JOURNAL OF VIROLOGY, May 2005, p. 5288–5295 Vol. 79, No. 9
0022-538X/05/$08.00+0 doi:10.1128/JVI.79.9.5288–5295.2005

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-

* Corresponding author. Mailing address: Modern Virology Research Centre, College of Life Sciences, Wuhan University, Wuhan 430072, People's Republic of China. Phone: 86-27-6875 2506. Fax: 86-27-6875 2897. E-mail: dguo@whu.edu.cn.

† S.H. and J.P. contributed equally to this work.

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

Primer	Sequence (5′ → 3′)	Position[a] (region)
Oligo(dT)	TTTTTTTTTTTTTTT	Poly(A) tail
SF8	CCAGGAAAAGCCAACCAACC	20–39 (leader)
SF9	CTCGATCTCTTGTAGATCTG	39–58 (leader)
SR10	CTTTCGGTCACACCCGGAC	259–241 (5′ UTR[b])
SR11	TCTGAAACATCAAGCGAAAAGG	22012–21991 (S)
SR12	TGTGCTTACAAGGGCACGCTAG	26097–26076 (3a and 3b)
SR13	AATGTTTGTTTCTGGGTTGAATG	26748–26726 (M)
SR14	CGCAGCTGATAGGTATGTCG	27505–27486 (7a)
SR15	ACAAGTACGTCTCTAAATGCAGCA	28091–28068 (8b)
SR16	GGTGTTGATTGGAACGCCCTG	28348–28328 (N)
SF17	TGTAAACGTTTTCGCAATTCCG	29421–29442 (3′ UTR)
SR18	TTTGTCATTCTCCTAAGAAGCTAT	29705–29725 (3′ UTR)

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

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 (4×10^7) were infected at a multiplicity of infection 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× 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 ^{32}P-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 SF8 (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).

Cloning and sequencing of the leader-body junction sequences of SARS CoV subgenomic mRNAs and minus-strand subgenomic RNAs. RT-PCR products were excised from the agarose gel and purified with a QIAquick gel extraction kit (Qiagen) as recommended by the manufacturer. Purified PCR fragment were ligated into the pGEM-T/T-Easy PCR cloning vector (Promega) or the pMD18-T PCR cloning vector (TakaRa) and transformed into *Escherichia coli* DH5α competent cells. Screening was done by colony PCR and restriction endonuclease digestion, and multiple independent cDNA clones were selected and sequenced for each species of subgenomic RNA. The sequencing reaction was carried out by using AmpliTaq DNA polymerase and universal primers with the Big Dye Terminator cycle sequencing ready reaction kit (PE Applied Biosystems) and analyzed on an ABI Prism 377 DNA sequencer (PE Applied Biosystems).

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

mRNA	ORF[a]	Name used in reference(s):			Junction sequence[b]
		25, 26	13	15	
mRNA 1	1a	1a	1a	1a	
	1b	1b	1b	1b	
mRNA 2	S	S	S	S	10
mRNA 2-1	S′				5/6[c]
mRNA 3	3a	3a	3	X1	11
	3b	3b	4	X2	
mRNA 3-1	3b				5/6
mRNA 4	E	E	E	E	8
mRNA 5	M	M	M	M	12
mRNA 6	6	6	7	X3	6
mRNA 7	7a	7a	8	X4	8
	7b	7b	9		
mRNA 8	8a	8a	10		11
	8b	8b	11	X5	
mRNA 9	N	N	N	N	9
	9b	9b	13		

[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.

