Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replications

(Kinetics of RNA synthesis/mRNA replication)

PHIROZE B. Sethna, SHAN-LING HUNG*, and DAVID A. BRIAN†

Department of Microbiology, University of Tennessee, Knoxville, TN 37996-0845

Communicated by Dorothy M. Horstmann, April 24, 1989 (received for review February 28, 1989)

ABSTRACT The genome of the porcine transmissible gastroenteritis coronavirus (TGEV) (1), like that of the avian infectious bronchitis coronavirus (2) and the mouse hepatitis coronavirus (MHV) (3), is infectious. It is therefore presumed that a single molecule of genomic RNA is sufficient for initiating infection, and much evidence now supports the hypothesis that genome replication occurs through a full-length minus-strand antigenome that also serves as the template for leader-primed transcription of subgenomic mRNA molecules (4-6). Since leader priming initiates at specific internal sites on the minus-strand antigenome and proceeds through the 5′ end of the molecule, mRNAs are made that form a 3′-coterminal nested set. Except for the smallest species, coronavirus mRNAs are structurally polycistronic but function primarily as monocistronic molecules with usually only the 5′-terminal open reading frame being translated (7-10). The subgenomic mRNAs of coronaviruses, if made by the leader-priming mechanism, would therefore be expected to have a 5′ untranslated leader sequence of ~80 bases that is identical to the 5′ end of the genome and a 3′ noncoding terminus of ~300 bases that is identical to the 3′ end of the genome (Fig. 1). This indeed seems to be the case from sequence data (11-14).

Since the promoter for synthesis of the minus-strand antigenome by the TGEV RNA-dependent RNA polymerase (15) presumably resides within the 276-base noncoding region at the 3′ end of the plus-strand genome (16), and the promoter for genome synthesis presumably resides within the 80-base antileader sequence (an estimated length) at the 3′ end of the minus-strand antigenome, it is natural to ask whether subgenomic mRNAs undergo replication as does the genome, since they possess 3′ and 5′ end sequences that are identical to those of the genome.

We have addressed this question by seeking the existence of subgenomic minus-strand RNA molecules in cells infected with TGEV. Synthetic oligodeoxynucleotide probes were used that specifically identified both full-length and subgenomic plus- and minus-strand RNA species, and the kinetics of synthesis of the most abundant species was measured. Evidence for mRNA replications in the form of replicative intermediates was found, and we propose that these function as a mechanism for mRNA amplification. We further propose that mRNA replications compete with replicating genome for a limiting factor during RNA synthesis, perhaps the viral RNA polymerase, much as do defective interfering RNA species of some defective viruses, and that this explains how coronaviruses readily establish persistent infections in cell culture.

MATERIALS AND METHODS

Preparation of RNA from Uninfected and Infected Cells. Clone 116 of the Purdue strain of TGEV was plaque-purified and grown on swine testicle cells as described (1). The virus was plaque-purified again by using infectious genomic RNA

Fig. 1. Structural relationships among the genome, mRNA transcripts, and synthetic oligodeoxynucleotide probes. The seven open reading frames (ORFS) deduced from primary sequence of the 3′ end of the TGEV genome are drawn to scale. They are identified as P for peplomer protein, M for matrix protein, N for nucleocapsid protein, HP for hydrophobic protein, and 7.7k, 27.7k, and 9.2k for potential nonstructural proteins of 7.7, 27.7, and 9.2 kDa, respectively. Their corresponding mRNAs are identified as p, m, n, hp, 7.7k, 27.7k, and 9.2k. The site from which the oligonucleotide probes were derived is indicated by an arrow. The 5′ leader sequence (presumed to be 80 bases) and the 3′ noncoding sequence (276 bases) are indicated by heavy lines and are drawn to scale. The 3′ poly(A) tail is indicated by a wavy line. The largest mRNA, which serves as template for synthesis of the viral polymerase, is presumably identical to the genome.

Abbreviations: TGEV, porcine transmissible gastroenteritis coronavirus; MHV, mouse hepatitis coronavirus.

*Present address: Department of Microbiology, School of Dentistry, University of Pennsylvania, Philadelphia, PA 19104.

