Identification of Novel Subgenomic RNAs and Noncanonical Transcription Initiation Signals of Severe Acute Respiratory Syndrome Coronavirus

Snawar Hussain,† Ji’an Pan,† Yu Chen, Yalin Yang, Jing Xu, Yu Peng, Ying Wu, Zhaoyang Li, Ying Zhu, Po Tien, and Deyin Guo*

Modern Virology Research Centre and National Key Laboratory of Virology, College of Life Sciences, Wuhan University, Wuhan, People’s Republic of China

Received 13 October 2004/Accepted 14 December 2004

The expression of the genomic information of severe acute respiratory syndrome coronavirus (SARS CoV) involves synthesis of a nested set of subgenomic RNAs (sgRNAs) by discontinuous transcription. In SARS CoV-infected cells, 10 sgRNAs, including 2 novel ones, were identified, which were predicted to be functional in the expression of 12 open reading frames located in the 3’ one-third of the genome. Surprisingly, one new sgRNA could lead to production of a truncated spike protein. Sequence analysis of the leader-body fusion sites of each sgRNA showed that the junction sequences and the corresponding transcription-regulatory sequence (TRS) are unique for each species of sgRNA and are consistent after virus passages. For the two novel sgRNAs, each used a variant of the TRS that has one nucleotide mismatch in the conserved hexanucleotide core (ACGAAC) in the TRS. Coexistence of both plus and minus strands of SARS CoV sgRNAs and evidence for derivation of the sgRNA core sequence from the body core sequence favor the model of discontinuous transcription during minus-strand synthesis. Moreover, one rare species of sgRNA has the junction sequence AAA, indicating that its transcription could result from a noncanonical transcription signal. Taken together, these results provide more insight into the molecular mechanisms of genome expression and subgenomic transcription of SARS CoV.

Severe acute respiratory syndrome (SARS) is an atypical form of pneumonia that was first recognized in Guangdong Province, China, in November 2002, and its causative agent was identified as novel a coronavirus (SARS CoV) (7, 9, 14). Coronaviruses are the largest RNA viruses, containing a single-stranded, plus-sense RNA ranging from 27 to 31.5 kb in size. The genomes of coronaviruses, possessing a 5’ cap structure and 3’ poly(A) tail, are polycistronic and are expressed through a poorly understood regulatory mechanism (11). The two large open reading frames (ORFs) (1a and 1b) at the 5’ end of the genome encode the viral replicase and are translated directly from the genomic RNA, while ORF 1b is expressed by -1 ribosomal frameshifting (26). The 3’ one-third of the genome comprises the genes encoding the structural and auxiliary proteins translated through six to nine nested and 3’-coterminal subgenomic RNAs (sgRNAs), but the number, composition, and expression strategies of the 3’-proximal ORFs vary greatly among coronaviruses, although four genes for the structural proteins S, E, M, and N are always included (11).

A unique feature for coronaviruses and some related viruses in the order Nidovirales is that the viral sgRNAs contain a common leader sequence of 55 to 92 nucleotides (nt), which is derived from the 5’ end of the genomic RNA (11). It has been shown that the synthesis of each subgenomic mRNA involves a discontinuous step by which the so-called 3’ body sequence is fused to the genomic 5’ leader sequence (22). The fusion of leader and body sequences during discontinuous transcription is determined, at least in part, by cis-acting elements termed transcription-regulatory sequences (TRS). These elements are located both at the 5’ end of the genome and at 5’-proximal sites corresponding to the individual transcription units (5). Although the mechanism for synthesis of sgRNAs is not fully understood, several models have been proposed. Two major models are leader-primed transcription (10, 12) and discontinuous transcription during minus-strand synthesis (19, 21), and the latter model has gained more support from recent evidence for the existence of transcriptionally active, subgenome-sized minus strands containing the antileader sequence and a transcription intermediate active in the synthesis of mRNAs (21, 22, 23, 24).