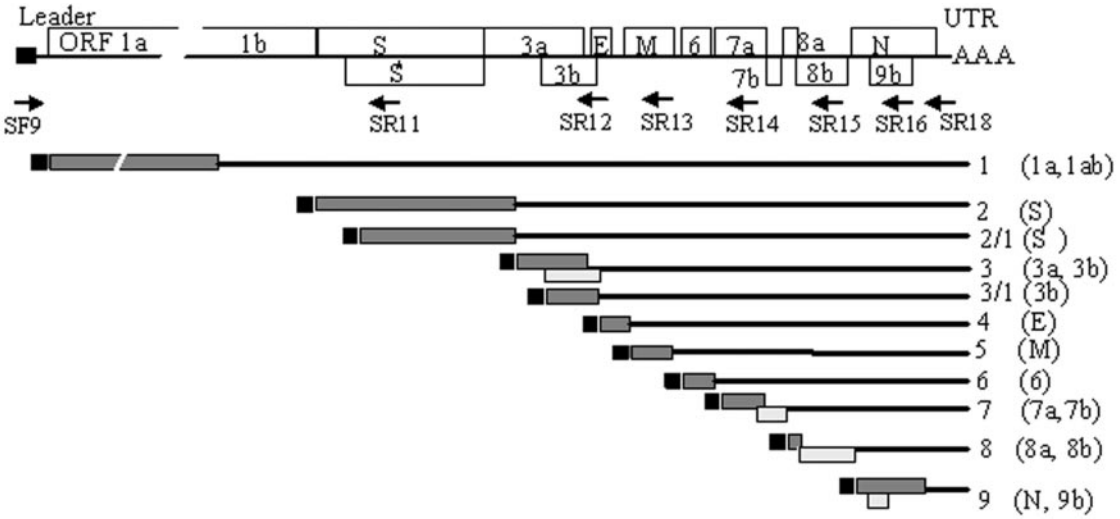

FIG. 1. Schematic representation of the genomic and subgenomic organizations of SARS CoV. The genome organization is based on the sequence of the SARS WHU isolate. In the upper panel, the genomic structure is shown. Known and potential ORFs are indicated by open boxes and are not to scale. The leader region is represented by a small solid box, and the poly(A) tail is represented by AAA. ORF 8a of isolate WHU contains a 2-nucleotide deletion and thus gives rise to a small ORF of 24 amino acids. Positions of forward (SF8 and -9) and reverse (SR11 to -18) primers used for cDNA synthesis and PCR amplification of different subgenomic RNAs are indicated by arrows under the genome. The bottom panel illustrates the 3′-coterminal nested set of mRNAs detected in this study. The small black boxes at the 5′ ends of the genomic and subgenomic RNAs represent the common leader sequence. The first (grey boxes) and second (open boxes) ORFs that are located at the 5′-proximal end and may be expressed from the mRNAs are shown.

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

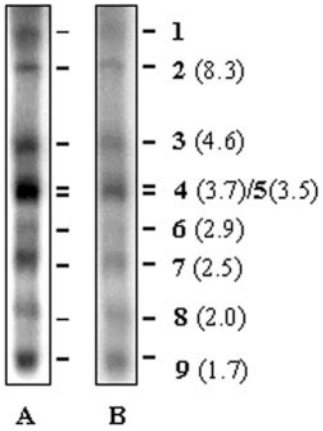

FIG. 2. Northern blot analysis of SARS CoV subgenomic mRNAs Twenty micrograms of total cellular RNA from SARS CoV-infected Vero E6 cells was separated by electrophoresis through a 1.2% denaturing agarose gel containing 2.2 M formaldehyde. The resolved RNA was transferred to a nylon membrane, and ^{32}P-labeled antisense and sense probes containing 305 nucleotides (positions 29421 to 29725) from the 3′ end of the SARS CoV genome were used to detect subgenomic mRNAs and minus-strand subgenomic RNAs, respectively. The mRNA designations and their approximate sizes (in parentheses) are indicated on the right. (A) Subgenomic mRNAs detected by the negative probe; (B) minus-strand subgenomic RNAs detected by the positive probe.