†To whom reprint requests should be addressed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
(1) and a virus stock was prepared by passing the virus five times, using a multiplicity of infection of ~1 at each passage. For RNA preparation, cells were grown to confluency in 850-cm² roller bottles (~3.8 × 10⁸ cells per bottle) and either mock-infected or infected with stock virus at a multiplicity of infection of 10. One hour after addition of inoculum, cells were rinsed twice with warm Earle’s balanced salt solution and either harvested for RNA extraction (0 hr postinfection) or refed with growth medium and incubated at 37°C until harvested at 0.5, 1, 1.5, 2, 4, 6, or 10 hr postinfection. Total cytoplasmic RNA was isolated as described (17) except that the lysis buffer contained 10 mM vanadyl ribonucleoside complex. The RNA precipitate was kept under ethanol (total volume of 24 ml for RNA from one roller bottle) at −20°C and resuspended by Vortex mixing immediately before sample removal. For quantitation by UV absorption, a fraction of the RNA was extracted with phenol/0.1% 8-hydroxyquinoline to remove residual vanadyl ribonucleoside complex, which also absorbs at 260-nm wavelength. Ten microliters (in ~100 µl of precipitate suspension) was used per lane for electrophoresis except when RNase-resistant forms were analyzed (described below).

Preparation of Virion RNA. Virus was purified from 1 liter of supernatant and RNA was extracted as described (16). One hundred microliters of precipitate suspension (from a total volume of 8 ml), containing ~20 ng of virion RNA, was electrophoresed per lane.

Preparation of Radiolabeled Oligodeoxynucleotide Probes. A stretch of G+G-rich sequence (57% G+C) was arbitrarily chosen from within the HP gene (Fig. 1) and a 26-mer oligodeoxynucleotide complementary to virus sense RNA and having the sequence 5'-CAGCATTGGAGAAAGCAGCTCCTG-3' (identified as probe 1) was synthesized by the phosphoramidite method. Probe 2 was likewise synthesized but has a sequence complementary to probe 1. Oligodeoxynucleotides were purified by size-exclusion chromatography through Sephadex G-25 NICK columns (Pharmacia) and the concentration of each oligodeoxynucleotide was determined by absorbency at 260 nm, with 1 A₂₆₀ unit equivalent to 20 µg. Oligodeoxynucleotide (100 ng, 11 pmol) was end-labeled by the forward reaction using [γ-³²P]ATP (ICN) and polynucleotide kinase (New England Biolabs) (18). Unincorporated [γ-³²P]ATP was removed with a Bio-Spin 6 column (Bio-Rad).

To determine the specific activity of the radiolabeled probe, it was assumed that all of the oligodeoxynucleotide was recovered, and the radioactivity was quantitated by spotting a sample onto Nytran membrane (Schleicher & Schuell) and then excising the spot for liquid scintillation counting in Scintiverse ( Fisher). Specific activity of the radiolabeled probes ranged from 1.6 to 3.5 × 10⁶ cpm/pmol.

Electrophoresis and Hybridization Analysis of RNA. RNA was removed as resuspended precipitate in ethanol, dried under vacuum (Savant SpinVac), dissolved in 37 µl of 1× Mops buffer (20 mM 4-morpholinopiperanesulfonic acid, pH 7.0/5 mM sodium acetate/1 mM EDTA)/50% (vol/vol) deionized formamide/2.2 M formaldehyde, denatured by heating at 65°C for 5 min, mixed with 8 µl of loading dye (50% glycerol/0.25% bromophenol blue/0.25% xylene cyanol/1 mM EDTA), and loaded onto a horizontal 1% agarose gel (20 × 25 cm and 0.5 cm thick) made in 1× Mops buffer containing 2.2 M formaldehyde. Electrophoresis was carried out at 140 V for 4 hr at room temperature in 1× Mops buffer. RNA was transferred to Nytran membrane (Schleicher & Schuell) by using a VacuBlot apparatus (LKB) and 20× SSC (1× SSC is 0.15 M NaCl/0.015 M sodium citrate) for 6 hr. From separate experimentation with radiolabeled RNA, the degree of transfer was found to range from 80% for the smallest RNA species to 60% for the largest (genome). RNA was UV-crosslinked (19) and the membrane was cut into two halves and prehybridized at 55°C for 2 hr in 5× SSC/0.1% Ficoll/0.1% polyvinylpyrrolidone/0.1% bovine serum albumin/50 mM sodium phosphate, pH 7.0/1% NaDodSO₄ containing 100 µg of sheared salmon sperm DNA and 50 µg of rRNA per ml.