The genomes of many SARS CoV isolates have been sequenced, and they consist of approximately 29,700 nucleotides (13, 15, 16). Fourteen ORFs have been identified, of which 12 are located in the 3’-proximal one-third of the genome (13, 25). The exact mechanisms of expression of the 3’-proximal ORFs are unknown, but by analogy with other coronaviruses, these ORFs are expressed through a set of sgRNAs (15). Rota and colleagues could readily identify six sgRNAs, and later Thiel et al. demonstrated the existence of eight sgRNAs in SARS CoV-infected cells (15, 26). However, the exact number and molecular mechanism underlying the synthesis of SARS CoV sgRNAs have not been clarified yet. Therefore, identification of new sgRNAs and characterization of the molecular details of the leader-body fusion in the sgRNAs will help elu-
cidate the regulatory mechanism of SARS CoV transcription and replication, and this knowledge could further be used for development of antiviral therapeutic agents and a vaccine for the cure and prevention of this newly emerged disease.

In this study, we showed the coexistence of both plus- and minus-strand sgRNAs in SARS CoV-infected cells and identified 10 sgRNAs, including two novel subgenomic mRNAs (named 2-1 and 3-1) with noncanonical leader-body fusion sites.

### MATERIALS AND METHODS

**Virus and cells.** African green monkey kidney (Vero E6) cells and baby hamster kidney (BHK) cells were grown and maintained in Dulbecco’s modified Eagle medium and modified Eagle medium (Gibco Invitrogen Corp.), respectively, supplemented with 10% heat-inactivated fetal bovine serum (Gibco Invitrogen Corp.) and 100 U of penicillin and 100 μg of streptomycin (Gibco Invitrogen Corp.) per ml. Vero cells (4 × 10⁹) were infected at a multiplicity of 0.1 with SARS coronavirus strain WHU, which was isolated from a blood sample from a patient admitted to a local hospital with characteristic signs and symptoms of SARS (29). The complete genome sequence of SARS virus isolate WHU was determined in the previous study (29) (GenBank accession number AY394850).

**Northern blotting.** The total cellular RNA from SARS CoV-infected Vero E6 cells was extracted by using Trizol reagent (Invitrogen) according to the manufacturer’s instructions. Twenty micrograms of extracted RNA was fractionated in a 1.2% denaturing agarose gel containing 2.2 M formaldehyde with 1 M MOPS (morpholinepropanesulfonic acid) buffer (17), transferred to a nylon membrane (Hybond-A; Amersham Pharmacia), and UV cross-linked. The Northern blot was probed overnight at 42°C with 3²P-labeled strand-specific single-stranded DNA probes according to the protocol of the manufacturer (Amersham Pharmacia). The signals were detected and analyzed with a PhosphorImager and Image Quant software (Molecular Dynamics). The membrane was stripped of the first probe according to the protocols provided with the Hybond A membrane and was reprobed with the second probe.

The negative probe was complementary to the 3’ ends (positions 29421 to 29725) of SARS CoV mRNAs and was used to detect plus-sense sgRNAs. The positive probe, complementary to the 5’ ends of viral antisense RNAs, was used to detect minus-strand subgenomic RNAs.

**Reverse transcriptase PCR (RT-PCR) of SARS CoV minus-strand RNA and subgenomic mRNAs.** One microgram of total cellular RNA, extracted from SARS CoV-infected Vero cells, was reverse transcribed into single-stranded cDNA with Moloney murine leukemia virus reverse transcriptase (Promega). Oligo(dT)₁₅ or the strand-specific oligonucleotide SR18 (Table 1) was used to prime cDNA synthesis from plus-sense RNAs, while oligonucleotide SF9 (Table 1), which is complementary to the antileader sequence, was used for cDNA synthesis from minus-sense viral RNAs under the conditions recommended by the manufacturer (Promega). A 0.2-μl amount of cDNA product from the RT step was used for PCR. Primers for PCR (Table 1) were originally designed on the basis of the published SARS virus genome sequences of strains BJ01 (accession number AY278488), HKU (accession number AY278491), and Urbani (accession number AY278741) by using Oligo 4.1 (National Biosciences).