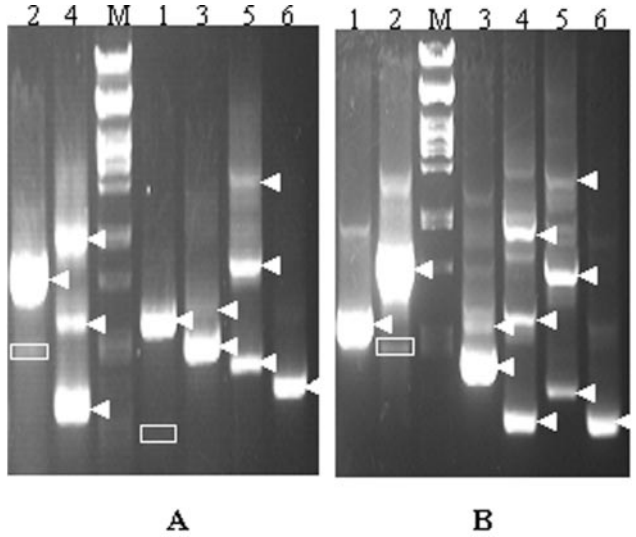

FIG. 3. RT-PCR analysis of SARS CoV subgenomic RNAs. One microgram of total cellular RNA from SARS CoV-infected Vero E6 cells was reverse transcribed and amplified by PCR with different combinations of forward (SF9) and reverse (SP11 to SP16 [lanes 1 to 6, respectively]) primers. The bands representing 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 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.

(SF8 and SF9) complementary to the antileader sequence and reverse specific primers (SR11 to -16) 200 to 600 nt downstream from the start codon of each ORF (Table 1 and Fig. 1). For cloning of the leader-body junction regions of plus-sense mRNAs, the cDNA was synthesized from total RNA of SARS CoV-infected cells with oligo(dT) and the strand-specific primer SR18. Using the common primer SP8 and one downstream specific primer, the junction sequences of various sgRNAs were amplified by PCR (Fig. 3A). As shown in Fig. 3A, all of the primer combinations amplified at least one major band of the expected size. It was not surprising that some primer combinations gave rise to multiple PCR bands, because the same primer combination could produce larger PCR fragments which correspond to the next-larger sgRNAs; e.g., primer SR14, intended to amplify the leader-body fusion sites of mRNA 7, could also amplify those of mRNAs 6 and 5 (Fig. 3A, lane 4), and primer SR15 could amplify the junction sequences of mRNAs 8, 7, and 6 (Fig. 3A, lane 5). It was expected that primer SR16, located in ORF N, could gave rise to multiple bands, but only one major band corresponding to the junction region of mRNA 9 was observed (Fig. 3A, lane 6). This was probably due to the high abundance of mRNA 9 and its preferred binding with the primer SR16.

The PCR fragments were individually isolated from the agarose gel and cloned for sequencing. At least 10 independent clones for each junction sequence were sequenced. The sequence data for the leader-body junction region (Fig. 4A) revealed the existence of all eight sgRNAs detected in the Northern blots (Fig. 2). The junction sequences underlined in Fig. 4A are identical to the conserved core elements in the intergenic TRS (Fig. 4B).

To confirm the existence of minus-sense sgRNAs, the same RNA preparation was used for cDNA synthesis with primer SF8, which hybridizes only with the antileader sequences 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 2-1 and 3-1 (Fig. 1 and 4). sgRNA 2-1 (minor band in Fig. 3A, lane 1) was obtained with primer SR11, which amplified the S junction region as the major band (Fig. 3A). The leader-body fusion site (ACGAGC) of sgRNA 2-1 (Fig. 4A) is located

