80**, 1339–1346.

Adkins,S., Siegel,R.W., Sun,J.H. and Kao,C.C. (1997) Minimal templates directing accurate initiation of subgenomic RNA synthesis *in vitro* by the brome mosaic virus RNA-dependent RNA polymerase. *RNA*, **3**, 634–647.

Almazan,F., Gonzalez,J.M., Penzes,Z., Izeta,A., Calvo,E., Plana-Duran.J. and Enjuanes,L. (2000) Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc. Natl Acad. Sci. USA*, **97**, 5516–5521.

Baker,S.C. and Lai,M.M.C. (1990) An *in vitro* system for the leader-primed transcription of coronavirus mRNAs. *EMBO J.*, **9**, 4173–4179.

Baric,R.S. and Yount,B. (2000) Subgenomic negative-strand RNA function during mouse hepatitis virus infection. *J. Virol.*, **74**, 4039–4046.

Baric,R.S., Stohlman,S.A. and Lai,M.M.C. (1983) Characterization of replicative intermediate RNA of mouse hepatitis virus: presence of leader RNA sequences on nascent chains. *J. Virol.*, **48**, 633–640.

Baric,R.S., Stohlman,S.A., Razavi,M.K. and Lai,M.M.C. (1985) Characterization of leader-related small RNAs in coronavirus-infected cells: further evidence for leader-primed mechanism of transcription. *Virus Res.*, **3**, 19–33.

Brian,D.A. and Spaan,W.J.M. (1997) Recombination and coronavirus defective interfering RNAs. *Semin. Virol.*, **8**, 101–111.

Chang,R.Y., Krishnan,R. and Brian,D.A. (1996) The UCUAAAC promoter motif is not required for high-frequency leader recombination in bovine coronavirus defective interfering RNA. *J. Virol.*, **70**, 2720–2729.

den Boon,J.A., Kleijnen,M.F., Spaan,W.J. and Snijder,E.J. (1996) Equine arteritis virus subgenomic mRNA synthesis: analysis of leader–body junctions and replicative-form RNAs. *J. Virol.*, **70**, 4291–4298.

Godeny,E.K., de Vries,A.A.F., Wang,X.C., Smith,S.L. and de Groot,R.J. (1998) Identification of the leader–body junctions for the viral subgenomic mRNAs and organization of the simian hemorrhagic fever virus genome: evidence for gene duplication during arterivirus evolution. *J. Virol.*, **72**, 862–867.

Guo,J., Henderson,L.E., Bess,J., Kane,B. and Levin,J.G. (1997) Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. *J. Virol.*, **71**, 5178–5188.

Guo,J., Wu,T., Anderson,J., Kane,B.F., Johnson,D.G., Gorelick,R.J., Henderson,L.E. and Levin J.G. (2000) Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *J. Virol.*, **74**, 8980–8988.

Haasnoot,P.C., Brederode,F.T., Olsthoorn,R.C. and Bol,J.F. (2000) A conserved hairpin structure in alfamovirus and bromovirus subgenomic promoters is required for efficient RNA synthesis *in vitro*. *RNA*, **6**, 708–716.

Joo,M. and Makino,S. (1992) Mutagenic analysis of the coronavirus intergenic consensus sequence. *J. Virol.*, **66**, 6330–6337.

Ke,N., Gao,X., Keeney,J.B., Boeke,J.D. and Voytas,D.F. (1999) The yeast retrotransposon Ty5 uses the anticodon stem–loop of the initiator methionine tRNA as a primer for reverse transcription. *RNA*, **5**, 929–938.

Kim,K.H. and Hemenway,C.L. (1999) Long-distance RNA–RNA interactions and conserved sequence elements affect potato virus X plus-strand RNA accumulation. *RNA*, **5**, 636–645.

Kirkegaard,K. and Baltimore,D. (1986) The mechanism of RNA recombination in poliovirus. *Cell*, **47**, 433–443.

Koev,G., Mohan,B.R. and Miller,W.A. (1999) Primary and secondary structural elements required for synthesis of barley yellow dwarf virus subgenomic RNA1. *J. Virol.*, **73**, 2876–2885.

Lai,M.M.C. and Cavanagh,D. (1997) The molecular biology of coronaviruses. *Adv. Virus Res.*, **48**, 1–100.

Lai,M.M.C., Patton,C.D. and Stohlman,S.A. (1982) Replication of mouse hepatitis virus: negative-stranded RNA and replicative form RNA are of genome length. *J. Virol.*, **44**, 487–492.

Lai,M.M.C., Baric,R.S., Brayton,P.R. and Stohlman,S.A. (1984) Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proc. Natl Acad. Sci. USA*, **81**, 3626–3630.

Meulenberg,J.J.M., de Meijer,E.J. and Moormann,R.J.M. (1993) Subgenomic RNAs of Lelystad virus contain a conserved leader–body junction sequence. *J. Gen. Virol.*, **74**, 1697–1701.

Miller,W.A. and Koev,G. (2000) Synthesis of subgenomic RNAs by positive-strand RNA viruses. *Virology*, **273**, 1–8.

Miller,W.A., Dreher,T.W. and Hall,T.C. (1985) Synthesis of brome mosaic virus subgenomic RNA *in vitro* by internal initiation on minus-sense genomic RNA. *Nature*, **313**, 68–70.