Radiolabeled oligodeoxynucleotide probes (~2 × 10⁶ Cerenkov cpm per 120-cm² membrane) were denatured at 90°C for 5 min and added to the prehybridization solution, and hybridization was carried out at 55°C for 16 hr. The membrane was then given three 10-min washes in 2× SSC at 25°C and a fourth wash for 30 min at 55°C, air-dried, and exposed to Kodak XAR-5 film at −70°C with an intensifying screen for 18–24 hr. Radioactive bands were excised for liquid scintillation counting in Scintiverse. The number of molecules of each RNA species per cell was determined from the specific activity of the individual probe and from our measured yield of 10 µg of RNA per 1.6 × 10⁶ cells.

Identification of Double-Stranded Replicative Forms. Intra-cellular RNA extracted 6 hr postinfection was used for detecting double-stranded RNA. RNA (60 µg) in ethanol suspension was pelleted, washed once with 80% ethanol, dried, dissolved in 40 µl of 10 mM Tris-HCl, pH 7.2/300 mM NaCl/10 mM MgCl₂/1 mM EDTA containing 10 µg of RNase A (Sigma) per ml, and incubated at 37°C for 30 min. Digestion was terminated by addition of NaDodSO₄ (final concentration, 2%) followed by phenol/chloroform and chloroform/isoamyl alcohol extraction, and RNA was precipitated by adding 2.2 volumes of ethanol. RNA was analyzed by electrophoresis and hybridization as described above. RNase-treated RNA (30 µg) and untreated RNA (10 µg) were analyzed in adjacent lanes.

RESULTS

There Exists a Subgenomic Minus-Strand Counterpart for Each Plus-Strand mRNA Species. Studies in this laboratory (16, 17, 20, 21) have determined the nucleotide sequence of the 3'-terminal 8.5 kilobases of the TGEV genome and have identified seven open reading frames (Fig. 1). The structural protein genes have also been sequenced in other laboratories. (22–24). Each open reading frame in Fig. 1 is preceded by the consensus intergenic sequence CYAAC, which is thought to function in leader priming of transcription, and thus an mRNA for each of these open reading frames can be expected (14, 16, 17, 25). Polyadenylated RNAs for six of the seven open reading frames were previously identified by metabolic labeling experiments (15), and the species encoding the major structural proteins as well as the 27.7-kDa putative nonstructural protein have been identified by in vitro translation studies (8).

To determine by a second experimental approach whether a plus-strand transcript is made for each of the seven open reading frames, RNA hybridization analyses were done using a single-stranded, minus-strand nucleic acid probe (probe 1) that is complementary to a region within the HP gene, the 3'-most open reading frame (Fig. 2, lanes 1–10). Because of the well documented 3' nested-set arrangement of coronavirus mRNAs (26–28), a 3'-end probe can be expected to identify all mRNAs, including the genome, which apparently functions as mRNA for synthesis of the RNA-dependent RNA polymerase (29, 30). Fig. 2 lanes 1–10 illustrate that probe 1 from within the HP gene does identify this mRNA (21) (identified as hp mRNA) and seven larger RNA species of the appropriate size to represent transcripts of the identified open reading frames and progeny genome. These are respectively named hp, n, m, 9.2k, 27.7k, 7.7k, p, and genome. The specificity of this approach was confirmed by using a 26-mer oligodeoxynucleotide probe, also having a G+G content of 57% but complementary to a region within the N gene. As would be predicted, the N-gene probe identified the n mRNA and all larger species but not the hp mRNA (data not shown).
Fig. 2. Identification of subgenomic plus- and minus-strand RNA species by probes derived from the HP gene sequence. RNA from purified virions or from the cytoplasm of uninfected or infected cells (hours postinfection are shown above lanes) was electrophoresed, blotted, and hybridized with radiolabeled probe 1 to detect plus-strand RNA species (lanes 1–10) or with probe 2 to detect minus-strand RNA species (lanes 11–20). α indicates antisense, or minus-strand, polarity; asterisks indicate uncharacterized RNA species (see text); three dots at the top of each lane identify the well.

To determine whether subgenomic minus-strand RNAs exist, probe 2, which is complementary to probe 1, was used on RNA prepared and analyzed the same way, and a counterpart to each plus-strand species was found (Fig. 2, lanes 11–20). That is, there exists a 5′-coterminal nested set of RNA species appearing as complements to the mRNA species. Four observations suggested that probes designed to detect minus strands were not merely detecting abundant mRNA molecules nonspecifically. (i) Three of the eight minus-strand species (αhp, α27.7k, and α7.7k) did not migrate with the same mobility in formaldehyde gels as did their presumed plus-strand counterparts [Fig. 2 and data (not shown) obtained by electrophoresis of plus and minus strands in alternating lanes]. Differences in migration rates would be expected if base compositions between the plus- and minus-strand counterparts differed significantly. A difference would also be expected if there were no poly(U) copy of the 3′ poly(A) tail on the mRNA. (ii) A probe from within the N gene (a complement of the 26-mer described above), designed to detect minus strand, did so and also detected the larger minus-strand species, but not αhp, as would be expected if the minus-strand RNAs formed a 5′-coterminal nested set (data not shown). (iii) The maximal abundance of minus-strand species occurred at 4 hr postinfection, whereas that of plus-strand species occurred at 6 hr (Fig. 3). (iv) Probes used to detect minus-strand species did not identify virion genomic RNA (Fig. 2, lane 20, and data not shown), whereas both probes used to detect plus-strand sequences did (Fig. 2, lane 1, and data not shown).