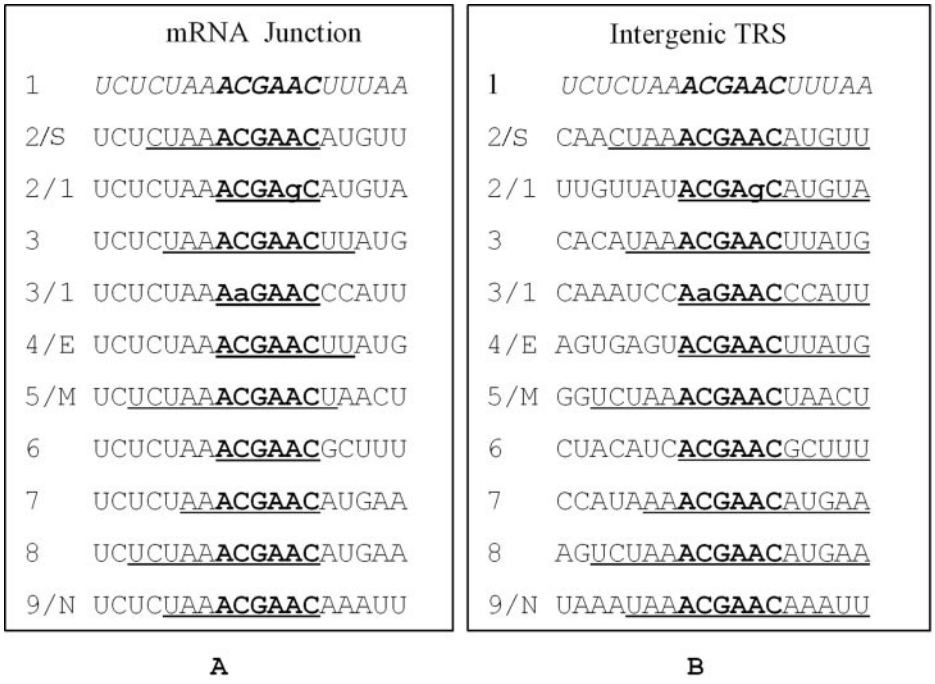

FIG. 4. Leader-body fusion sites of subgenomic mRNAs and their corresponding intergenic sequences. The 5′ genomic leader TRS is in italic. The hexanucleotide core sequence of the TRS is indicated in boldface, and the mismatched nucleotides with the leader core sequence (ACGAAC) are in lowercase. (A) Leader-body junction sites of subgenomic mRNAs in comparison with the genomic leader TRS. The junction sequences in subgenomic RNAs are underlined. (B) The TRS in the intergenic regions. The body sequences that are fused with the 5′ leader are underlined.

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′) 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′ 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′ 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′-ACGAAC-3′), 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.