### TABLE 1. Oligonucleotides used for RT-PCR analysis of SARS CoV subgenomic RNAs

<table>
<thead>
<tr>
<th>Oligo(dT)</th>
<th>Sequence (5’ → 3’)</th>
<th>Positiona (region)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF8</td>
<td>CCGGAAAAAGCCCAACCCACC</td>
<td>20–39 (leader)</td>
</tr>
<tr>
<td>SF9</td>
<td>CTCGAGTCTTGTAGATCTTG</td>
<td>39–58 (leader)</td>
</tr>
<tr>
<td>SR10</td>
<td>CTTTCCGTACACCCCCGGAC</td>
<td>259–241 (5’ UTRb)</td>
</tr>
<tr>
<td>SR11</td>
<td>TCTGAAACATCAACCGGAAAAAG</td>
<td>22012–21991 (S)</td>
</tr>
<tr>
<td>SR12</td>
<td>TGTGCTTACAAGGCGAGCGCTAG</td>
<td>26097–26076 (3a and 3b)</td>
</tr>
<tr>
<td>SR13</td>
<td>AATTGTTTGTTTCTGGGTTGAATG</td>
<td>26748–26726 (M)</td>
</tr>
<tr>
<td>SR14</td>
<td>CGCAGCTGATAGGTATGTCG</td>
<td>27505–27486 (7a)</td>
</tr>
<tr>
<td>SR15</td>
<td>ACAAGTACGTCTCTAAATGCAG</td>
<td>28091–28068 (8b)</td>
</tr>
<tr>
<td>SR16</td>
<td>GGTGTGATTGGAACGCCTG</td>
<td>28348–28328 (N)</td>
</tr>
<tr>
<td>SF17</td>
<td>TGTAAACGTTCCTCAATGCGCA</td>
<td>29421–29442 (3c)</td>
</tr>
<tr>
<td>SR18</td>
<td>TTTGTCATTCTCCTAAGAAGC</td>
<td>29705–29725 (3a)</td>
</tr>
</tbody>
</table>

a Numbering refers to the nucleotide coordinates of the SARS virus isolate WHU sequence (accession no. AY394850).
b UTR, untranslated region.

c Junction sequence is six nucleotides but contains one mismatch.

### TABLE 2. Names of SARS-CoV mRNAs and ORFs or genes

<table>
<thead>
<tr>
<th>mRNA</th>
<th>ORP</th>
<th>Name used in reference(s):</th>
<th>Junction sequence</th>
<th>mRNA 1</th>
<th>1a</th>
<th>1a</th>
<th>1a</th>
<th>1a</th>
<th>1b</th>
<th>1b</th>
<th>1b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>25, 26</td>
<td>mRNA 2</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>mRNA 2-1</td>
<td>S’</td>
<td>5/6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>mRNA 3</td>
<td>3a</td>
<td>3a</td>
<td>3</td>
<td>X1</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>mRNA 3-1</td>
<td>3b</td>
<td>3b</td>
<td>4</td>
<td>X2</td>
<td>5/6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>mRNA 4</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>mRNA 5</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>mRNA 6</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>X3</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>mRNA 7</td>
<td>7a</td>
<td>7a</td>
<td>8</td>
<td>X4</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>mRNA 8</td>
<td>8a</td>
<td>8a</td>
<td>10</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>mRNA 9</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Genes that might be expressed from the mRNA.
b Number of nucleotides in the mRNA leader-body junction sequence.
c The junction sequence is six nucleotides but contains one mismatch.
RESULTS

Features of the genomic structure of SARS CoV isolate WHU. In the late period of the SARS outbreak, we isolated a SARS CoV isolate (named WHU) from a blood specimen from a SARS patient hospitalized in Hubei Province. The genome of SARS CoV WHU was completely sequenced, and the sequence was deposited in GenBank (accession number AY394850). It consisted of 29,725 nucleotides, excluding the poly(A) tail, and showed the typical genotypic features of the SARS CoV isolates that prevailed during the late epidemic period (29). This virus isolate was used throughout the studies reported here, and the sequence coordinates were based on the genomic sequence of WHU. The nomenclature of the mRNAs and genes followed the recommendations of the International Coronavirus Study Group (6), similar to those of Thiel et al. (26) and Snijder et al. (25). To avoid confusion due to different names for the same gene, a comparison with published names is shown in Table 2.