Three species of minus-strand RNA for which we have no explanation at the present time were identified with probe 2 (asterisks in Fig. 2). These are a broad band of high molecular

Fig. 3. Kinetics of plus-strand (A) and minus-strand (B) RNA synthesis. Radiolabeled bands in Fig. 2 were excised and radioactivity was measured by scintillation spectroscopy. Copy numbers were determined as described in Materials and Methods.
weight RNA migrating between the antigenome and the well of origin, a species located between a7.7k and αp, and a species located between ahp and αn.

Subgenomic, Double-Stranded Replicative Forms Correspon- 
d in Size to mRNA Species. If a minus-strand counterpart of each mRNA exists that might be involved in the replication of mRNA, then a corresponding RNase-resistant double-stranded form should be found in cytoplasmic RNA. To identify double-stranded forms, cytoplasmic RNA isolated 6 hr postinfection was digested with pancreatic RNase A in the presence of 0.3 M NaCl and analyzed by RNA hybridization after electrophoresis in formaldehyde/agarose gels. Both plus and minus strands corresponding in size to mRNA species were found, indicating that subgenomic replicative forms were present (Fig. 4A). Furthermore, within the double-stranded forms, there were very few full-length (i.e., mRNA-length) plus-strand RNAs (Fig. 4A, lanes 2 and 6), but full-length minus-strand RNAs were abundant (lane 4). These results indicate that the replicative forms were most probably derived from replicative intermediate structures having a single minus-strand template and multiple plus-strand tails (Fig. 4B). Such structures would generate only short fragments of protected plus-strand molecules following RNase digestion, and these fragments would be too small to be resolved by electrophoresis in a 1% agarose gel. Replicative intermediates having this structure have been characterized for genome-length RNA during picornavirus and togavirus replication, and they were found to occur at the time of peak plus (genome)-strand synthesis (31, 32).

Both Plus- and Minus-Strand Subgenomic RNA Species Are, in General, Synthesized at a Rate Inversely Related to Their Length. Two points emerge from our quantitative analysis of plus- and minus-strand RNA synthesis. First, with the exception of the shortest (hp) mRNA, the shorter RNA species were made at a higher rate than longer species. n mRNA and its complement were made most rapidly, followed by hp, then m, and so on in order of increasing length. This can be observed by noting slopes throughout the first 6 hr for plus-strand RNA synthesis (Fig. 3A) and throughout the first 4 hr for minus-strand RNA synthesis (Fig. 3B). Second, although we were unable to measure turnover rates with our methods, the number of plus-strand molecules at all times exceeded the number of minus strands for any given species. At 6 hr postinfection, the time of peak plus-strand synthesis, this ratio ranged from 2 for genome RNA to 10 for n mRNA (Fig. 3A). At 4 hr postinfection, the time of peak minus-strand synthesis, the molar ratio of subgenomic minus-strand species to antigenome ranged from 1 for αp to 8 for αn (Fig. 3B).

**DISCUSSION**

The processes of RNA transcription and replication are one and the same for picornaviruses and those plus-strand viruses for which only one mRNA molecule is made, which is identical or nearly identical to genomic RNA (33). For the plus-strand togaviruses that synthesize a subgenomic plus-strand mRNA molecule as well as genome-length mRNA, transcription and replication are separate processes with regard to the subgenomic mRNA but are apparently the same for genome-length mRNA (34). In the case of togaviruses, it is not clear why the subgenomic mRNA, which has a 3' terminus that is identical to the 3' end of the genome, does not undergo replication to generate a subgenomic minus strand. It has been proposed that the full-length genome, which is known to circularize, does so and allows the 5' end to interact with the 3' end of the molecule in such a way as to enable polymerase initiation for minus-strand synthesis (34). The subgenomic mRNA is missing whatever is required at the 5' terminus (possibly a stable double-hairpin structure) for circularization or for polymerase recognition, or both, and minus-strand synthesis does not take place.