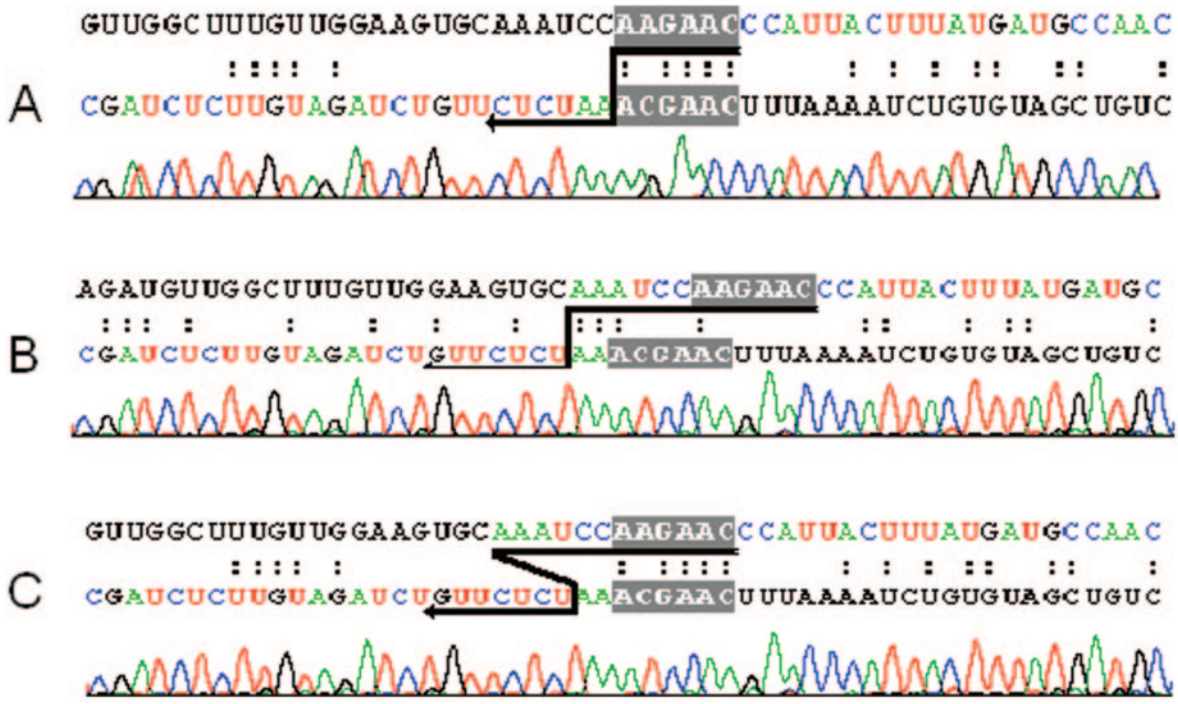

FIG. 5. Junction sequences of SARS CoV mRNA 3-1 and models for template switch. The upper strand in the alignments represents the intergenic region of mRNA 3-1, and the lower strand is the genomic leader sequence. Dots indicate identity between the sequences. The conserved hexanucleotide core sequence 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.

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 subgenome-length 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

(CS-L) and 3 nt upstream from the complementary body core sequence of mRNA 3-1. Two models could be proposed for the synthesis of this rare RNA: (i) the AAA is used directly as a transcription-regulating signal, and the complementarity between CS-L and CS-B takes place in the AAA region during the template switch step (Fig. 5B), and (ii) the interaction and complementarity between CS-L and CS-B are the same as those of mRNA 3-1, but the RNA polymerase can slide 4 nucleotides back on the leader template (Fig. 5C). In another coronavirus, mouse hepatitis virus, the UUA sequence was characterized as a noncanonical site for subgenomic RNA synthesis (28), implying that the AAA sequence in a specific sequence context in SARS CoV might also suffice for subgenomic synthesis and thus supporting the former model.

Although we have shown that the mRNAs 2-1 and 3-1 identified in this study could be functional messages, we have not identified their natural expression product in SARS CoV-infected cells due to the strict control on, and later prohibition against, using living SARS virus. According to the sequence of mRNA 2-1, it can lead to translation of a truncated S protein (S′). 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 Jiamin 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