The genomic structure of SARS CoV isolate WHU (Fig. 1) is similar to those of other isolates but has a deletion of two nucleotides which correspond to nucleotides 28 and 29 of ORF 8a in SARS CoV Tor2 and Urbani (13, 15). The deletion was confirmed by sequencing multiple cDNA clones synthesized with viral RNAs prepared from different virus passages. This 2-nt deletion leads to a shifted ORF 8a of only 24 amino acids, with its first seven codons identical to those of other isolates. The genomic region of ORF 8a is the hot spot for deletions and additions during SARS CoV evolution (4, 8). The 2-nt deletion apparently did not influence virus replication in infected cells.

Coexistence of both plus- and minus-sense subgenomic RNAs in SARS CoV-infected cells. The expression of genetic information in members of the coronavirus family involves the synthesis of a variable number of subgenomic mRNAs depending on particular species of coronavirus (11). Sequence analysis showed that the SARS CoV genome contains at least 14 ORFs, of which 12 are located in the 3′-one-third of the genome and were predicted to be expressed from sgRNAs (Fig. 1). However, the initial analysis could identify only five sgRNAs (15), which are probably not sufficient to express the 12 ORFs. By Northern blot analysis with a radiolabeled strand-specific probe which detected plus-sense RNAs (Fig. 2A), we detected eight subgenomic mRNAs with approximate sizes of 8.3, 4.6, 3.7, 3.5, 2.9, 2.5, 2.0, and 1.7 kb. Similar results were obtained by Thiel et al. during the period of this study (26). Subsequently, we used a strand-specific DNA probe to detect minus-sense strand RNAs in infected cells. As shown in Fig. 2B, similar patterns of minus-sense sgRNAs could be seen, indicating the existence of minus-sense RNAs of subgenome length in SARS CoV-infected cells.

To further demonstrate that the bands detected in the Northern blot analysis are subgenome-length RNAs and not degradative or truncated products of genome-length RNAs and to provide conclusive evidence of the coexistence of both plus- and minus-sense sgRNAs, the leader-body junctions and surrounding regions of all of the sgRNAs detected by Northern blotting were amplified by RT-PCR and sequenced.

The rationale for the cloning is that the subgenomic mRNAs of coronaviruses are 3′-coterminal to the viral genome and possess a common leader sequence of about 70 nucleotides derived from the 5′ end of the viral genome. Thus, cloning of the junction region of the RNA leader and body sequences would reveal the existence of the corresponding sgRNAs. To clone each possible sgRNA, we designed two common primers.
RNA preparation was used for cDNA synthesis with primer intergenic TRS (Fig. 4B). Fig. 4A are identical to the conserved core elements in the Northern blots (Fig. 2). The junction sequences underlined in revealed the existence of all eight sgRNAs detected in the sequence data for the leader-body junction region (Fig. 4A) clones for each junction sequence were sequenced. The sequence confirmed the existence of minus-sense sgRNAs, the same RNA preparation was used for cDNA synthesis with primer SF8, which hybridizes only with the antileader sequence of the minus-strand sgRNAs. The subsequent PCR amplification and cloning were similar to those for plus-strand mRNAs. PCR fragment analysis gave rise to a band pattern (Fig. 3B) similar to that of plus-sense RNAs (Fig. 3A), and sequence analysis confirmed the identity of the individual junction sequences in the minus-strand RNAs. More minor bands were observed in some primer combinations, such as primer SR13 or 15 (Fig. 3B, lane 3, 4, and 5), and sequencing results showed that they represented either larger sgRNAs or nonspecific amplifications of cellular mRNAs or SARS CoV sequence. Taken together, the Northern blotting and sequencing results showed the coexistence of both plus- and minus-strand sgRNAs. Identification of novel subgenomic RNAs. The Northern blots could readily detect eight subgenomic RNA bands, and the sequencing confirmed their existence. PCR amplification is generally more sensitive to detect sgRNAs with low abundances. By carefully sequencing the minor bands amplified by RT-PCR (Fig. 3), two novel sgRNAs which could not be revealed in Northern blots were found, and these were named respectively) primers. The bands representing the specific SARS CoV sequences are indicated by arrowheads. The bands which revealed two novel subgenomic RNAs are boxed. Lanes 1, mRNA 2 (arrowhead) and mRNA 2-1 (boxed faint band in panel A); lanes 2, mRNA 3 (major band) and mRNA 3-1 (boxed); lanes 3, mRNA 5 (lower major band) and 4 (upper minor band); lanes 4, mRNA 7 (lower major band), mRNA 6 (middle band), and mRNA 5 (upper band); lanes 5, mRNA 8 (lower band), mRNA 7 (middle band), and mRNA 6 (upper band); lane 6, mRNA 9. (A). The cDNA used for PCR was made with oligo(dT) or SR18 primer, and thus the sequence of corresponding plus-strand RNA was amplified. (B). The cDNA used for PCR was made with primer SF8, which is complementary to the antileader sequence, and therefore the sequence of corresponding minus-strand RNA was amplified.
inside the S gene, 384 nucleotides downstream from the authentic core sequence (ACGAAC) for mRNA 2/S. In sgRNA 2-1, the first AUG codon is followed immediately by a stop codon, UAA, and the second AUG is 43 nt downstream and in the same reading frame of the S gene, which could result in the synthesis of a truncated S protein (named S\(_{\text{H11032}}\)) missing the N-terminal 143 amino acids. The corresponding ORF is named ORF 2b. By fusion of the green fluorescent protein gene with the 5\(_{\text{H11032}}\) part of sgRNA 2-1, a fusion protein could be detected, indicating the translatability of this sgRNA (data not shown). However, the existence of protein S\(_{\text{H11032}}\) in infected cells is yet to be determined.