Nagy,P.D. and Simon,A.E. (1997) New insights into the mechanisms of RNA recombination. *Virology*, **235**, 1–9.

Nagy,P.D. and Simon,A.E. (1998) *In vitro* characterization of late steps of RNA recombination in turnip crinkle virus. I. Role of motif1-hairpin structure. *Virology*, **249**, 379–392.

Nagy,P.D., Zhang,C. and Simon,A.E. (1998) Dissecting RNA recombination *in vitro*: role of RNA sequences and the viral replicase. *EMBO J.*, **17**, 2392–2403.

Pasternak,A.O., Gultyaev,A.P., Spaan,W.J.M. and Snijder,E.J. (2000) Genetic manipulation of arterivirus alternative mRNA leader–body junction sites reveals tight regulation of structural protein expression. *J. Virol.*, **74**, 11642–11653.

Pedersen,K.W., van der Meer,Y., Roos,N. and Snijder,E.J. (1999) Open reading frame 1a-encoded subunits of the arterivirus replicase induce endoplasmic reticulum-derived double-membrane vesicles which carry the viral replication complex. *J. Virol.*, **73**, 2016–2026.

Plotch,S.J., Bouloy,M., Ulmanen,I. and Krug,R.M. (1981) A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell*, **23**, 847–858.

Reeder,R.H. and Lang,W.H. (1997) Terminating transcription in eukaryotes: lessons learned from RNA polymerase I. *Trends Biochem. Sci.*, **22**, 473–477.

Sawicki,D.L., Wang,T. and Sawicki,S.G. (2001) The RNA structures engaged in replication and transcription of the A59 strain of mouse hepatitis virus. *J. Gen. Virol.*, **82**, 385–396.

Sawicki,S.G. and Sawicki,D.L. (1995) Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands. *Adv. Exp. Med. Biol.*, **380**, 499–506.

Sethna,P.B., Hung,S.L. and Brian,D.A. (1989) Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicon. *Proc. Natl Acad. Sci. USA*, **86**, 5626–5630.

Siegel,R.W., Adkins,S. and Kao,C.C. (1997) Sequence-specific recognition of a subgenomic RNA promoter by a viral RNA polymerase. *Proc. Natl Acad. Sci. USA*, **94**, 11238–11243.

Sit,T.L., Vaewhongs,A.A. and Lommel,S.A. (1998) RNA-mediated *trans*-activation of transcription from a viral RNA. *Science*, **281**, 829–832.

Snijder,E.J. and Meulenberg,J.J.M. (1998) The molecular biology of arteriviruses. *J. Gen. Virol.*, **79**, 961–979.

Spaan,W.J.M, Delius,H., Skinner,M., Armstrong,J., Rottier,P.J.M., Smeekens,S., van der Zeijst,B.A.M. and Siddell,S.G. (1983) Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. *EMBO J.*, **2**, 1839–1844.

Thiel,V., Herold,J., Schelle,B. and Siddell.S.G. (2001) Infectious RNA transcribed *in vitro* from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J. Gen. Virol.*, **82**, 1273–1281.

Tijms,M.A., van Dinten,L.C., Gorbalenya,A.E. and Snijder,E.J. (2001) A zinc finger-containing papain-like protease couples subgenomic mRNA synthesis to genome translation in a positive-stranded RNA virus. *Proc. Natl Acad. Sci. USA*, **98**, 1889–1894.

van der Meer,Y., van Tol,H., Krijnse Locker,J. and Snijder,E.J. (1998) ORF1a-encoded replicase subunits are involved in the membrane association of the arterivirus replication complex. *J. Virol.*, **72**, 6689–6698.

van der Most,R.G., de Groot,R.J. and Spaan,W.J.M. (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.

van Dinten,L.C., den Boon,J.A., Wassenaar,A.L.M., Spaan,W.J.M. and Snijder,E.J. (1997) An infectious arterivirus cDNA clone: identific ation of a replicase point mutation that abolishes discontinuous mRNA transcription. *Proc. Natl Acad. Sci. USA*, **94**, 991–996.

van Marle,G., Dobbe,J.C., Gultyaev,A.P., Luytjes,W., Spaan,W.J.M. and Snijder,E.J. (1999a) Arterivirus discontinuous mRNA transcription is guided by base pairing between sense and antisense transcription-regulating sequences. *Proc. Natl Acad. Sci. USA*, **96**, 12056–12061.

van Marle,G., van Dinten,L.C., Spaan,W.J.M., Luytjes,W. and Snijder, E.J. (1999b) Characterization of an equine arteritis virus replicase mutant defective in subgenomic mRNA synthesis. *J. Virol.*, **73**, 5274–5281.

Wang,J., Carpenter,C.D. and Simon,A.E. (1999) Minimal sequence and structural requirements of a subgenomic RNA promoter for turnip crinkle virus. *Virology*, **253**, 327–336.

Wilson,K.S. and von Hippel,P.H. (1995) Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proc. Natl Acad. Sci. USA*, **92**, 8793–8797.

Worobey,M. and Holmes,E.C. (1999) Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.*, **80**, 2535–2543.

Zhang,G., Slowinski,V. and White,K.A. (1999) Subgenomic mRNA regulation by a distal RNA element in a plus-strand RNA virus. *RNA*, **5**, 550–561.

Received July 16, 2001; revised October 25, 2001; accepted November 1, 2001