1. **An, S., A. Maeda, and S. Makino.** 1998. Coronavirus transcription early in infection. J. Virol. **72:**8517–8524.
2. **Baric, R. S., and B. Yount.** 2000. Subgenomic negative-strand RNA function during mouse hepatitis virus infection. J. Virol. **74:**4039–4046.
3. **Callebaut, P., I. Correa, M. Pensaert, G. Jimenez, and L. Enjuanes.** 1988. Antigenic differentiation between transmissible gastroenteritis virus of swine and a related porcine respiratory coronavirus. J. Gen. Virol. **69:**1725–1730.
4. **Chinese SARS Molecular Epidemiology Consortium.** 2004. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science **303:**1666–1669.
5. **De Vries, A. F., M. C. Horzinek, P. J. M. Rottier, and R. J. de Groot.** 1997. The genome organization of the Nidovirales: similarities and differences between arteri-, toro-, and coronaviruses. Semin. Virol. **8:**33–47.
6. **Enjuanes, L., W. Spaan, E. Snijder, and D. Cavanagh.** 2000. Nidovirales, p. 827–834. *In* M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carsten, M. K. Estes, S. M. Lemon, D. J. McGeoch, J. Maniloff, M. A. Mayo, C. R. Pringle, and R. B. Wickner (ed.), Virus taxonomy. Classification and nomenclature of viruses. Academic Press, New York, N.Y.
7. **Fouchier, R. A. M., T. Kuiken, M. Schutten, G. Van Amerogen, G. J. van Doornum, B. G. van Hoogen, M. Peiris, W. Lim, K. Stohr, and A. D. Osterhaus.** 2003. Aetiology: Koch′s postulates fulfilled for SARS virus. Nature **423:**240.
8. **Guan, Y., B. J. Zheng, Y. Q. He, X. L. Liu, Z. X. Zhuang, C. L. Cheung, S. W. Luo, P. H. Li, L. J. Zhang, Y. J. Guan, K. M. Butt, K. L. Wong, K. W. Chan, W. Lim, K. F. Shortridge, K. Y. Yuen, J. S. Peiris, and L. L. Poon.** 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science **302:**276–278.
9. **Ksiazek, T. G., D. Erdman, C. S. Goldsmith, S. R. Zaki, T. Peret, S. Emery, S. X. Tong, C. Urbani, J. A. Comer, and W. Lim, P. E. Rollin, S. F. Dowell, A. E. Ling, C. D. Humphrey, W. J. Shieh, J. Guarner, C. D. Paddock, P. Rota, B. Fields, J. DeRisi, J. Y. Yang, N. Cox, J. M. Hughes, J. W. LeDuc, W. J. Bellini, L. J. Anderson, and the SARS Working Group.** 2003. A novel coronavirus associated with severe acute respiratory syndrome. N. Engl. J. Med. **348:**1953–1966.
10. **Lai, M. M., and D. Cavanagh.** 1997. The molecular biology of coronaviruses. Adv. Virus Res. **48:**1–100.
11. **Lai, M. M. C., and K. V. Holmes.** 2001. Coronaviruses, p. 1163–1185. *In* D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, and S. E. Straus (ed.), Fields virology. Lippincott Williams & Wilkins, Philadelphia, Pa.
12. **Makino, S., S. A. Stohlman, and M. M. C. Lai.** 1986. Leader sequences of murine coronavirus mRNAs can be freely reassorted: evidence for the role of free leader RNA in transcription. Proc. Natl. Acad. Sci. USA **83:**4204–4208.
13. **Marra, M. A., S. J. M. Jones, C. R. Astell, R. A. Holt, A. Brooks-Wilson, Y. S. N. Butterfield, J. Khattra, J. K. Asano, S. A. Barber, S. Y. Chan, A. Cloutier, S. M. Coughlin, D. Freeman, N. Girn, O. L. Griffin, S. R. Leach, M. Mayo, H. McDonald, S. B. Montgomery, P. K. Pandoh, A. S. Petrescu, A. G. Robertson, J. E. Schein, A. Siddiqui, D. E. Smailus, J. E. Stott, G. S. Yang, F. Plummer, A. Andonov, H. Artsob, N. Bastien, K. Bernard, T. F. Booth, D. Bowness, M. Czub, M. Drebot, L. Fernando, R. Flick, M. Garbutt, M. Gray, A. Grolla, S. Jones, H. Feldmann, A. Meyers, A. Kabani, Y. Li, S. Normand, U. Stroher, G. A. Tipples, S. Tyler, R. Vogrig, D. Ward, B. Watson, R. C. Brunham, M. Krajden, M. Petric, D. M. Skowronski, C. Upton, and R. L. Roper.** 2003. The genome sequence of the SARS-associated coronavirus. Science **300:**1399–1404.