The second novel sgRNA (3-1) (minor band in Fig. 3A, lane 2, and B, lane 2) corresponded to ORF 3b, which had been predicted to be expressed from mRNA 3 (25, 26). The leader-body fusion site (AAGAAC) for subgenomic mRNA 3-1 is 10 nucleotides upstream of the AUG start codon of ORF 3b and has a mismatch (underlined) with the leader core sequence (CS-L) (ACGAAC) of SARS CoV. Therefore, the existence of sgRNA 3-1 may indicate that ORF 3b could be expressed from a separate mRNA other than mRNA 3. The expression of 3b from sgRNA 3-1 was subsequently verified by fusion with the green fluorescent protein gene (data not shown).

The leader-body fusion sites of both sgRNA 2-1 (ACGAGC) and sgRNA 3-1 (AAGAAC) (Fig. 4A) have one nucleotide difference (underlined) from the core sequence (ACGAAC) in the leader TRS (TRS-L) of SARS CoV (26) but are identical to the core sequence of TRS-B (Fig. 4), which is consistent with previous findings that the core sequence in subgenomic mRNAs is derived from the body TRS but not from the leader TRS (20, 27, 30).

After sequencing of 12 independent clones of the junction region of mRNA 3-1, one clone showed a variant fusion site within or upstream of the AAA sequence motif (Fig. 5B), which is three nucleotides preceding the body core sequence (CS-B) (AAGAAC) for mRNA 3-1. This variant sequence was confirmed by sequencing another set of independent clones and may not represent a random event of template switching around the CS-B (AAGAAC) during the discontinuous transcription.

**Uniqueness and stability of the junction sequences of the subgenomic RNAs.** Comparison of the leader-body junction sequences of different sgRNAs showed that each subgenomic mRNA has a unique fusion site that is different from all others (Fig. 4A). The length of complementary sequences between TRS-L and TRS-B varies from 6 nucleotides in mRNA 6 to 12 nucleotides in mRNA 5/M, although they all contain the conserved hexanucleotide sequence (5\(\text{H11032}\)-ACGAAC-3\(\text{H11032}\)), except for the sequences of mRNAs 2-1 and 3-1, which have one nucleotide mismatch. The junction sequences of mRNAs 4 and 7 as well as mRNAs 3 and 8 have the same length but carry different extra nucleotides flanking the hexanucleotide core sequence (Fig. 4A). The uniqueness of the fusion sites of SARS CoV could play a regulatory role in controlling the abundances of different mRNAs (30).

We also analyzed viral RNAs prepared from viruses at different passages for 4 months. Sequencing of multiple cDNA clones of each mRNA showed that the fusion sites were stable for all mRNAs and did not change over time with virus passage.
DISCUSSION

Coronaviruses possess the largest RNA genomes, and the genes located on the 3′ parts of the genomes are expressed through a nested set of subgenomic mRNAs that are both 3′ and 5′ coterminal to the viral genome. There are about 12 ORFs in the 3′ one-third of the SARS CoV genome, and in this report we have characterized 10 subgenomic mRNAs, 2 of which have not been reported previously.