The picture is different, however, for coronaviruses. Although one mRNA species (the largest) is apparently identical to genomic RNA, the subgenomic mRNAs have both 5' and 3' termini that are identical to those of the genome (11–14), so theoretically there is no reason why they should

---

**Fig. 4.** (A) RNase-resistant species in the replicative form. Cytoplasmic RNA obtained from cells 6 hr postinfection was digested with RNase A in 0.3 M NaCl, electrophoresed in a denaturing gel, transferred, and hybridized with probes from the HP gene sequence. Lane 1, 10 μg of untreated RNA, and lane 2, 30 μg of RNase-treated RNA, were hybridized with probe 1. Lane 3, 10 μg of untreated RNA, and lane 4, 30 μg of RNase-treated RNA, were hybridized with probe 2. Lanes 5 and 6 are extended exposures of lanes 1 and 2. (B) Model showing the relationship between the replicative intermediate (RI) and resulting double-stranded replicative form (RF) following RNase digestion for a replicative intermediate of 2.5 kilobases. Sites to which probes 1 (on the plus strand) and 2 (on the minus strand) would bind are depicted by a filled circle and an open circle, respectively.
not replicate the same as the genome, provided the promoters for minus- and plus-strand synthesis lie within these terminal sequences. Experiments with MHV suggested that this is probably not the case, since only genome-length minus-strand RNA and genome-length replicative forms were found in infected cells (4, 5).

Our data show that TGEV exhibits a major difference from MHV with regard to the behavior of subgenomic molecules. Subgenomic minus strands were found in TGEV-infected cells that cannot, on the basis of kinetics of appearance, be breakdown products of antigenic RNA. We propose that they arose by the action of replicase on mRNA templates. Furthermore, RNase-resistant replicative forms were found that corresponded in size to TGEV mRNAs and bore structures that most probably arose from parental replicative intermediates having a single minus-strand template and multiple plus-strand tails. Such structures indicate that subgenomic minus strands, in turn, serve as templates for the synthesis of new mRNAs. Proof for our assertions regarding the origin and function of subgenomic minus strands, however, will require further evidence showing subgenomic double-stranded forms to be replicationally and transcriptionally active.

It is not clear why there is such a striking difference between TGEV and MHV with regard to the existence of subgenomic minus-strand RNAs. It is known that much divergence exists among coronavirus species in both the primary structure of genes of homologous proteins and in the arrangement of genes along the genome (reviewed in ref. 35). In many cases insertions or deletions appear both in genes and in noncoding sequences of closely related strains. High-frequency recombination among coronaviruses undoubtedly contributes to some gene rearrangements (6). It is possible, therefore, that polymerase recognition signals that are functional on TGEV mRNA species for the synthesis of minus strands have been lost or altered on MHV subgenomic mRNAs.

The phenomenon of subgenomic mRNA replication does not rule out the mechanism of leader-primed transcription but, rather, suggests a second compatible mechanism by which mRNA can be produced. Conceivably, mRNA replication could begin with nascent products of leader-primed transcription or with mRNA carried into the cell as part of the infecting virion. We predict that leader-primed transcription would be the more important source, however, because the number of mRNAs incorporated into virions is small (data not shown). For any given subgenomic mRNA species there is <1 copy incorporated for every 10 copies of genome.

Two biological consequences can be predicted from the replication of subgenomic mRNAs. The first is that it may be a mechanism for rapidly amplifying mRNA levels for the synthesis of structural proteins that are required in highest numbers during virus replication. To test this hypothesis, it will be important to determine whether there are structural or functional differences between mRNAs produced by the leader-priming and replication mechanisms. The second consequence is that replicating mRNAs would most probably compete with replicating genome for limiting factors required in RNA replication, possibly the RNA polymerase, and behave as defective interfering particle RNAs having internal deletions (reviewed in refs. 36 and 37). This has been proposed to result in attenuation of viral cytopathogenesis and allow the establishment of persistent viral infections. This may explain how coronaviruses can so readily establish persistent infections in cell culture.

We thank Sushma Abraham, Tom Kienzle, Joanne Maki, and Brenda Hogue for many helpful discussions, and Stuart Riggby and Jerry Weir for manuscript suggestions. This work was supported by the National Institutes of Health (AI14367), the Department of Agriculture (85-CRSR-2-2623), and the University of Tennessee, College of Veterinary Medicine, Center of Excellence Program in Livestock Diseases and Human Health.