14. **Peiris, J. S., C. M. Chu, V. C. Cheng, K. S. Chan, I. F. Hung, L. L. Poon, K. I. Law, B. S. Tang, T. Y. Hon, C. S. Chan, K. H. Chan, J. S. Ng, B. J. Zheng, W. L. Ng, R. W. Lai, Y. Guan, and K. Y. Yuen.** 2003. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. Lancet **361:**1767–1772.
15. **Rota, P. A., M. S. Oberste, S. S. Monroe, W. A. Nix, R. Campagnoli, J. P. Icenogle, S. Penaranda, B. Bankamp, K. Maher, M. H. Chen, S. X. Tong, A. Tamin, L. Lowe, M. Frace, J. L. Derisi, Q. Chen, D. Wang, D. D. Erdman, T. C. T. Peret, C. Burns, T. G. Ksiazek, P. E. Rollin, A. Sanchez, S. Liffick, B. Holloway, J. Limor, K. McCaustland, M. Olsen-Rasmussen, R. Fouchier, S. Gunther, A. D. M. E. Osterhaus, C. Drosten, M. A. Pallansch, L. J. Anderson, and W. J. Bellini.** 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science **300:**1394–1399.
16. **Ruan, Y. J., C. L. Wei, A. L. Ee, V. B. Vega, H. Thoreau, S. T. Su, J. M. Chia, P. Ng, K. P. Chiu, L. Lim, T. Zhang, C. K. Peng, E. O. Lin, N. M. Lee, S. L. Yee, L. F. Ng, R. E. Chee, L. W. Stanton, P. M. Long, and E. T. Liu.** 2003. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. Lancet **361:**1779–1785.
17. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Sanchez, C. M., G. Jimenez, M. D. Laviada, I. Correa, M. C. Sune, J. Bullido, F. Gebauer, C. Smerdou, P. Callebaut, J. M. Escribano, and L. Enjuanes.** 1990. Antigenic homology among coronaviruses related to transmissible gastroenteritis virus. Virology **174:**410–417.
19. **Sawicki, S. G., and D. L. Sawicki.** 1990. Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. J. Virol. **64:**1050–1056.
20. **Sawicki, S. G., and D. L. Sawicki.** 1995. Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands. Adv. Exp. Med. Biol. **380:**499–506.
21. **Sawicki, S. G., and D. L. Sawicki.** 1998. A new model for coronavirus transcription. Adv. Exp. Med. Biol. **440:**215–220.
22. **Sawicki, D. L., T. Wang, and S. G. Sawicki.** 2001. The RNA structures engaged in replication and transcription of the A59 strain of mouse hepatitis virus. J. Gen. Virol. **82:**385–396.
23. **Schaad, M. C., and R. S. Baric.** 1994. Genetics of mouse hepatitis virus transcription: evidence that subgenomic negative strands are functional templates. J. Virol. **68:**8169–8179.
24. **Sethna, P. B., S. L. Hung, and D. A. Brian.** 1989. Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. Proc. Natl. Acad. Sci. USA **86:**5626–5630.
25. **Snijder, E. J., P. J. Bredenbeek, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. Poon, Y. Guan, M. Rozanov, W. J. Spaan, and A. E. Gorbalenya.** 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J. Mol. Biol. **331:**991–1004.
26. **Thiel, V., K. A. Ivanov, A. Putics, T. Hertzig, B. Schelle, S. Bayer, B. Weissbrich, E. J. Snijder, H. Rabenau, H. W. Doerr, A. E. Gorbalenya, and J. Ziebuhr.** 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. J. Gen. Virol. **84:**2305–2315.
27. **Van Marle, G., J. C. Dobbe, A. P. Gultyaev, W. Luytjes, W. J. Spaan, and E. J. Snijder.** 1999. Arterivirus discontinuous mRNA transcription is guided

by base pairing between sense and antisense transcription-regulating sequences. Proc. Natl. Acad. Sci. USA **96:**12056–12061.
28. **Zhang, X., and R. Liu.** 2000. Identification of a noncanonical signal for transcription of a novel subgenomic mRNA of mouse hepatitis virus: implication for the mechanism of coronavirus RNA. Virology **278:**75–85.
29. **Zhu, Y., M. Liu, W. Zhao, J. Zhang, X. Zhang, K. Wang, C. Gu, K. Wu, Y. Li, C. Zheng, G. Xiao, H. Yan, J. Zhang, D. Guo, P. Tien, and J. Wu.** 2005. Isolation of virus from a SARS patient and genome-wide analysis of genetic mutations related to pathogenesis and epidemiology from 47 SARS CoV isolates. Virus Genes **30:**93–102.
30. **Zuniga, S., I. Sola, S. Alonso, and L. Enjuanes.** 2004. Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. J. Virol. **78:**980–994.