The synthesis of subgenomic mRNAs of coronaviruses involves a discontinuous step in which the 5′ leader and 3′ body sequences of mRNA are joined through the transcription-regulating sequences in the 3′ end of the leader and in the intergenic region preceding each mRNA body (30). Although the molecular details of the discontinuous RNA transcription are not completely known, the discovery of transcriptionally active, subgenomic-size minus strands containing antileader sequence (23, 24) favors the model of discontinuous transcription during the minus-strand synthesis (19). In this report, we have shown the coexistence of both plus- and minus-strand subgenomiclength RNAs in SARS CoV-infected cells, consistent with previous findings for other coronaviruses (1, 2). Therefore, the present data are more compatible with the discontinuous minus-strand synthesis model.

The proposed coronavirus discontinuous transcription mechanism implies a close interaction between leader TRS (TRS-L) and complementary body TRS (cTRS-B) in the intergenic region (30). The eight sgRNAs (mRNAs 2 to 9) of SARS CoV, which are easily detected by Northern blotting, possess junction sequences of 6 to 12 nucleotides, all containing the canonical core sequence (5′-ACGAAC-3′). The 100% identity of leader and body core sequences for these eight sgRNAs made it impossible to judge the origin of the junction sequences (from TRS-L or TRS-B) and the template switch site within the TRS. However, identification of two novel SARS CoV sgRNAs with noncanonical fusion sites shed light on these questions. The CS-Bs of both mRNA 2-1 (ACGAGC) and mRNA 3-1 (AAGAAC) (Fig. 4) contain one-nucleotide mismatches (underlined) with the CS-L (ACGAAC), but the sequence patterns of CS-B were retained in the junction region, indicating that the junction sequences of coronavirus sgRNAs originate from the CS-B, and this, in turn, supports the discontinuous minus-strand synthesis model. While the mismatch in mRNA 3-1 is at the second position of the hexanucleotides, the template switch can be envisaged to take place at the 3′ end of the nascent minus-strand RNA (Fig. 5A), again reinforcing the model of discontinuous transcription at minus-strand synthesis (19).

In this study, a rare species of mRNA 3-1 which contains a junction of only three nucleotides (AAA) was discovered (Fig. 5B and C). Although transcription of this subgenomic RNA could represent a rare event for SARS CoV, it did give more evidence for the use of noncanonical transcriptional signals in synthesis of sgRNAs. The template switching takes place at the sequence motif AAA, just preceding the leader core sequence.

FIG. 5. Junction sequences of SARS CoV mRNA 3-1 and models for template switch. The upper strand in the alignments represents the intergenic region of mRNA 3-1, and the lower strand is the genomic leader sequence. Dots indicate identity between the sequences. The conserved hexanucleotide core sequence is shaded, and the possible site for the template switch is indicated by arrow. The nucleotides in color are derived directly from the sequence profiles. (A) Leader-body fusion site of mRNA 3-1; (B and C) junction sequence and models of template switch of a rare variant of mRNA 3-1.
A similar truncated S protein has been reported for porcine respiratory coronavirus (3, 18). Currently, we are making efforts to construct an infectious cDNA clone of SARS CoV, and the use of reverse genetics will be helpful to elucidate the molecular mechanism of the discontinuous transcription and to reveal the biological functions of the new sgRNAs and their encoded proteins in the viral life cycle and pathogenesis.

ACKNOWLEDGMENTS

We thank Wei Jing and Lianwei Li for excellent technical assistance. We also thank other members of the SARS Task Force of Wuhan University, especially Congyi Zheng, Huimin Yan, Gengfu Xiao, and Jianmin Zhang, for fruitful collaboration.

This study was supported by China 973 basic research program grant 2003CB514102, China NSFC grant 30270313, and special SARS research funding from Hubei Province and Wuhan University. S.H. is supported by the China Scholarship Council and HEC Pakistan. D.G.’s lab is supported by the startup package and Luojia professorship program of Wuhan University.

REFERENCES

